Towards Efficient Neural Decoder for Dexterous Finger Force Predictions

Jiahao Fan and Xiaogang Hu

***Abstract*—*Objective*: Dexterous control of robot hands requires a robust neural-machine interface capable of accurately decoding multiple finger movements. Existing studies primarily focus on single-finger movement or rely heavily on multi-finger data for decoder training, which requires large datasets and high computation demand. In this study, we investigated the feasibility of using limited single-finger surface electromyogram (sEMG) data to train a neural decoder capable of predicting the forces of unseen multi-finger combinations. *Methods*: We developed a deep forest-based neural decoder to concurrently predict the extension and flexion forces of three fingers (index, middle, and ring-pinky). We trained the model using varying amounts of high-density EMG data in a limited condition (i.e., single-finger data). *Results*: We showed that the deep forest decoder could achieve consistently commendable performance with 7.0% of force prediction errors and R^2 value of 0.874, significantly surpassing the conventional EMG amplitude method and convolutional neural network approach. However, the deep forest decoder accuracy degraded when a smaller amount of data was used for training and when the testing data became noisy. *Conclusion*: The deep forest decoder shows accurate performance in multi-finger force prediction tasks. The efficiency aspect of the deep forest lies in the short training time and small volume of training data, which are two critical factors in current neural decoding applications. *Significance*: This study offers insights into efficient and accurate neural decoder training for advanced robotic hand control, which has the potential for real-life applications during human-machine interactions.**

***Index Terms*—Hand function, finger force prediction, deep forest, convolutional neural network.**

I. Introduction

OUR ability to independently control individual fingers empowers us to accomplish intricate tasks, spanning from simple grasping activities to complex manipulation tasks. In an effort to restore impaired or lost hand function, the development of assistive devices, such as prosthetic hands and exoskeleton gloves, has progressed significantly. Although these advanced devices can mimic the complex movements of a biological human hand [1]–[4], decoding user's motion intention into executable control commands remains a substantial challenge.

Surface electromyogram (EMG) signals are commonly used as the source of neural control for robotic hands. While EMG-based pattern recognition has attained prominent performance in identifying a finite set of intended movements (e.g. hand gesture recognition) [5]–[7], a major limitation of the pattern recognition is the lack of proportional activation of the recognized classes [8]. A second approach, termed proportional control, enables users to control individual actuators in a continuous manner, which has more potential to enable dexterous control of robotic hands [9].

Most previous studies on proportional finger control have primarily focused on predicting the motor output of an individual finger, which may have limited application as many daily tasks require dexterous control of multiple fingers. Relatively few studies have explored neural decoders that can predict forces of multiple fingers concurrently [10]–[13]. When leveraging high-density EMG (HD-EMG), it has been demonstrated that both the global (i.e., EMG amplitude) [14], [15] and microscopic (i.e., neural drive information) HD-EMG features can predict forces at individual finger levels. The global-feature-based method directly maps EMG amplitude to forces using simple regression models. However, global EMG amplitude features are susceptible to interference such as signal crosstalk and amplitude drift, leading to suboptimal performance at the individual finger level. In contrast, neural drive information utilizes motor unit firing activities obtained through motor unit decomposition. This information more accurately reflects the motor commands, thus providing accurate decoding of finger motor intent [11], [12], [16]. However, extracting neural drive information via motor unit decomposition was computationally intensive, limiting its practicality in real-life implementations.

Recent studies [17]–[19] have shown that individual finger flexion or extension can be characterized by localized spatial activation patterns in the HD-EMG global features, suggesting the presence of finger-specific mapping between EMG channels and their corresponding motor outputs. However, extrapolating these finger-specific channel mappings identified in single-finger tasks to untrained multi-finger tasks have proven to be unsatisfactory [11], [12]. To achieve optimal multi-finger force prediction, EMG from multi-finger movements must be included in the training data to provide information about the intricate co-activation patterns between fingers. However, given the multitude of possible finger combinations, it is often inefficient and sometimes impractical to exhaustively acquire data for all finger combinations to train the neural decoders. Therefore, it is invaluable to develop a method that can be trained on a limited dataset, such as single-finger data, and subsequently extrapolate to movements outside the training set (e.g., combinations of fingers). Such an approach can substantially reduce the time required for system training. However, implementing such a method may

Jiahao Fan and Xiaogang Hu are with the Department of Mechanical Engineering, Pennsylvania State University, University Park, 16802, PA, U.S.
*Corresponding author: Xiaogang Hu (xxh120@psu.edu)

This study was supported in part by the National Science Foundation (CBET-2246162, IIS-2330862, IIS-2319139) and the Department of Defense (W81XWH2110185).

be challenging, as the interdependencies among fingers [15], [16] may not be as pronounced in EMG during single-finger movements as in the movements of the unseen multiple-finger combinations.

