Temporal Dilation of Deep LSTM for Agile Decoding of sEMG: Application in Prediction of Upper-Limb Motor Intention in NeuroRobotics

Tianyun Sun, Qin Hu, *Student Member, IEEE*, Paras Gulati, and S. Farokh Atashzar, *Member, IEEE*

***Abstract*—The spectrotemporal information content of surface electromyography has shown strong potential in predicting the intended motor command. During the last decade, with accelerated exploitation of powerful deep-learning techniques aligned with advancements in active prostheses and neurorobots, a great deal of interest has been drawn to the development of intelligent myoelectric prostheses with an ultimate resolution of upper-limb gestures prediction. Recent research involves Deep CNNs, RNNs, and hybrid frameworks, which have shown promising results. However, deep-learning models have almost always been challenged by the structural complexity, the large number of trainable parameters, concerns of overfitting, and prolonged training time, which complicate the practicality and limit the outcomes. In this letter, for the first time, we propose temporal-dilation in the LSTM module of a hybrid Deepnet model for sEMG-based gesture detection, hypothesizing improved accuracy and training agility. We also analyze the effect of dilation-aggressiveness. We conduct systematic and statistical analysis on the efficacy of the proposed approach in comparison to recent literature, including our previous work. this letter shows that the proposed temporally-dilated LSTM model wins over the recent deep-learning techniques in terms of accuracy, and more significantly, it reduces the training time while increasing the convergence speed, with the ultimate goal of maximizing practicality and translational value for neurorobotic systems.**

***Index Terms*—Electromyography, machine learning, medical robotics, multi-layer neural network, prosthetics, recurrent neural networks, and convergence.**

I. Introduction

IT IS estimated that more than 2.1 million people in the US suffer from the loss of a biological limb, and the number is expected to be double by 2050 [1]. Significant research efforts have been conducted to develop human-machine interfaces for neurorobotic systems [2] such as those in assistive exoskeletons [3]–[7], and powered prostheses [8]–[13], with the ultimate goal of augmenting human performance and bypassing physiological and pathological barriers, including loss of a biological limb. In this regard, the mechatronic designs of neurorobotic systems have been significantly improved [11], [14] from simple old-fashioned cosmetic non-functional prostheses to sophisticated multifunction prostheses. Some commercial examples are i-Limb [15], Bebionic hand [16], and LUKE arm [17]. Despite the significant progress related to the mechatronic design (e.g., recent soft bionic hands [18]–[20] and soft exosuits [21]–[23]), the existing commercial neurorobotic technologies suffer from an inert human-machine interface (HMI), which has resulted in excessive cognitive and visual loads on the user, low quality of control, and accordingly, high technology rejection rate among the users [24]–[27]. The lack of adequate accuracy, resolution, adaptability, agility, and intelligence in interfacing with human neurophysiology has resulted in limited performance and heterogeneous response, especially under unstructured and highly dynamic conditions, in real life.

Advanced biological signal processing and machine-learning techniques have shown great potential for revolutionizing the field of human-machine integration, motivated by the unmet need to have intuitive, agile, transparent, scalable, adaptable, affordable, and bidirectional means of interfacing between human neurophysiology and machine intelligence. In this regard, surface electromyography (sEMG), a non-invasive biosignal collection technique, has been used to control advanced myoelectric prostheses [28].

To control upper-limb prostheses, hand gesture recognition (HGR) has attracted a great deal of interest [29]. Classical approaches utilize the quantifiable features and the corresponding patterns for classifications. Spectral and temporal features have been used in the literature to perform classical machine-learning techniques on sEMG [30]–[36]. Examples of temporal features are Mean Absolute Value, Zero Crossing, and Waveform Length. Examples of spectral features are Cepstral Coefficient, Marginal Discrete Wavelet Transform. In addition, spectrotemporal features have been extracted from spectrograms generated by short-time Fourier transform (STFT) and continuous wavelet transform (CWT).

Deep-learning methods have been recently introduced to process multichannel sEMG for HGR [37]–[39]. High accuracies have been reported on a large number of gestures, such as in [40], [41]. Both Recurrent Neural Networks (RNN) [42], [43] and Convolutional Neural Networks (CNN) [44]–[46] have been

Manuscript received February 24, 2021; accepted June 14, 2021. Date of publication June 23, 2021; date of current version July 9, 2021. This letter was recommended for publication by Associate Editor Dr. Tommaso Lenzi and Editor Prof. Pietro Valdastri upon evaluation of the reviewers' comments. This work was supported by the US National Science Foundation under Grants 2037878 and 2031594. *(Corresponding author: S. Farokh Atashzar.)*

Tianyun Sun, Qin Hu, and Paras Gulati are with the Department of Electrical and Computer Engineering, New York University (NYU), New York, NY 10003 USA (e-mail: ts3907@nyu.edu; qh503@nyu.edu; paras.gulati@nyu.edu).

S. Farokh Atashzar is with the Department of Electrical and Computer Engineering, New York University (NYU), New York, NY 10003 USA, with the Department of Mechanical and Aerospace Engineering, New York University, New York, NY 10003 USA, and also with the NYU Center for Urban Science and Progress (CUSP), New York University Wireless, Brooklyn, NY 11201 USA (e-mail: f.atashzar@nyu.edu).

Digital Object Identifier 10.1109/LRA.2021.3091698

used in literature for this task. In this regard, to capture both short-term and long-term memories as required to handle time series of sEMG data, Long Short Term Memory (LSTM) is commonly used to deal with vanishing/exploding gradient problems [47], [48]. In addition, hybrid RNN-CNN models have been evaluated for sEMG-based HGR [49]–[51], taking advantage of both techniques. With all the diversity in terms of structures, the utilized deep-learning models typically take a large amount of training data to increase the number of classified gestures. However, while deep-learning models have the potential to bring high performance on large datasets, they are generally intense in terms of the training process due to great structural complexity. This would result in prolonging the training time, reducing the agility of the training process, affecting the convergence, and in some cases, overfitting. The mentioned problem also reduces the flexibility in evaluating various network architectures to find the most optimal design in a systematic manner. In addition, the higher the size of the network, the higher the possibility of overfitting, calling for larger datasets and data collection.

