

A Recurrent Neural Network Provides Stable Across-Day Prosthetic Control for a Human Amputee with Implanted Intramuscular Electromyographic Recording Leads

Caleb J. Thomson, *Student Member, IEEE*, Gregory A. Clark, and Jacob A. George, *Member, IEEE*

Abstract— Upper-limb prosthetic control is often challenging and non-intuitive, leading to up to 50% of prostheses users abandoning their prostheses. Convolutional neural networks (CNN) and recurrent long short-term memory (LSTM) networks have shown promise in extracting high-degree-of-freedom motor intent from myoelectric signals, thereby providing more intuitive and dexterous prosthetic control. An important next consideration for these algorithms is if performance remains stable over multiple days. Here we introduce a new LSTM network and compare its performance to previously established state-of-the-art algorithms—a CNN and a modified Kalman filter (MKF)—in offline analyses using 76 days of intramuscular recordings from one amputee participant collected over 425 calendar days. Specifically, we assessed the robustness of each algorithm over time by training on data from the first (one, five, ten, 30, or 60) days and then testing on myoelectric signals on the last 16 days. Results indicate that training on additional datasets from prior days generally decreases the Root Mean Squared Error (RMSE) of intended and unintended movements for all algorithms. Across all algorithms trained with 60 days of data, the lowest RMSE for unintended movements was achieved with the LSTM. The LSTM also showed less across-day variance in RMSE of unintended movements relative to the other algorithms. Altogether this work suggests that the LSTM algorithm introduced here can provide more intuitive and dexterous control for prosthetic users, and that training on multiple days of data improves overall performance on subsequent days, at least for offline analyses.

I. INTRODUCTION

In the United States 1.6 million individuals have lost a limb due to either dysvascular diseases like diabetes (54%) or to trauma (45%) [1]. For nearly one in every 200 individuals, limb-loss is a life-long struggle with chronic pain, depression, and functional disability [1], [2]. State-of-the-art prostheses have the capability to mimic the complex movements of the human hand. However the majority of commercial prostheses use only two electromyographic (EMG) electrodes to control up to two degrees of freedom [3], [4]. Unsatisfied with the current standard of care, up to 50% of upper-limb amputees abandon their prostheses [5], citing ineffective control as a primary reason [6].

Even though the physical hand is missing after an amputation, amputees still retain the neural circuits and, in the

case of most transradial amputees, the forearm musculature that are used to control the hand. Neural and EMG activity from the residual nerves and muscles can be recorded through implanted electrodes and then correlated to motor intent under a supervised learning approach.

A variety of different algorithms have been used to correlate these bioelectric signals to motor intent in order to classify discrete hand grasps or regress continuous joint angles. These algorithms generally fall into the broad categories of Wiener filters, population vectors, probabilistic methods, and recursive Bayesian decoders [7]. Previous works by this group have used a modified Kalman filter (MKF) [8] and convolutional neural networks (CNNs) [9], [10] to regress continuous joint angles and control multiple degrees of freedom of a prosthesis in real-time.

Our group and others have shown that recurrent neural networks—that learn temporal dependencies in the bioelectric signals (e.g., long-short term memory networks; LSTMs)—can provide more accurate prosthetic control [11], [12]. However, the vast majority of prosthetic control algorithms were trained and tested using data collected within the span of a single day. The few prior works that have investigated across-day performance used algorithms that were trained on a single dataset from one day [11], [13]. Here, we specifically sought to answer the question if training on aggregate data from prior days improves or degrades performance on subsequent days. Training on more data generally improves performance [14], but day-to-day variations in training data have also been shown to degrade performance [13].

We demonstrate that training on data from multiple prior days improves overall performance on subsequent days and reduces the variance of algorithm performance day to day. We also show LSTM networks result in the best overall performance, regardless of the amount of training data used. This work constitutes an important step towards the development of robust prosthetic control and has broad implications for how training data should be collected for neuroprostheses.

II. METHODS

A. Human Subjects and Implanted Devices

Data from one transradial amputee was used in this study.