Deep neural network models have demonstrated proficiency in mapping complex non-linear relations [20], [21]. For instance, an earlier work [22] utilized a deep convolutional neural network (CNN) that combined time and frequency domain features for wrist force prediction. Similarly, CNN demonstrated the effectiveness of predicting multi-directional wrist torques [23]. Others also employed a multi-task learning framework [24], [25], incorporating a deep learning approach to predict both myoelectric patterns and the resultant force. An earlier study [26] developed a more complex transformer structure to predict the joint kinematics of the hands.

However, there are several limitations of neural network models, First, these models often require a large volume of training data, which may not be feasible. Second, neural networks typically require extensive hyperparameter tuning to achieve optimal performance. Third, the optimization of neural network weights through iterative backpropagation is also notably time-consuming. From the perspective of efficiency, here we introduced a deep forest (DF) [27] based neural decoder to concurrently predict the force exerted by multiple fingers. DF leverages hierarchical cascade layers of decision trees to achieve deep representation learning, which enables it to capture the finger-specific activation patterns from HD-EMG for force prediction. In addition, compared to neural networks, DF has minimal hyper-parameters to tune, and the model structure could be adapted in a data-driven manner, thus being considered more efficient.

We used the DF-based decoder to simultaneously and accurately predict the flexion and extension forces of multiple fingers, using a global HD-EMG feature. We restricted our training data to single-finger data and then evaluated the model on untrained multi-finger combinations. Assessments on the force prediction of three fingers (index, middle, and ring-pinky) across seven participants revealed that the DF decoder consistently exhibited high accuracy, even with limited training data, in comparison to both the traditional EMG amplitude method and convolutional neural network. These findings underscore the potential of the DF decoder as an efficient tool for EMG-based neural decoding.

II. METHODS

A. Experimental Setup

Seven neurologically intact participants were recruited in the study. All subjects gave informed consent with protocols approved by the Institutional Review Board of the Pennsylvania State University.

The experiments included performing isometric finger flexion and extension tasks using the index, middle, pinky, and ring fingers. As shown in Fig. 1, two 8×16 electrode arrays were replaced on the anterior and posterior sides of the forearm of the subjects to record EMG signal from the flexor digitorum superficialis (FDS) and extensor digitorum communis (EDC) muscles. The single-electrode diameter and the inter-electrode distance of the electrode arrays are 3-mm and 10-mm, respectively. The reference channel was placed at the wrist. The HD EMG was acquired using an EMG-USB2+ system (OT Bioelettronica) with a gain of 1000 and a sampling rate of 2048 Hz. To measure the force during finger flexion and extension, we used miniature load cells (SM-200N, Interface) for the index, middle, ring, and pinky fingers individually. The sampling rate for recording the force data was set at 1000 Hz. During the experiments, the forearm was positioned in a neutral position and supported by two stiff foam pads. Prior to each trial, the offsets of the individual load cells were carefully adjusted, to ensure a positive force reading represented flexion and a negative reading represented extension.

B. Experiment Procedure

For each participant, the first step involved measuring the maximum voluntary contraction (MVC) of the flexion and extension force exerted by each finger. The flexion and extension MVC were measured separately for each finger. These MVC values were respectively used for flexion and extension force normalizations. A positive force indicates the flexion force, while a negative force denotes the extension force. In particular, the subjects were asked to perform finger flexion and extension at MVC following the force trajectory displayed on the monitor. Given the significant enslavement between the ring and pinky fingers [28], the participant was requested to extend and flex these two fingers simultaneously. These two fingers were treated as one finger (referred to hereafter as the 'ring'). Consequently, the force exerted by the ring-pinky finger was computed as the summation of the ring and pinky forces, which were then displayed to the participants.

During the experiment, the subjects followed a predefined force target (Fig.1 (B) and Fig.1 (C)), ranging from 0 to 50% MVC for each finger. The 50% MVC was selected to alleviate muscle fatigue that might occur at higher force levels. Two types of trials were performed by the participants. The first type was the single-finger trial, where the participants were instructed to flex or extend an individual finger according to the predefined single trapezoid for a duration of 21 seconds(Fig. 1 (B)). During the single-finger trial, the subjects were asked to avoid co-contraction with the unintended fingers. The second type was the multi-finger trial, where subjects were asked to flex and extend a minimum of two fingers, following a sequence of trapezoids for a duration of 27 seconds (two fingers involved) or 42 seconds (three fingers involved) (Fig. 1 (C)). For a specific duration, one finger was instructed to maintain the intended force level, while co-activation was allowed for other fingers. The selection of target fingers was randomized across the multi-finger trials. Overall, 12 single-finger trials (3 fingers $\times$ 4 trials) and 16 multi-finger trials were performed by each subject.