In this letter, we introduce the application of temporal dilation of LSTM[52] in a hybrid model to increase the long-term memory of LSTM and to make the model less data-hungry while increasing the "temporal reach," resulting in higher accuracy and much lower training time and iteration. In order to extend access to deep temporal dynamics without increasing complexity, dilated LSTM utilizes dilated recurrent cells that skip some temporal connections, see Fig. 3. This new feature extends the range for modeling temporal dependencies, alleviating conventional vanishing and exploding gradient problems. The temporal dilation makes the structure less complex at the same time stronger for modeling long dependencies.

Unlike conventional dilations of the kernels in CNN, which has been used in several domains, temporal dilation in RNN was suggested in 2017 [52] and thus have not been examined in a different application specifically for processing electrophysiological signals and in the area of neurorobotics. In this letter, for the first time, we have shown that temporal dilation can significantly reduce the processing time for training the neural network model for sEMG decoding in the task of human hand gesture prediction. We have shown that temporal dilation can address an unmet need to accelerate the calibration process for sEMG decoding, which allows for taking a major step towards using deep learning models in practice for myoelectric control. Using the proposed dilation, the model can reach older memories and deeper temporal structures and dependencies of the data without increasing the size of the LSTM, reducing the potential need for a very high number of trainable parameters. To the best of our knowledge, this is the first time that dilated RNN is proposed in the context of sEMG data classification. In addition, we propose and investigate the concept of "aggressiveness" of the temporal dilation and evaluate three degrees of aggressiveness. Finally, we perform the statistical tests to evaluate the importance and significance of the observations.

Fig. 1. Considered hand gestures from Ninapro DB2 set B.

II. Material and Methods

A. Data Acquisition

In this work, in order to benchmark the results with respect to the literature, we utilize the publicly available Ninapro dataset, which has been widely used to evaluates various methods for sEMG-based human-machine interfacing. We use the "exercise set B" from the second subset of the dataset (DB2) [35], which contains 17 hand movements shown in Fig 1. The data for DB2 was collected from 40 healthy adults, 12 females and 28 males with ages ranging from 29.9 ± 3.9 years old. The equipment used was Delsys Trigno Wireless sEMG System with 12 wireless electrodes, sampled at two kHz. The subjects were asked to perform different hand gestures for six repetitions. The subjects were told to hold the gesture for five seconds followed by a three-second rest. The sEMG signals were post-processed by relabeling to minimize the mistakes in auto-labeling. The relabeled dataset was named the Posterior dataset, whereas the original data is called the Prior dataset. In our experiment, we follow the recommendations provided by the dataset [35] and utilize repetitions 1, 3, 4, and 6 for training purposes and the remaining repetitions 2 and 5 for testing the trained model. This process results in having 544 seconds ($<$10 minutes) of data, required to train the system for a new subject.

B. Data Preprocessing

In this letter, the preprocessing includes (a) Z-score normalization based on training data only, (b) signal windowing using a sliding window of 300 ms (we also provide results for 200 ms and 100 ms) with a step size of 10 ms. Signal samples are considered to be the features. The model input is a two-dimensional space (600×12) where 600 represents the timestamps (corresponding to 300 ms at 2 kHz), and 12 represents the number of channels. Based on the literature, the recommended window length is 300 ms for real-time implementation in the context of myoelectric control. This topic has been investigated widely in the literature, and a comparative study is given in [34]. The window length of 300 ms has been adopted as a standard in the research community. Although an optimal window length for real-time gesture classification is still an open topic, the current literature follows the suggestion of window lengths $\leq$300 ms [53]–[56]. We also conduct a comparative study on the effect of reduction in window length, as given in Section II-B. It should be noted that the time taken by the model to predict a gesture ($\approx$2 ms) is negligible as compared to the window.

Remark: In this letter, we also evaluate the effect of smaller sliding window sizes of 100 ms and 200 ms on the performance

TABLE I
RESULTS FOR DIFFERENT SLIDING WINDOW SIZES

	100ms	**200ms**	**300ms**
4-layer 3rd Order Dilation	74.7%	79.0%	82.4%
4-layer Hybrid Baseline	72.5%	77.6%	80.1%
SVM	26.2%	26.9%	30.7%

of our best hybrid baseline and dilated models. Smaller windows could reduce the response time of a prosthesis system. However, we observed that the performance of the proposed system stays reasonably high even for smaller windows, up to 79%. To compare the performance of the system with classical approaches, we also implemented classical SVM operating on window sizes of 300 ms, 200 ms, and 100 ms. For the SVM, we extracted 192 features from each 300 ms sliding window, consisting of 48 temporal features and 144 frequency features. The 48 (4×12) time features include four moments (mean, variance, skewness, and kurtosis) from each channel, and the 144 ($4 \times 3 \times 12$) frequency features contain the same four moments from three conventional neural frequency ranges from each channel. The aforementioned frequency bands are 0.5–12 Hz, 12–35 Hz, and 35+ Hz. In the implemented SVM, the feature dimension is reduced using Principal Component Analysis (PCA) before training. As can be observed in Table I, SVM fails to decode complex neurophysiological features for the high number of gestures (i.e., 17). Also, it is observed that the proposed dilated model performs better than the conventional model in terms of accuracy, in addition to the speed of convergence which will be discussed later. •

C. Early Stopping and Model Checkpoint

In this letter early stopping criteria [48], [57], [58] is adopted to avoid overfitting and to skip unnecessary training iterations. For this, a patience factor of 30 was considered which makes sure that the model keeps on training for 30 more iterations even after it achieves the plateau of accuracy improvement. Similar to the literature if no improvement is observed after the duration of patience factor, the training stops and the prior best model was taken. This is used to compare the pattern of convergence of the proposed method with the conventional model.