J. A. George (jacob.george@utah.edu) is with the Electrical Engineering Department, University of Utah, Salt Lake City, UT 84112, and the Division of Physical Medicine and Rehabilitation, University of Utah, Salt Lake City, UT 84132 USA.

*This work was funded by: NIH, NIDCR, NICHD, Office of the Director, Award Number DP5OD029571. Additional support provided by DARPA, BTO, Hand Proprioception and Touch Interfaces program, Space and Naval Warfare Systems Center, Pacific, Contract No. N66001-15-C-4017;

C. J. Thomson (corresponding author; caleb.j.thomson@utah.edu) and G. A. Clark are with the Department of Biomedical Engineering, University of Utah, Salt Lake City, UT 84112 USA

The participant, referred to as Subject S6 in prior work, was a 57-year-old male whose left foot and left forearm were amputated 13 years prior due to trauma [15]. The participant had 32 intramuscular electromyographic recording electrodes (iEMGs; Ripple LLC, Salt Lake City, UT, USA) implanted into their residual arm muscles for 425 days. Additional information about the implanted devices and surgical procedures can be found in [16]. Informed consent and experimental protocols were carried out in accordance with the University of Utah Institutional Review Board.

B. Data Collection

A total of 76 datasets were collected from the participant over the span of the study. Each dataset consisted of prerecorded movements of a 12-degree-of-freedom virtual bionic arm or 6-degree-of-freedom physical prosthesis and EMG recordings while the participant actively attempted to mimic the prerecorded movements with their phantom limb. In general, datasets consisted of participants mimicking individual movements of flexion and extension of the five fingers and wrist, thumb intrinsic movement, and pronation and supination of the wrist. Each movement was repeated five to ten times. EMG recordings were sampled at 1 kHz using the Grapevine System (Ripple LLC, Salt Lake City, UT, USA). The 32 channels of continuous EMG signals were band-pass filtered with cutoff frequencies of 15 Hz (sixth-order high-pass Butterworth filter) and 375 Hz (second-order low-pass Butterworth filter). Notch filters were applied at 60, 120, and 180 Hz. Differential EMG signals were calculated for all possible pairs of channels, resulting in 496 (32 choose two) differential recordings. The mean absolute value over a sliding 300-ms window was then calculated at 30 Hz for all of the single-ended channels and differential pairs. The final training data consisted of 12 kinematic recordings from the predetermined movements of the prosthesis and the mean absolute value of 528 EMG channels. Additional information on the data collection and processing can be found in [6].

B. Training Data

A total of 76 datasets were collected from the participant. The last 16 datasets (datasets 61-76) were used exclusively for testing the algorithms (Fig. 1). Each algorithm was trained with five different amounts of data: 1) a single prior dataset (dataset 60); 2) five prior datasets (datasets 56-60); 3) ten prior datasets (datasets 51-60); 4) 30 prior datasets (datasets 31-60); and 5) 60 prior datasets (datasets 1-60). When training with multiple datasets, the data from each dataset were concatenated together and treated as a single large dataset.

C. Motor-decode Algorithms

Three motor-decode algorithms were implemented in MATLAB 2020B. Two have been presented previously, a MKF [8] and a shallow CNN [9]. The third algorithm included in the comparison was a novel LSTM network. The input to the LSTM consisted of the 528 EMG channels at the current point in time. The LSTM architecture consisted of a single LSTM layer with 128 hidden units, three fully connected layers with 1056 units, ReLu activation between layers and a regression output. The regression output resulted in 12 continuous values, one for each possible degree of freedom of the prosthesis. The LSTM and CNN were trained using Stochastic Gradient Descent with a learning rate of 0.01 and 0.00001, respectively. For the MKF, 100% of the training data was used to train the algorithm. For the CNN and LSTM, 60% of the training data was used for training and the remaining 40% was used for validation. Both the CNN and LSTM were trained for a maximum of 2000 epochs. The final output of the CNN was not modified with a threshold as has been done in the group's previous work [9].