This article has been accepted for publication in IEEE Transactions on Biomedical Engineering. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TBME.2024.3353145

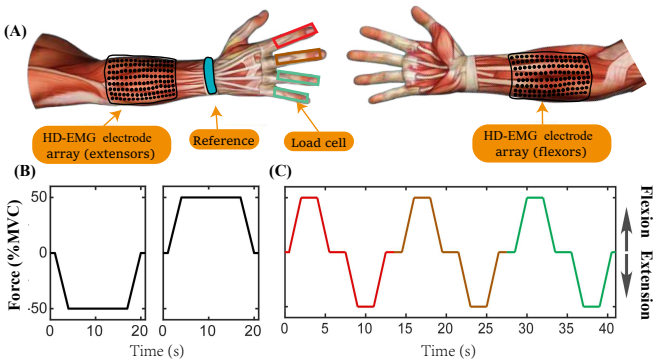

Fig. 1: The experiment setup and protocol. (A): Monopolar EMG signals were recorded from the finger extensor and flexor, respectively, with two 8×16 electrode arrays, and the flexion/extension forces of the index, middle, ring, and pinky fingers were recorded. The trapezoidal force target from the single-finger extension and flexion trial (B) and the multi-finger trial (C). The force target of the multi-finger trial was shown with different colors to represent the three fingers, i.e. index (red), middle (brown), and ring (green) as the instructed finger, respectively.

C. Data Preprocess

The data analysis was performed using MATLAB (The MathWorks, Inc.), scikit-learn[1], and Pytorch[2], running on a computer equipped with an Intel i7-12700k CPU and an RTX 3070 Ti GPU. The pre-processing of the recorded EMG and force data involved several steps. Initially, the raw EMG signals were filtered using a high-pass filter (Butterworth zero-phase shift with an order of 4). Next, the Root Mean Square (RMS) was computed for each channel utilizing a smoothing window of 0.5 seconds and a moving step of 0.1 seconds. The extracted RMS served as the input feature for the subsequent computations. For deep forest, the dimension of the input was 256 (2 electrode arrays $\times$ 128 channels). For CNN, we reshaped the input to a spatial layout of $2\times8\times16$. The recorded force was separately normalized by the MVC values for both flexion and extension. The normalized force was smoothed using the same window size and moving step.

In the supplementary material, we have provided example data for both single-finger and multi-finger trials, as illustrated in Fig. S1 and Fig. S2. The rationale behind using single-finger trials to predict multi-finger trials is based on the observation that the localized activation patterns in both single-finger and multi-finger movements exhibit overlap when the subject is instructed to perform the same finger-force task. Although the activation of an intended finger in a multi-finger movement can be influenced by the adjacent fingers due to co-activation, it is still expected to exhibit similar activation patterns as observed in the corresponding single-finger trials. Our goal is to capture these localized activation patterns during single-finger force production and utilize them to accurately predict multi-finger forces, even in scenarios where multiple fingers are simultaneously activated.

[1]https://scikit-learn.org/stable/
[2]https://pytorch.org/

D. Deep Forest

As illustrated in Fig. 2, each finger corresponded to a Deep Forest (DF), with each layer consisting of two random forests (RF) and two completely random forests (C-RF). This finger-specific model design offers scalability, because when a new finger needs to be incorporated, there's no need to retrain the models. In practice, these finger-specific models could run in parallel to ensure real-time processing during inference. In contrast to deep neural networks (DNNs) that necessitate determining the complexity of the network prior to training, the complexity of DF (e.g. the layer number) was determined in a data-driven manner and was automatically adjusted based on the information in the data. Namely, the addition of more layers is interrupted if there is no performance improvement.

During training, both RF and C-RF were fitted on the input data by selecting features that lead to the highest impurity reduction at each node within their respective decision trees. However, C-RF introduced an additional degree of randomness by not just randomly selecting the feature subsets in the splitting, but also randomly assigning split points within these features. This combination of different types of random forests in each estimator encourages diversity, which could potentially improve the model's overall performance. The estimation output of a forest layer was concatenated with the original input feature ($x \in \mathbb{R}^{256}$ (2 electrode arrays $\times$ 128 channels)), forming an augmented feature to be fed into the next layer. This simple feature augmentation promotes effective in-model feature transformation.

Though DF shares a similar concept of layer-by-layer processing with DNNs, it fundamentally differs in its essence as a non-parametric, tree-based ensemble method. Unlike DNNs, DF doesn't require the differentiation of layers or back-propagation for training. As previously mentioned, the number of layers is dynamically added until optimal performance is reached, ensuring a structured, step-wise training procedure for efficient learning and optimal model performance. It was worth noting that we did not undertake any hyper-parameter searching for the DF. The only hyper-parameters explicitly assigned were the number of RF and C-RF ($n = 2$) in each layer, with each forest comprising 100 trees. Beyond that, all the hyper-parameters were used in their default settings as described earlier [27].