III. MODEL STRUCTURE

Although deep-learning architectures for sEMG processing have shown high performance while leveraging large amounts of data, they are yet far from optimal in terms of both complexities of the structure and efficiency of the training process. In this letter, we propose the concept of temporal dilation with multi-degree aggressiveness, for sEMG classification, with the goal of improving accuracy, training speed, and reducing structural complexity. A faster training process can significantly help with practicality, especially since, in many cases, re-calibration is needed for neurorobotic systems and in general human-machine interfaces. Also, the computational power can be limited for outside of research lab uses of the prostheses. Thus it is always important to reduce the complexity and support a fast-rate training process.

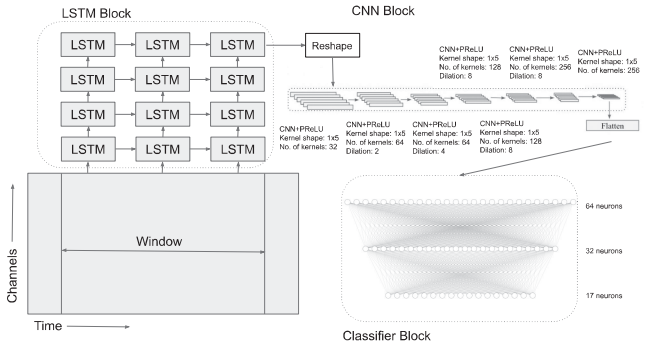

Fig. 2. Baseline hybrid model using a conventional LSTM.

It should be noted that the concept of dilation was initially implemented for CNN architectures [59], and recently it has been suggested for RNN [52]. In this work, we exploit the concept of temporal dilation and evaluate the effect of various degrees of dilation aggressiveness on the performance of the model. We also compare the performance of the proposed temporally-dilated method with two baseline models.

A. The Baseline Models for Comparison

For our baseline models, we utilize a hybrid architecture consisting of two modules, the LSTM module, and the CNN module [51]. For the LSTM module, we use four LSTM layers (for model 1) and three LSTM layers (for model 2), each with 600 LSTM cells and 128 hidden units. The final output of the LSTM module is given to a 1D CNN module, which includes seven blocks, each containing a Batch Normalization Layer and a Parametric Rectified Linear Unit (PReLU) layer following the Convolutional layers. In the end, we utilize a 3-layer fully-connected network, which subsequently contains 64, 32, and 17 cells, for encapsulating the information and conduction of the classification. Baseline model 1 is shown in Fig 2.

B. High-Order Aggressive Temporal Dilation

We propose three levels of aggressiveness for the LSTM dilation. We start with the first order dilation, defined as when we reduce the number of LSTM cells by half for each LSTM layer as compared to the previous layer. As a result, in our case, we have 600, 300, 150, and 75 LSTM cells for the first, second, third, and fourth layers, respectively. Fig. 3 visualizes the first order dilation. It is important to make sure that the remaining LSTM cells are all timewise aligned with the right margin such that the final output of the LSTM block is the latest cell and thus contains the most encapsulated memory. As an effort to further reduce the complexity and focus on deeper temporal dependencies in the data, higher-order dilation aggressiveness is evaluated in this letter. As a result, we utilize the second-order dilation (where we reduce the number of LSTM cells to one-forth of the previous layer) and a very aggressive third-order dilation (where each layer is reduced to one-eighth of the previous layer). In addition,

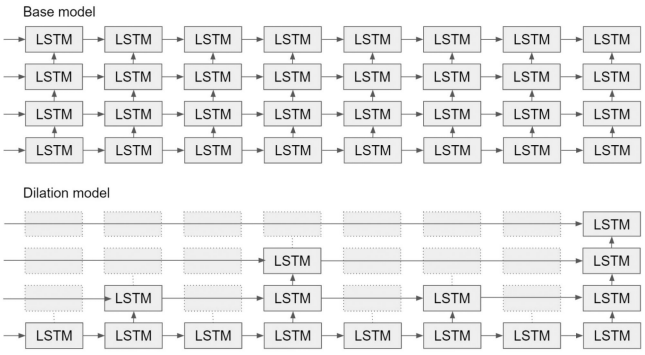

Fig. 3. Comparison of basic LSTM and first-order dilation models. Dilation reduces complexity of the model.

TABLE II
MODEL DESCRIPTIONS

ID	Name	Description
1	4-Layer, Baseline Model	Basic 4-layer LSTM model
2	4-Layer, 1st Order Dilation	4 layers, reduce to $\frac{1}{2}$ for each layer
3	4-Layer, 2nd Order Dilation	4 layers, reduce to $\frac{1}{4}$ for each layer
4	4-Layer, 3rd Order Dilation	4 layers, reduce to $\frac{1}{8}$ for each layer
5	3-Layer, Baseline Model	Basic 3-layer LSTM model
6	3-Layer, 2nd Order Dilation	3 layers, reduce to $\frac{1}{4}$ for each layer

to compare the effects of dilation on the various depth of the network, we conduct the dilation on 3-layer LSTM as well. A list of all models is given in Table II. In the next step, we evaluate the accuracy, the number of iterations for convergence, and the average training time for all models.