D. Performance Metrics

The control algorithms were tested on 16 novel, unseen datasets (datasets 61-76). Performance was measured by the root mean squared error (RMSE) between the algorithm predictions and the target kinematics that the participant was attempting to mimic. The RMSE was divided into two categories: intended movement RMSE and unintended movement RMSE, as described in [8]. Intended movement RMSE measures the ability of the algorithm to replicate the participants desired movements (e.g., flexion or extension of a single DOF), whereas unintended movement RMSE measures the ability to eliminate cross-talk such that only the intended degree of freedom is active. The median RMSE from the intended and unintended movements was calculated for each dataset, such that comparisons were made against the RMSEs of each algorithm across all of the testing days ($N = 16$).

A one-way non-parametric ANOVA (Kruskal-Wallis) was used to compare the three motor-decode algorithms (MKF, CNN, and LSTM). Separate ANOVAs were performed for intended and unintended movement RMSE. If any significance was found, subsequent pairwise comparisons were made using the Tukey-Kramer correction for multiple comparisons. Variance between motor-decode algorithms trained on 60 datasets was compared using a nonparametric two-sample F-test for equal variances (Ansari-Bradley test).

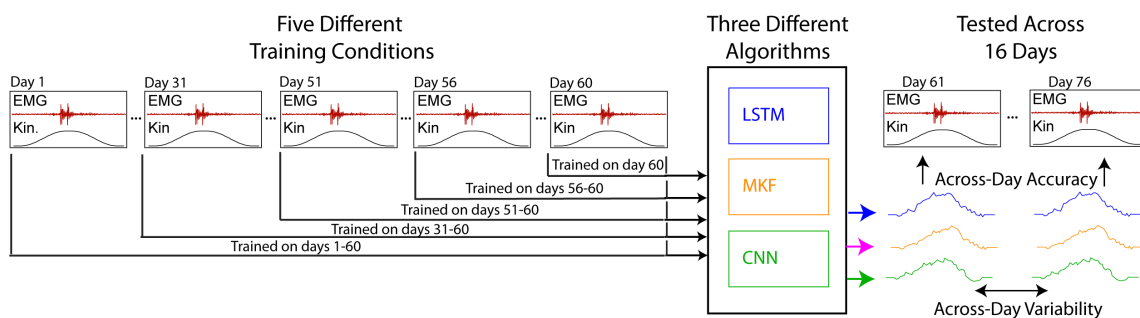


Figure 1. Three different algorithms (LSTM, MKF and CNN) were trained using data from the one, five, ten, 30, or 60 most recent datasets. Data consisted on EMG recordings (EMG) and the participants desired kinematics (Kin). The algorithms were tested using data from the 16 unseen datasets. Performance metrics included across-day accuracy relative to the ground truth kinematics and across-day variability.

III. RESULTS

A. Training on prior datasets increases the accuracy of control algorithms on subsequent days

We found that training on additional datasets from prior days improved the motor-decode algorithms' accuracy. The CNN and LSTM had less intended movement RMSE when trained on the most recent 60 datasets compared to training on the single most recent dataset (p 's < 0.05, Fig. 2). The additional training datasets did not significantly change the intended movement RMSE of the MKF. Furthermore, the additional training datasets significantly reduced the unintended movement RMSE of the MKF and LSTM (training on one day vs training on 60 days; p 's < 0.05, Fig. 3), and trended towards reducing the unintended movement RMSE of the CNN ($p = 0.08$, Fig. 3).

B. MKF produces more accurate intended movements when training data is limited

We found that when only a single dataset was used for training, the MKF generally performed the best of the three algorithms. The MKF had significantly lower intended movement RMSE than the LSTM ($p < 0.05$) and trended towards having lower intended movement RMSE than the CNN ($p = 0.11$).

C. CNN and LSTM produce more accurate intended movements when training data is abundant

In contrast, when trained on 60 datasets the CNN and LSTM generally performed the best. The LSTM and CNN had significantly less intended movement RMSE than the MKF (p 's < 0.05, Fig. 2). No significant differences were found between the intended movement RMSEs of the LSTM and the CNN when trained on 60 datasets.