E. CNN

CNN has emerged as an effective approach for EMG-force prediction [21], [22]. We implemented a two-branch Convolutional Neural Network (CNN) architecture (Fig. 3) to specifically analyze the activation patterns of the flexor and extensor muscles. Given the relatively small size of our dataset, we used a lightweight CNN architecture to prevent overfitting while ensuring efficient learning. Since the employed hyper-parameters may significantly impact the model's performance, we have tested various combinations of layer numbers $N \in \{1, 2, 3, 4\}$ and filter sizes $f \in \{3, 5, 7\}$. We found that $N = 2$

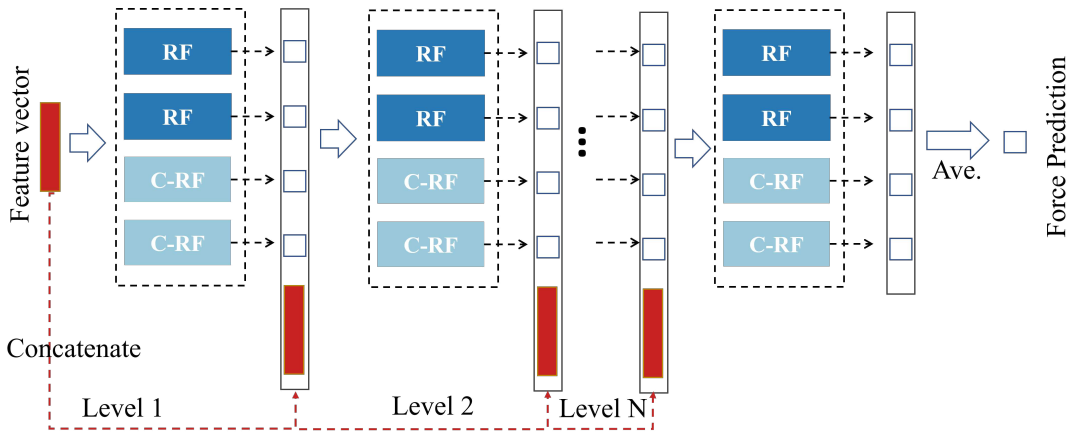

Fig. 2: The cascade structure of a deep forest. RF:random forest; C-RF: completely random forest; Ave.: Average operation

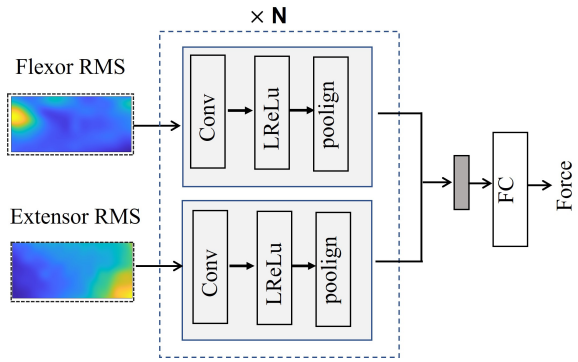

Fig. 3: The two-branch CNN model. Conv: 2d Convolutional Layer; LReLU, Leakey Rectified Linear Unit; Pooling: Max Pooing. FC: fully connected layer. The feature vectors extracted from the flexor and extensor EMG signals were fused to make a final prediction. The gray block represents the feature fusion operation.

and $f = 3$ yielded the best results on the validation dataset. Consequently, these settings were employed throughout our data analysis. The details of the utilized CNN can be found in Section 2 of the supplementary material.

Our initial strategy was to train a finger-specific model using the available single-finger data. However, due to data scarcity, the performance of the finger-specific CNN was notably inferior in comparison with the concurrent prediction models. As a remedy, we trained the CNN model capable of concurrent prediction of three-finger forces. This model, featuring a three-dimensional output space, was able to efficiently capture and represent the complex dynamics of multiple finger outputs.

The CNN was trained employing the Mean Squared Error (MSE) loss function along with $\mathcal{L}_2$ regularization, with a weight of 10^{-4}. The training process was set to a maximum of 300 iterations, using a learning rate of 0.001. To monitor the training progress and perform early stopping, 20% of the training data was randomly selected for validation after each training iteration. If the validation loss did not decrease for 10 consecutive iterations, the training process was stopped. The model that yielded the lowest validation loss was selected as the final model for testing and evaluation.