IV. EXPERIMENTS AND RESULTS

The results of statistical comparisons are shown in Fig. 4. To investigate the significance and importance of dilation aggressiveness, we perform statistical analysis on all the models across 40 subjects. The significant threshold for p-value is considered to be 0.05. We also applied Bonferroni correction to the observed p-values. Bonferroni is used in the literature [60] to reduce the probability of false positives and to prevent data from incorrectly appearing to be statistically significant, leading to more conservative test results. Accordingly for the results in Fig. 4, we used the following significance markers: (a) corrected p-values between 0.05 and 1 are considered to be not significant (ns); (b) corrected p-values between 0.01 and 0.05 are marked by $*$; (c) corrected p-values between 0.001 and 0.01 are marked by $**$; (d) corrected p-values between 0.0001 and 0.001 are marked by ***; and (e) corrected p-values smaller than 0.0001 are marked by ****. Table III shows the results of comparison. For statistical comparisons, we utilized paired tests since the exact same group of subjects was tested in different distributions. To compare the accuracy, we performed D'Agostino-Pearson test for normality and since it passes, we used the t-test for comparison. To compare the number of converged iterations and to compare the average

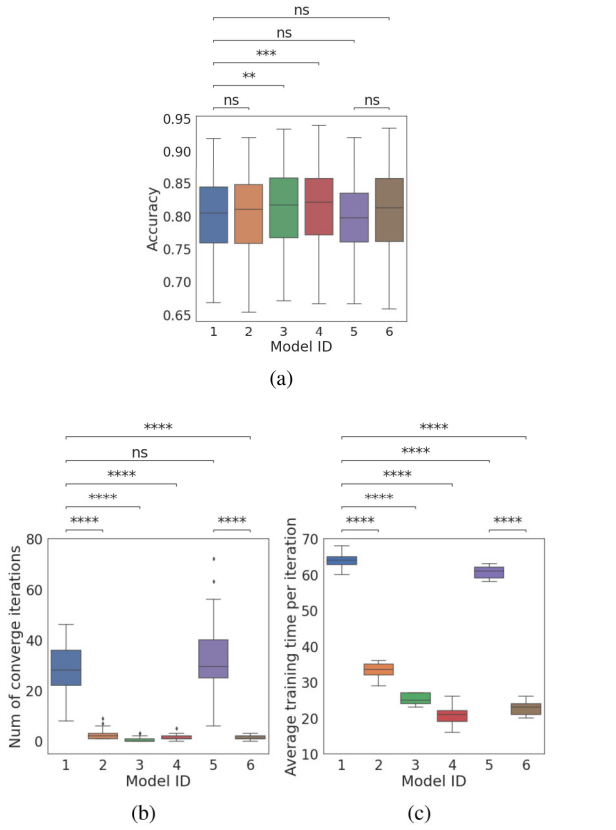

Fig. 4. (a) Accuracy boxplots for all models (x-axis is the ID from Table II and black horizontal line in IQR is the median); (b) Number of iterations to converge; (c) Average training time per iteration. N = 40 in all box plots.

TABLE III
STATISTICAL TESTS WITH BONFERRONI CORRECTION

Model Pairs	Accuracy	# Converged Iteration	Iteration Time
1 v.s. 2	Not significant	2.116e-07	1.960e-07
1 v.s. 3	0.0026	2.125e-07	1.991e-07
1 v.s. 4	0.00034	2.126e-07	1.963e-07
1 v.s. 5	Not significant	Not significant	5.137e-06
1 v.s. 6	Not significant	2.124e-07	1.936e-07
5 v.s. 6	Not significant	2.127e-07	2.010e-07

Note: model pairs are denoted using model IDs in Table II.

training time per iteration between the models, we utilized the Wilcoxon signed-rank test.

We also evaluate the model performance in terms of accuracy, sensitivity, specificity, number of iterations to converge, training time per iteration, number of trainable parameters of the six models (with and without dilation) using boxplots. In this work, sensitivity is the ratio between the true positive predictions of a gesture and the real number of samples of that gesture in the dataset. Specificity is the ratio between true negative predictions of a gesture and the real number of samples in the dataset that does not belong to the understudy gesture. Each boxplot presents the performance distribution across 40 subjects for each model. The median performance of each model is denoted by a black horizontal line in the middle of each interquartile range (IQR).

The length of each whisker was set to 1.5 times IQR which is the difference between the values at the 75th percentile and 25the percentile. Thus, each of the whiskers covers subject accuracies that are two standard deviations away from the IQR of each model. Fig. 4(a) shows the boxplots for the model accuracy of all subjects for different models (the x-axis is the model ID, from Table II). The details of statistical tests are listed in Table III.

When comparing the 4-layer hybrid baseline model with the third-degree dilation, we can observe an increase of performance from a median accuracy of 80.4%, median sensitivity of 87.74%, and median specificity of 90.32% for the 4-layer hybrid baseline model to a median accuracy of 82.0%, median sensitivity of 88.57%, and median specificity of 91.13% for the 4-layer third-order dilated model, showing a statistically significant difference (Bonferroni corrected p-value=0.00 034). More interestingly, although the shallower 3-layer hybrid baseline model has lower performance (median accuracy of 79.6%, median sensitivity of 85.08%, and median specificity of 90.88%) when compared with the 4-layer baseline, adding the second-order dilation to the 3-layer hybrid baseline model results in better performance (median accuracy of 81.2%, median sensitivity of 86.92%, and median specificity of 91.66%) when compared with the 4-layer baseline. This is despite the fact that the 4-layer hybrid baseline has a higher number of trainable parameters and layers, and in general, higher complexity. This highlights the importance of dilation.

Regarding training efficacy, we evaluate two factors, i.e., the number of iterations needed to reach 95% of final accuracy and the average time consumption of each iteration. Results are shown in Fig. 4(b) and 4(c). It could be clearly seen that dilation has a consistently outstanding ability to cut down the number of iterations (by a factor of 7), and the difference (when compared with the hybrid baseline model) is statistically significant. Among the proposed dilation methods, the second-order dilation in the 3-layer model, and the third-order dilation in the 4-layer model, provide the best performance in terms of the number of iterations.