D. LSTM has substantially less unintended movement and less across-day variability compared with the CNN and MKF

We found that the LSTM consistently resulted in less cross-talk. That is, across all amounts of training data, the

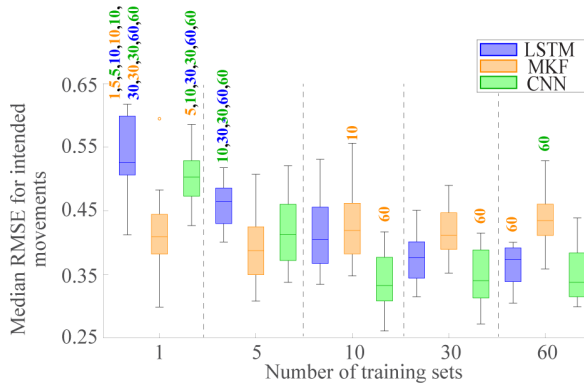


Figure 2. The median RMSE for intended movements of the testing datasets was reduced with increased amounts of training data for the CNN and LSTM. The MKF showed little change in intended movement RMSE with increased training data. Numbers listed above a given condition represent statistically significant differences relative to other conditions (p 's < 0.05, pairwise rank-sum tests). The number denotes the compared training condition (number of training sets) and the color denotes the compared control algorithm. The MKF was generally better when trained with only a single dataset and was significantly worse when trained using the full 60 datasets.

LSTM consistently had significantly less unintended movement RMSE than the CNN and MKF (p 's < 0.05, Fig. 3). Even though all algorithms had less unintended when trained with the full 60 datasets, the LSTM still outperformed the CNN and MKF (p 's < 0.05, Fig. 4). Furthermore, the LSTM also showed less across-day variance in unintended movement RMSE than the CNN and MKF (p 's < 0.05, Fig. 4).

IV. DISCUSSION

Neural and electromyographic recordings from implanted devices change day to day such that training data from prior days does not perform well on subsequent days [13]. Despite these day-to-day variations, we found that training on multiple days of data improved the accuracy of all algorithms. That is, including additional data from prior days (even if the old data is no longer the most accurate) resulted in less unintended movements for the MKF, CNN and LSTM. The CNN and LSTM (but not MKF) also demonstrated less error in intended movements when additional data from prior days was used for training. Overall, when trained on the most recent 60 datasets, the LSTM had the best performance, suggesting that it may lead to the best real-time, human-in-the-loop prosthetic control.

Previous work from this group and others have shown that neural networks improve with increased amounts of data [10], [17]. The improvements seen with the CNN and LSTM are consistent with these prior results and provide additional motivation to leverage multi-day recordings when training prosthetic control algorithms, even if recordings vary day to day.

Our findings also suggest that LSTMs may be a favorable control algorithm with similar performance to CNNs in intended movements, but better performance at eliminating unintended movements. This differs from the results presented in [11] where the researchers found CNNs to perform better than LSTMs. This may be due to the fact that the networks used in [11] were deeper than the those presented here, with the LSTM having four LSTM layers, and the CNN having two convolutional layers. In addition, both the LSTM and CNN had two main dataflows, one for EMG and one for recent

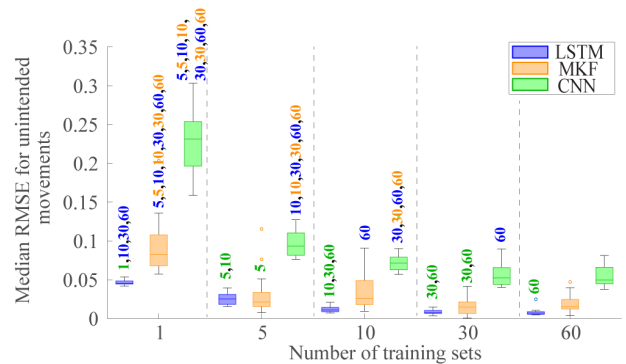


Figure 3. The median RMSE for unintended movements of the testing datasets was reduced with increasing amounts of training for all motor decode algorithms. The LSTM shows the smallest unintended movement RMSE for each number of training datasets. Numbers listed above a given condition represent statistically significant differences relative to other conditions (p 's < 0.05, pairwise rank-sum tests). The number denotes the compared training condition (number of training sets) and the color denotes the compared control algorithm.