F. EMG-amplitude Method

The conventional EMG-amplitude (referred to as AMP hereafter) method was also applied as a benchmark for comparison. In order to enhance its performance in differentiating individual fingers, steps were taken to refine the channel selection process. This refinement is based on the understanding that not all channels are necessarily tied to the activation of a specific finger. Thus, only the most informative channels were chosen, restricted to the top 60 channels with the highest amplitude for the extensor or flexor muscle of each finger. This channel selection represents approximately 50% of the total channels of the extensor or flexor. The force prediction errors between EMG amplitudes using all channels and the top 60 channels are presented in Fig. S3 of the Supplementary Materials. If the multi-finger trial was available (only in sectionIII-A), a further refinement procedure was undertaken to select the most informative EMG channels from the top 60 channels for individual fingers [11], [13]. For a given finger, the EMG amplitude of the top 60 channels and the forces were calculated using a 0.5-second average window. Subsequently, a linear regression analysis was performed between all the 60 EMG channels and the forces recorded in multi-finger trials. The corresponding R^2 values were obtained and averaged across all multi-finger trials. An EMG channel was retained if the R^2 value between the EMG amplitude of the channel and the force of the corresponding finger was larger than that of other fingers. Otherwise, the channel was removed from the channel pool. After the channel refinement, a linear regression model was computed as follows:

$$F_i = aA_{i,\text{flx}} + bA_{i,\text{ext}} + c_i \tag{1}$$

where F_i represents the force of the $i-th$ finger, $A_{i,\text{flx}}$ and $A_{i,\text{ext}}$ denote the sum of the EMG amplitudes for the selected flexor and extensor channels related to the $i-th$ finger, respectively. The parameters a and b are coefficients that are learned through the regression, and c_i is the intercept.

This article has been accepted for publication in IEEE Transactions on Biomedical Engineering. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TBME.2024.3353145

G. Validation Protocol

Firstly, we evaluated the performance of the three methods (Deep Forest, EMG-amplitude, and CNN) in predicting multi-finger force when multi-finger trials were included during the training. For each subject, we used all the single-finger data and $3/4$ of the multi-finger trials for training, while keeping the hold-out multi-finger trials for testing. This process was repeated four times using a 4-fold cross-validation approach over the testing multi-finger trials.

In the main experiments, we focused on training the model using only single-finger trials and testing it on all the multi-finger trials. To ensure reproducibility, we repeated this process five times to account for any potential randomness or variation in the results.

We further investigated the robustness of the proposed methods under conditions of reduced training data and varied levels of noise. In particular, we evaluated model performance when the volume of training data was reduced to $\{75\%, 50\%, 25\%\}$ of its original amount. In addition, the resistance to noise is also examined, while the quality of training data can be controlled during the data acquisition phase, the quality of real-world testing data may be influenced by various uncontrollable factors such as ambient noise, movement artifacts, or changes in electrode-skin contact over time. Therefore, it is vital to assess the performance of the trained models under different noise conditions in the testing phase. Hence we have evaluated the robustness of the proposed deep forest against various levels of noise. We introduced Gaussian Noise into our testing data at varying signal-to-noise ratios (SNRs) to emulate real-world scenarios where the signal might be compromised. The levels of noise were defined by the SNRs in {5, 10, 15, 20}. For each SNR, the Gaussian noise was added to each channel of the original EMG, and then fed into the trained models for evaluation.

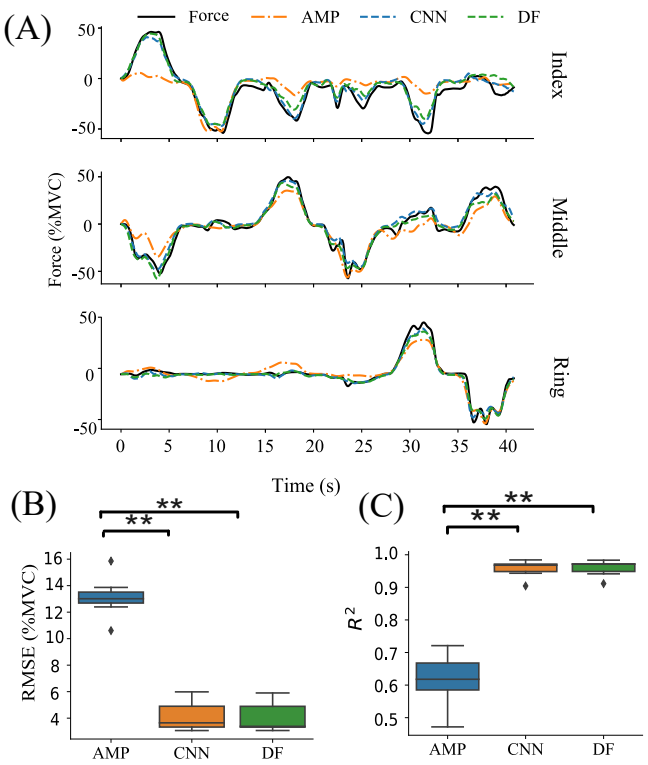

Fig. 4: Decoding accuracy of different approaches when single-finger and multi-finger data were used for training. The error bar represents standard error and the solid diamond denotes outliers (A) An example trial illustrating the estimated force relative to the recorded force using different methods (Deep Forest, EMG-amplitude, and CNN). (B) Root Mean Square Error (RMSE) and (C) R^2 were obtained by three methods using data that included multi-finger trials under 4-fold validation. *:$p < 0.05$. **: $p < 0.01$