For the average training time per iteration (which depends on the type of the processor), we take the average iteration time of each model training on all subjects under the same environment using Nvidia Quadro RTX 8000 GPU. As can be seen, the dilation has consistently resulted in a shorter average training time per iteration. In this regard, the third-order dilation for the 4-layer model has the minimum average training time (three times faster than the hybrid baseline). For the 3-layer model, significant reduction is also achieved when the 2nd-order dilation is applied.

As a result, it can be mentioned that the third-order dilation on the 4-layer model resulted in the best performance, the lowest number of iterations, and a lower average time per iteration. As an indicator, in Table IV we also calculate a converge time as the multiplication of the number of iterations to converge and time per iteration. The hybrid baseline model takes almost 1800 s while our best model takes around 84 s, indicating roughly a 20 times faster speed in converge time using our high-performance machine. This highlights the importance of the proposed temporal dilation, which gives higher access to older memories of the signal, allowing the model to better understand the underlying temporal dynamics of the time series, which directly relates to the underlying neurophysiology in the context of sEMG-based gesture classification for neurorobotic prostheses.

TABLE IV
RESULTS FOR MODEL PERFORMANCE AND AGILITY

Model ID	1	2	3	4	5	6
Median Acc	80.4%	80.9%	81.6%	**82.0%**	79.6%	81.2%
# Iterations to Converge	29	5	**3**	4	32	4
Time per Iteration	63s	40s	33s	**21s**	60s	30s
Converge Time	1827s	200s	99s	**84s**	1920s	120s
# parameters for LSTM	466,944	466,944	466,944	466,944	335,360	335,360

Note: Acc=Accuracy, #=Number of.

For better visualization, we also include the comparison of accuracies between the 4-layer baseline model and the third-order dilation for all subjects, separately, in Fig 5. This shows the behavior of the proposed method compared to the conventional method for 40 subjects.

It can also be seen in Fig. 5 that the proposed model has a simpler convergence pattern which highlights much lower complexity. For most of the subjects, the conventional method follows a nonlinear and unpredictable pattern of convergence. For example, in many iterations, the performance drops before it goes back up again (which is common in complex networks and increases the likelihood of suboptimal local minima). However, the proposed approach resulted in an almost-monotonically increasing performance which is a major need for re-calibration, in practice, when new data is added. This also helps to fine-tune the parameters and allows for a more practical optimization. The results highlight the significance of temporal dilation for the processing of sEMG data. This research is one step toward the ultimate implementation of deep-learning models for neurorobotic systems in the daily lives of the users, with the potential to transform the performance of human-machine interfaces.

V. COMPARATIVE STUDY

The goal of this comparative study is to specifically highlight the effect of dilation without the potential effect from the CNN module of the hybrid model. Thus we compare (a) pure 4-layer LSTM (without the CNN module) using various numbers of layers (from 1 to 4), (b) pure 4-layer dilated LSTM using various orders of aggressiveness for dilation (from 1 to 3), (c) classical SVM, (d) 2-layer multilayer perceptron (MLP), and (e) 2-layer CNN.

The tested MLP has 24, 17 filters in its first and second layer, respectively. The CNN model consists of two CNN blocks, each of which has one convolutional layer, one batch normalization, and one PReLU. A max-pooling layer with a kernel size of 2×2 was applied in between two convolutional blocks. The first convolutional layer has 16 filters with a kernel size of 15×5; the second convolutional layer has 24 filters, each having a kernel size of 15×5. For the SVM model, we extracted 192 time and frequency features from each 300 ms sliding window (containing 600 timestamps of sEMG signals). The number of

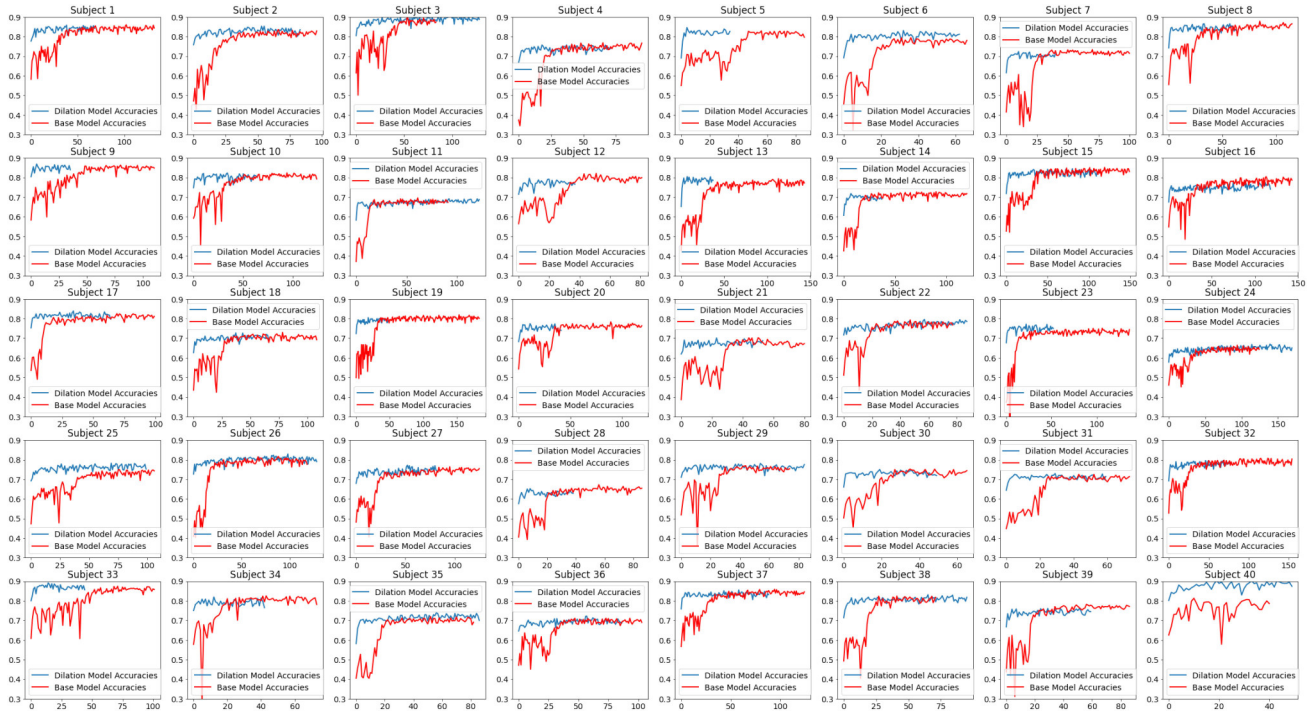

Fig. 5. Validation accuracy per iteration (blue line is the proposed and red line is the conventional technique).