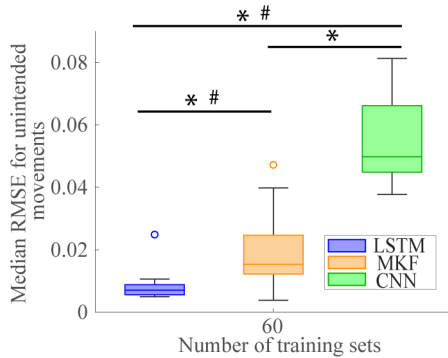


Figure 4. The median RMSE for unintended movements of the testing datasets under the best possible training condition (i.e., the most data possible). The LSTM had significantly less unintended movement RMSE relative to the CNN and MKF (*, p 's < 0.05, pairwise rank-sum tests). The LSTM also had significantly less variance than the CNN and the MKF (#, p 's < 0.005, pairwise Ansari-Bradley tests).

kinematic position. The LSTM and CNN in [11] were also trained on only a single dataset under a different training paradigm.

Prior work directly compared the CNN and MKF reported here on various activities of daily living and found they performed similarly [18]. However, clinical considerations favored the MKF since it is faster to train and less computationally expensive than the CNN [18]. Consistent with that, the MKF was the quickest control algorithm to train (Table 1). That said, the amount of time needed to train the LSTM and CNN would likely be appropriate when training on datasets across multiple days since training could take place between days while the participant uses an algorithm trained on subsequent days. Furthermore, the prediction times of the LSTM and CNN algorithms are both within the actuation and control speed of the prostheses (33-ms update speed). The LSTM also provided better across-day control than the MKF, specifically with unintended movements. Less variability would likely be advantageous in allowing users to adapt to and learn with their prosthesis.

The results presented here consist of offline analyses from one amputee participant. It is not clear if the results seen here will be applicable to the general population of prosthesis users. It is also unclear how multi-day training and the LSTM will perform during real-time human-in-the-loop control. Future work will validate these approaches with multiple participants performing activities of daily living.

V. CONCLUSION

This work demonstrates strengths of LSTM networks for extracting motor intent from biological signals. In addition,

Table 1. Computational Time Taken to Train the Algorithms in Minutes (mean \pm STD)

Number of training datasets	LSTM	MKF	CNN
1	14.59 \pm 1.61	0.12 \pm 0.01	17.13 \pm 0.35
5	111.91 \pm 13.14	0.84 \pm 0.00	130.50 \pm 0.44
10	269.13 \pm 28.46	2.12 \pm 0.01	318.72 \pm 1.00
30	856.23	7.91 \pm 0.12	1146.63
60	1822.24	16.17 \pm 0.28	2434.34

this work highlights the importance of training prosthetic control algorithms on multiple datasets across many days – a feature that is uniquely beneficial to non-linear algorithms like the LSTM. Altogether, the LSTM trained with 60 prior datasets provided significantly better performance than two state-of-the-art algorithms previously reported.