H. Performance Metrics

Root mean square error (RMSE) and coefficient of determination (R^2) were used to evaluate the force estimation performance of the different decoding methods. To quantify the finger separation, the false active rate and false active rate were computed based on different levels of the threshold. Specifically, a finger was classified as 'active' if the recorded force exceeded a particular MVC threshold (set at $\{5\%, 10\%, or 15\%\}$) during a specific timeframe. Conversely, if the force fell below this threshold, the finger was deemed to be in a 'rest' state. The false rest rate refers to the percentage of 'rest' samples classified as 'active,' whereas the false active rate corresponds to the proportion of 'active' samples mistakenly identified as 'rest'. The reported values are computed by pooling all the obtained subject-wise results. Shapiro–Wilk test was first applied to verify the normality of the obtained metrics. Results showed that the Gaussian distribution assumption was satisfied. Therefore, the repeated measures analysis of variance (ANOVA) was used to evaluate the effects of specific factors on the obtained performance metrics. Paired t-tests with Bonferroni-Holm correction were then applied for multiple comparisons if necessary. The significance level was set as $p < 0.05$ in this study.

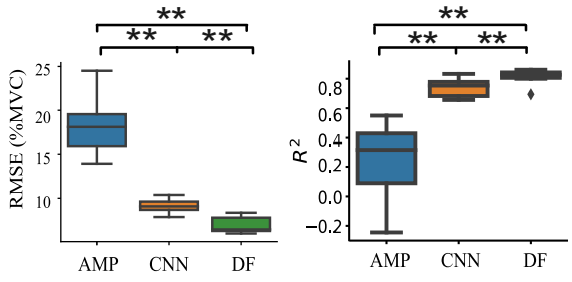

Fig. 5: Decoding accuracy of different approaches when only single-finger data were used for training. The error bar represents standard error and the solid diamond denotes outliers*:$p < 0.05$. **: $p < 0.01$

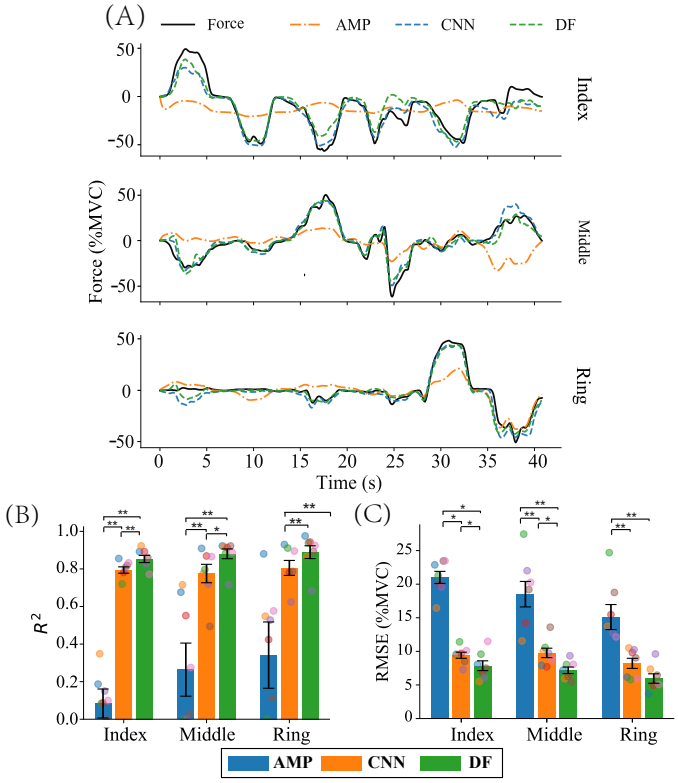

Fig. 6: The average force prediction results across subjects and trials. (A) An exemplar trial of multi-finger extension and flexion, (B) R^2 and (C)RMSE; *:$p < 0.05$, **: $p < 0.01$