TABLE V
RESULTS FOR COMPARING BASELINE LSTM AND DILATED LSTM MODELS WITH CONVENTIONAL MODELS

	Average Accuracy
4-layer 3rd Order Dilation (pure LSTM)	**79.7%**
4-layer 2nd Order Dilation (pure LSTM)	79.6%
4-layer 1st Order Dilation (pure LSTM)	79.3%
4-layer Baseline	75.3%
2-layer CNN	74.6%
2-layer LSTM	70.2%
1-layer LSTM	68.4%
2-layer MLP	66.2%
SVM	30.7%

hidden units on each LSTM layer is set to be 64. Here the baseline model is the pure 4-layer LSTM for comparison purposes. The results are summarized in Table V.

Observation 1: We observe that the performance of the shallower LSTM models (one-layer: 68.4%; two-layer: 70.2%) has deteriorated compared to 4-layer baseline given the same window size of 600 cells, as one or two LSTM layers are not enough to extract the complicated underlying neurophysiological features from the sEMG signals.

Observation 2: The CNN underperforms our baseline model by having an accuracy of 74.6% while having a similar size in terms of the number of trainable parameters.

Observation 3: The 2-layer MLP also returns a lower accuracy of 66.2%. The MLP was also structurally comparable (≈140 k trainable parameters) to the baseline model.

Observation 4: SVM, having an average accuracy of 30.7%, fails to decode complex neurophysiological features for the higher number of gestures (i.e., 17).

This comprehensive comparative study clearly shows the performance boost by adding dilation to LSTM layers as well as the superiority of the dilated LSTM models compared to all the commonly-used conventional deep learning models in this study. It is important to note that, when comparing time-series deep learning models (e.g., LSTM) with other modern or conventional neural networks, it can be mentioned that time-series models have shown a superior generalizability performance due to the capability of extracting underlying temporal dynamics, we have recently shown this feature [50]. Another advantage of LSTM models over other neural network models is that LSTM models can utilize information from the previous inputs to predict the upcoming inputs, significantly enhancing the time resolution. However, the bottlenecks for using time-series modeling is the very slow convergence speed due to the complexity of the model, which is targeted in this current paper using aggressive dilation. Additionally, it should be noted that the 4-layer pure LSTM dilated model achieves similar performance compared to the 4-layer hybrid model. The performance resemblance indicates that adding simple dilation (which reduces the complexity) has an equivalent positive effect on accuracy, as adding seven CNN blocks (which increases the complexity). Considering the above and the statistical results regarding 20 times faster convergence time, this letter illustrates that temporal dilation can open new avenues for translating advanced machine intelligence from research labs into the daily lives of the users.

VI. Conclusion

The performance of neurorobotic systems, including powered prosthetic technologies, significantly depends on the accuracy and agility of the human-machine interfaces. Deep-learning models have shown good potential for providing a high spatiotemporal resolution in the detection of human intention. However, due to the complexity, the training process of such systems has been a major bottleneck, especially since re-training and re-calibration are needed for practical uses. In this letter, for the first time, we propose the concept of high-order temporal dilation of LSTM for sEMG processing. We have shown that the proposed architecture not only can increase the performance but also reduces the need for having deeper networks while significantly dropping the number of iterations and the time for each iteration. In addition, the proposed technique results in a monotonic convergence pattern of the neural network, which would help better optimize the design for neurorobotic systems. Statistical validation supports the increased accuracy and efficacy of the proposed architecture. The current study uses offline analysis which can be considered as one limitation.

The paper follows the recommended threshold in the literature for the duration of the signal windowing, i.e., ≈300 ms, to predict the intention. Although our primary goal is not to reduce the window length, we have shown that the proposed model can keep the accuracy ≈80% even for a 200 ms window, while it significantly reduces the training time (by 20 times). Smaller sliding window sizes could further reduce the response time of a prosthetic system, enhancing agility and practicality. Optimizing the window length is one of our future lines of research. In this letter, we do not aim and claim to decode the transient phases of motor intention. We have recently investigated the possibility of transient phase decoding using decomposed high-density sEMG. As an ongoing line of research, we are investigating the potential use of dilated LSTM models for enhancing the performance of transient phase decoding.

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, 2008.

[2] S. F. Atashzar *et al.*, "Haptics-enabled interactive neurorehabilitation mechatronics: Classification, functionality, challenges and ongoing research," *Arch. Phys. Med. Rehabil.*, vol. 57, pp. 1–19, 2019.

[3] A. S. Gorgey, "Robotic exoskeletons: The current pros and cons," *World J. Orthop.*, vol. 9, pp. 112–119, 2018.

[4] M. Sanchez-Villamañan *et al.*, "Compliant lower limb exoskeletons: A comprehensive review on mechanical design principles," *J. Neuroeng. Rehabil.*, vol. 16, no. 1, p. 55, May 2019.

[5] H. Kazerooni, "A review of the exoskeleton and human augmentation technology," in *Proc. Dyn. Syst. Control Conf.*, vol. 43352, 2008, pp. 1539–1547.

[6] K. Tsiakas *et al.*, "A taxonomy in robot-assisted training: Current trends, needs and challenges," *Technologies*, vol. 6, no. 4, p. 119, Dec. 2018.