REFERENCES

- [1] K. Ziegler-Graham *et al.*, “Estimating the Prevalence of Limb Loss in the United States: 2005 to 2050,” *Arch. Phys. Med. Rehabil.*, vol. 89, no. 3, pp. 422–429, Mar. 2008.
- [2] C. G. Bhuvaneshwar *et al.*, “Reactions to Amputation: Recognition and Treatment,” *Prim. Care Companion CNS Disord.*, vol. 9, no. 4, pp. 0–0, Aug. 2007.
- [3] “Mobius Bionics LLC. LUKE Arm System.” [Online]. Available: <https://www.mobiusbionics.com/wp-content/uploads/2019/09/Mobius-Bionics-LUKE-Product-Spec-Sheet.pdf>. [Accessed: 19-Jul-2021].
- [4] “Motion Control Division of Fillauer. Utah Arm U3 and U3+ User Guide.” [Online]. Available: <https://fillauer.com/wp-content/uploads/2020/09/1910025-USER-GUIDE-Utah-Arm-U3-and-U3-Plus-Rev-C-08-24-2020.pdf>. [Accessed: 19-Jul-2021].
- [5] E. A. Biddiss *et al.*, “Upper limb prosthesis use and abandonment: A survey of the last 25 years,” *Prosthet. Orthot. Int.*, vol. 31, no. 3, pp. 236–257, Sep. 2007.
- [6] E. Biddiss *et al.*, “Upper-Limb Prosthetics: Critical Factors in Device Abandonment,” *Am. J. Phys. Med. Rehabil.*, vol. 86, no. 12, pp. 977–987, Dec. 2007.
- [7] K. Anam *et al.*, “Myoelectric control systems for hand rehabilitation device: A review,” in *2017 4th International Conference on Electrical Engineering, Computer Science and Informatics (EECSI)*, 2017, pp. 1–6.
- [8] J. A. George *et al.*, “Intuitive neuromyoelectric control of a dexterous bionic arm using a modified Kalman filter,” *J. Neurosci. Methods*, vol. 330, p. 108462, Jan. 2020.
- [9] J. A. George *et al.*, “Improved Training Paradigms and Motor-decode Algorithms: Results from Intact Individuals and a Recent Transradial Amputee with Prior Complex Regional Pain Syndrome,” in *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2018, pp. 3782–3787.
- [10] J. George *et al.*, “Inexpensive surface electromyography sleeve with consistent electrode placement enables dexterous and stable prosthetic control through deep learning,” *MEC20 Symp.*, Jul. 2020.
- [11] H. Dantas *et al.*, “Deep Learning Movement Intent Decoders Trained With Dataset Aggregation for Prosthetic Limb Control,” *IEEE Trans. Biomed. Eng.*, vol. 66, no. 11, pp. 3192–3203, Nov. 2019.
- [12] M. Jabbari *et al.*, “EMG-Based Hand Gesture Classification with Long Short-Term Memory Deep Recurrent Neural Networks,” in *2020 42nd Annual International Conference of the IEEE Engineering in Medicine Biology Society (EMBC)*, 2020, pp. 3302–3305.
- [13] S. M. Wendelken, “Dexterous Control of a Hand Prosthesis Using Neuromuscular Signals from Implanted or Surface Electrodes,” Ph.D. dissertation, University of Utah, Salt Lake City, UT, 2017.
- [14] A. Phinyomark *et al.*, “EMG Pattern Recognition in the Era of Big Data and Deep Learning,” *Big Data Cogn. Comput.*, vol. 2, no. 3, p. 21, Sep. 2018.
- [15] J. A. George *et al.*, “Biomimetic sensory feedback through peripheral nerve stimulation improves dexterous use of a bionic hand,” *Sci. Robot.*, vol. 4, no. 32, p. eaax2352, Jul. 2019.
- [16] D. M. Page *et al.*, “Motor Control and Sensory Feedback Enhance Prosthesis Embodiment and Reduce Phantom Pain After Long-Term Hand Amputation,” *Front. Hum. Neurosci.*, vol. 12, 2018.
- [17] M. Zia ur Rehman *et al.*, “Multiday EMG-Based Classification of Hand Motions with Deep Learning Techniques,” *Sensors*, vol. 18, no. 8, p. 2497, Aug. 2018.
- [18] M. D. Paskett *et al.*, “Activities of daily living with bionic arm improved by combination training and latching filter in prosthesis control comparison,” *J. NeuroEngineering Rehabil.*, vol. 18, no. 1, p. 45, Dec. 2021.