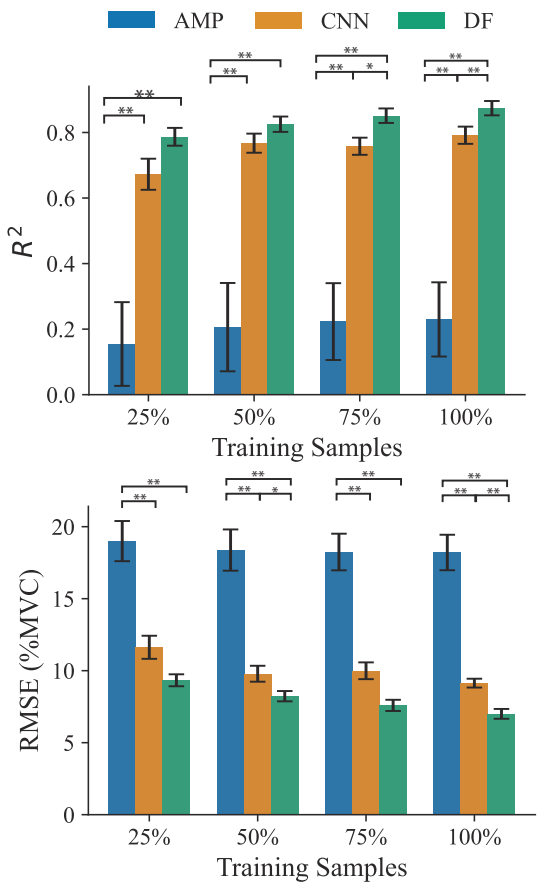

Fig. 8: The averaged force prediction performance v.s. the utilized training data. *:$p < 0.05$, **: $p < 0.01$.

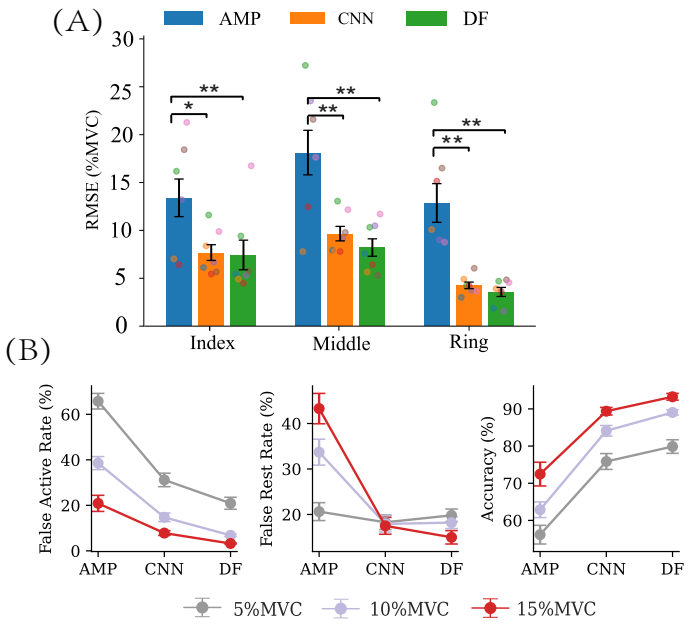

Fig. 7: (A) The force estimation error of unintended fingers. (B) The accuracy of active v.s. rest classification. *:$p < 0.05$, **: $p < 0.01$

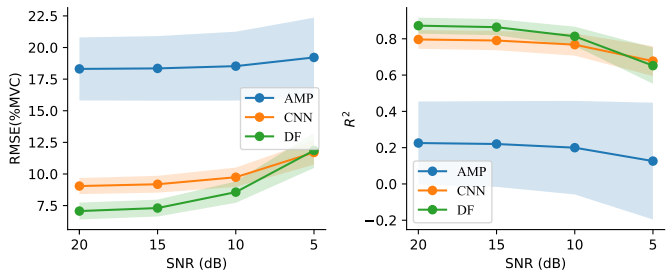

Fig. 9: The performance of force estimation v.s. the signal-to-noise ratio of the testing data

III. Results

A. Multi-finger Force Prediction Using Both single-finger and Multi-finger Data

Fig. 4 (B) and (C) display the results of multi-finger force prediction using three methods when multi-finger data was involved during training. The CNN and DF methods exhibit exceptional performance, achieving low RMSE of 4.16 ± 0.5 and 4.11 ± 0.5 (mean $\pm$ standard error, %MVC) respectively, with corresponding R^2 values of 0.957 ± 0.01 and 0.58 ± 0.01. In comparison, the EMG-amplitude method yields an RMSE value of 13.12 ± 1.2 and an R^2 value of 0.617 ± 0.06, which is consistent with previous studies [13]. A one-way repeated measures ANOVA indicated significant differences in both RMSE ($F(2, 12) = 111.9$, $p < 0.01$) and R^2 ($F(2, 12) = 97.0$, $p < 0.01$). Subsequent *post-hoc* analysis demonstrated that both CNN and DF methods yielded significantly lower RMSE and higher R^2 compared to AMP methods. However,

the performance between the DF and CNN methods did not exhibit any significant differences.