[7] A. Rajavenkatanarayanan *et al.*, "A survey of assistive technologies for assessment and rehabilitation of motor impairments in multiple sclerosis," *Multimodal Technol. Interaction*, vol. 3, no. 1, p. 6, Feb. 2019.

[8] D. Farina *et al.*, "Man/machine interface based on the discharge timings of spinal motor neurons after targeted muscle reinnervation," *Nature Biomed. Eng.*, vol. 1, no. 2, pp. 1–12, 2017.

[9] J. L. Betthauser *et al.*, "Stable responsive EMG sequence prediction and adaptive reinforcement with temporal convolutional networks," *IEEE Trans. Biomed. Eng.*, vol. 67, no. 6, pp. 1707–1717, Jun. 2019.

[10] M. Bretan *et al.*, "A robotic prosthesis for an amputee drummer," Dec. 2016, *arXiv:1612.04391*.

[11] A. Fougner, Ø. Stavdahl, P. J. Kyberd, Y. G. Losier, and P. A. Parker, "Control of upper limb prostheses: Terminology and proportional myoelectric control-a review," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 20, no. 5, pp. 663–677, Sep. 2012.

[12] S. Srinivasan *et al.*, "On prosthetic control: A regenerative agonist-antagonist myoneural interface," *Sci. Robot.*, vol. 2, no. 6, p. eaan2971, May 2017, doi: 10.1126/scirobotics.aan2971.

[13] M. K. Burns *et al.*, "Dynamic control of virtual hand grasp using spatiotemporal synergies," *IEEE Access*, vol. 7, pp. 112327–112338, 2019.

[14] J. T. Belter *et al.*, "Mechanical design and performance specifications of anthropomorphic prosthetic hands: A review," *J. Rehabil. Res. Dev.*, vol. 50, no. 5, pp. 599–618, 2013.

[15] "Össur. Life Without Limitations," Accessed: Jul. 01, 2021. [Online]. Available: https://www.ossur.com/enus/prosthetics/arms/i-limb-ultra

[16] *Bebionic Homepage*. [Online]. Available: https://www.ottobockus.com/campaign/bebionic.html

[17] "LUKE Arm detail page . mobius bionics," [Online]. Available: https://www.mobiusbionics.com/lukearm/

[18] J. Fras and K. Althoefer, "Soft biomimetic prosthetic hand: Design, manufacturing and preliminary examination," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2018, pp. 1–6.

[19] H. Zhao *et al.*, "Optoelectronically innervated soft prosthetic hand via stretchable optical waveguides," *Sci. Robot.*, vol. 1, no. 1, 2016, Art. no. eaai7529.

[20] R. P. Rocha *et al.*, "Fabrication and characterization of bending and pressure sensors for a soft prosthetic hand," *J. Micromechanics Microengineering*, vol. 28, no. 3, 2018, Art. no. 034001.

[21] P. Polygerinos *et al.*, "Soft robotic glove for combined assistance and at-home rehabilitation," *Robot. Auton. Syst.*, vol. 73, pp. 135–143, 2015.

[22] P. Polygerinos *et al.*, "Towards a soft pneumatic glove for hand rehabilitation," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2013, pp. 1512–1517.

[23] A. T. Asbeck, S. M. M. De Rossi, I. Galiana, Y. Ding, and C. J. Walsh, "Stronger, smarter, softer: Next-generation wearable robots," *IEEE Robot. Automat. Mag.*, vol. 21, no. 4, pp. 22–33, Dec. 2014.

[24] E. Biddiss and T. Chau, "The roles of predisposing characteristics, established need, and enabling resources on upper extremity prosthesis use and abandonment," *Disabil. Rehabil.: Assistive Technol.*, vol. 2, no. 2, pp. 71–84, 2007.

[25] K. Østlie *et al.*, "Prosthesis rejection in acquired major upper-limb amputees: A population-based survey," *Disabil. Rehabil.: Assistive Technol.*, vol. 7, no. 4, pp. 294–303, 2012.

[26] A. D. Roche *et al.*, "Prosthetic myoelectric control strategies: A clinical perspective," *Curr. Surg. Rep.*, vol. 2, no. 3, pp. 1–11, 2014.

[27] E. Biddiss *et al.*, "Consumer design priorities for upper limb prosthetics," *Disabil. Rehabil.: Assistive Technol.*, vol. 2, no. 6, pp. 346–357, 2007.

[28] M. A. Oskoei and H. Hu, "Myoelectric control systems-a survey," *Biomed. Signal Process. Control*, vol. 2, no. 4, pp. 275–294, 2007.

[29] W. Geng *et al.*, "Gesture recognition by instantaneous surface EMG images," *Sci. Rep.*, vol. 6, no. 1, pp. 1–8, 2016.

[30] X. Zhang and P. Zhou, "Sample entropy analysis of surface EMG for improved muscle activity onset detection against spurious background spikes," *J. Electromyogr. Kinesiol.*, vol. 22, no. 6, pp. 901–907, 2012.

[31] A. Phinyomark *et al.*, "Evaluation of EMG feature extraction for hand movement recognition based on euclidean distance and standard deviation," in *Proc. ECTI-CON2010: ECTI Int. Conf. Elect. Eng./Electron., Comput., Telecommun. Inf. Technol.*, 2010, pp. 856–860.

[32] A. Phinyomark *et al.*, "Feature reduction and selection for EMG signal classification," *Expert Syst. Appl.*, vol. 39, no. 8, pp. 7420–7431, 2012.

[33] A. Phinyomark *et al.*, "EMG feature evaluation for improving myoelectric pattern recognition robustness," *Expert Syst. Appl.*, vol. 40, no. 12, pp. 4832–4840, 2013.

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

[35] M. Atzori *et al.*, "Electromyography data for non-invasive naturally-controlled robotic hand prostheses," *Sci. Data*, vol. 1, no. 1, pp. 1–13, 2014.

[36] R. N. Khushaba and S. Kodagoda, "Electromyogram (EMG) feature reduction using mutual components analysis for multifunction prosthetic fingers control," in *Proc. 12th Int. Conf. Control Automat. Robot. Vis.*, 2012, pp. 1534–1539.

[37] U. C. Allard *et al.*, "A convolutional neural network for robotic arm guidance using sEMG based frequency-features," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2016, pp. 2464–2470.
[38] M. Atzori *et al.*, "Deep learning with convolutional neural networks applied to electromyography data: A resource for the classification of movements for prosthetic hands," *Front. Neurorobot.*, vol. 10, p. 9, Sep. 2016.
[39] Y. Du *et al.*, "Surface EMG-based inter-session gesture recognition enhanced by deep domain adaptation," *Sensors*, vol. 17, no. 3, p. 458, Feb. 2017.
[40] S. Tam, M. Boukadoum, A. Campeau-Lecours, and B. Gosselin, "A fully embedded adaptive real-time hand gesture classifier leveraging HD-sEMG and deep learning," *IEEE Trans. Biomed. Circuits Syst.*, vol. 14, no. 2, pp. 232–243, Apr. 2020.
[41] X. F. Zhou *et al.*, "Gesture recognition with EMG signals based on ensemble RNN," *Guangxue Jingmi Gongcheng/Opt. Precis. Eng.*, vol. 28, no. 2, pp. 424–442, 2020.
[42] A. Samadani, "Gated recurrent neural networks for EMG-based hand gesture classification. a comparative study," in *Proc. 40th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, 2018, pp. 1–4.
[43] M. Hioki and H. Kawasaki, "Estimation of finger joint angles from sEMG using a neural network including time delay factor and recurrent structure," *Int. Scholarly Res. Notices*, vol. 2012, 2012. [Online]. Available: https://downloads.hindawi.com/archive/2012/604314.pdf
[44] W. Wei *et al.*, "A multi-stream convolutional neural network for sEMG-based gesture recognition in muscle-computer interface," *Pattern Recognit. Lett.*, vol. 119, pp. 131–138, 2019.
[45] T. Triwiyanto, I. P. A. Pawana, and M. H. Purnomo, "An improved performance of deep learning based on convolution neural network to classify the hand motion by evaluating hyper parameter," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 28, no. 7, pp. 1678–1688, Jul. 2020.
[46] U. Côté-Allard *et al.*, "Deep learning for electromyographic hand gesture signal classification using transfer learning," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 27, no. 4, pp. 760–771, Apr. 2019.
[47] M. Jabbari *et al.*, "Emg-based hand gesture classification with long short-term memory deep recurrent neural networks," in *Proc. 42nd Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, 2020, pp. 3302–3305.
[48] F. Quivira *et al.*, "Translating sEMG signals to continuous hand poses using recurrent neural networks," in *Proc. IEEE EMBS Int. Conf. Biomed. Health Informat.*, 2018, pp. 166–169.
[49] Y. Hu *et al.*, "A novel attention-based hybrid CNN-RNN architecture for sEMG-based gesture recognition," *PLoS one*, vol. 13, no. 10, 2018, Art. no. e0206049.
[50] P. Gulati, Q. Hu, and S. F. Atashzar, "Toward deep generalization of peripheral EMG-based human-robot interfacing: A hybrid explainable solution for neurorobotic systems," *IEEE Robot. Automat. Lett.*, vol. 6, no. 2, pp. 2650–2657, Apr. 2021.
[51] E. Rahimian *et al.*, "Semg-based hand gesture recognition via dilated convolutional neural networks," in *Proc. IEEE Glob. Conf. Signal Inf. Process.*, 2019, pp. 1–5.
[52] S. Chang *et al.*, "Dilated recurrent neural networks," Oct. 2017, *arXiv:1710.02224*.
[53] S. J. A.-L. Hassan, F. Hussein and I. K. Ibraheem, "Teleoperated robotic ARM movement using electromyography signal with wearable myo armband," *J. King Saud Univ.-Eng. Sci.*, vol. 32, no. 6, pp. 378–387, 2020.
[54] Z.-J. Lyu, S.-T. Tang, T.-L. Chia, W.T. Shi, and C.-Y. Yang, "A bionic hand controlled by hand gesture recognition based on surface EMG signals: A preliminary study," *Biocybernetics Biomed. Eng.*, vol. 38, no. 1, pp. 126–135, 2019.
[55] A. Aranceta-Garza and B. A. Conway, "Differentiating variations in thumb position from recordings of the surface electromyogram in adults performing static grips, a proof of concept study," *Front. Bioeng. Biotechnol.*, vol. 7, May 2019, doi: 10.3389/fbioe.2019.00123.
[56] M. F. Wahid, R. Tafreshi, and R. Langari, "A multi-window majority voting strategy to improve hand gesture recognition accuracies using electromyography signal," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 28, no. 2, pp. 427–436, Feb. 2020.
[57] C. Maufroy and D. Bargmann, "Cnn-based detection and classification of grasps relevant for worker support scenarios using sEMG signals of forearm muscles," in *Proc. IEEE Int. Conf. Syst., Man, Cybern.*, 2018, pp. 141–146.
[58] U. Côté-Allard *et al.*, "Transfer learning for sEMG hand gestures recognition using convolutional neural networks," in *Proc. IEEE Int. Conf. Syst., Man, Cybern.*, 2017, pp. 1663–1668.
[59] A. van den Oord *et al.*, "Wavenet: A generative model for raw audio," Sep. 2016, *arXiv:1609.03499*.
[60] M. Jafari and N. Ansari-Pour, "Why, when and how to adjust your p values?" *Cell J.*, vol. 20, no. 4, pp. 604–607, Jan. 2019.