Fig. 4 (C) shows that both the CNN and DF methods can fit the recorded force almost perfectly, while the EMG amplitude method also demonstrates a good fit, especially for the ring finger. However, we observed substantial over- or under-fit in the index and middle fingers. These results provide initial validation of the feasibility of these methods for concurrent prediction of flexion and extension forces of three fingers.

B. Multi-finger Force Prediction Using Single-finger data

When the multi-finger data was excluded from the training data, the force estimation accuracy of all the methods decreased substantially (Fig.5). However, DF method still maintains commendable accuracy with an average R^2 of 0.874, while the AMP method showed a much lower mean R^2 value of 0.182. This suggests that the AMP method is not able to effectively establish the mapping between EMG signals and force outputs when the regression function was trained solely on single-finger data. On the other hand, the DF method demonstrates its robustness and ability to generalize in the absence of multi-finger data. We observed that the DF method exhibits superior performance by achieving the lowest average RMSE and the highest R^2 value among the tested methods. A one-way repeated measures ANOVA showed a significant difference in RMSE ($F(2,12) = 52.71, p < 0.01$) and R^2 ($F(2,12) = 22.62, p < 0.01$)) value among methods. Further *post-hoc* comparison showed that the difference in performance between the DF method and the other two methods was statistically significant ($p < 0.05$).

Fig.6 offers a detailed examination of the force estimation performance for individual fingers. For the intended fingers, the DF method consistently showed the lowest estimation error, with RMSE of the index ($7.86 \pm 0.72\%$ MVC (mean $\pm$ standard error)), middle ($7.18 \pm 0.49\%$MVC), and ring fingers ($5.95 \pm 0.71\%$MVC). The two-way (*method* (DF vs. CNN vs. AMP) $\times$ *finger* (index vs. middle vs. ring)) repeated measures ANOVA revealed a significant effect of the method on the RMSE ($F(2,12) = 130.91, p < 0.01$ and $R^2 (F(2,12) = 41.38, p < 0.01)$ with no interaction effect. The significant effect on *fingers* was found on the RMSE ($F(2,12) = 4.80, p < 0.01$) with no significant effect on R^2 ($F(2,12) = 2.23, p > 0.05$). *Post-hoc* comparisons showed that DF methods achieved significantly lower RMSE and higher R^2 on index and middle fingers ($p < 0.05$) compared to AMP and CNN methods. However, the performance differences between CNN and DF on ring fingers were not significant ($p > 0.05$). A representative trial of the flexion-extension task of multi-finger was depicted in Fig.6(A). Both the DF and CNN methods demonstrated a precise fit to the recorded force, not only during specific finger activation but also during simultaneous multi-finger movements.

We evaluated the estimation errors for forces exerted by unintended fingers. The root mean square error (RMSE) for these unintended fingers was illustrated in Fig.7 (A). A two-way (*method* (DF vs. CNN vs. AMP) $\times$ *finger* (index vs. middle vs. ring)) repeated measures ANOVA revealed significant effects of both the *method* (F(2,12)=26.09, $p < 0.001$) and the *fingers* (F(2,12)=7.89, $p < 0.05$) on the RMSE without interaction effect. Notably, a *post-hoc* comparison revealed that the RMSE for the DF and CNN methods was significantly lower than that of the AMP method ($p < 0.05$) on three fingers. Yet, there were no significant differences in RMSE between DF and CNN methods. This observation underscores the ability of both DF and CNN to effectively distinguish individualized finger forces during multi-finger motor output.

We further quantified the active and rest states of finger muscle contraction based on the predicted force and a predetermined threshold. The average rates of false activation, false rest, and detection accuracy are presented in Fig. 7 (B). The repeated measures ANOVA revealed a significant difference in the false activation rate across the three methods, irrespective of the threshold used (5% MVC: $F(2,12) = 51.63, p < 0.01$; 10% MVC: $F(2,12) = 58.71, p < 0.01$; 15% MVC: $F(2,12) = 15.78, p < 0.01$). Subsequent paired comparisons with *post-hoc* demonstrated that the false activation rate of the DF method was significantly lower than the other two methods ($p < 0.05$). Similarly, the detection accuracy also showed significant differences across the three methods, irrespective of the threshold used (5% MVC: $F(2,12) = 29.15, p < 0.01$; 10% MVC: $F(2,12) = 71.12, p < 0.01$; 15% MVC: $F(2,12) = 26.09, p < 0.01$), subsequent *post-hoc* tests indicated that the DF method achieved significantly higher accuracy than both AMP ($p < 0.05$) and CNN ($p < 0.05$) under all thresholds. The false rest rate showed significant differences between methods under 10% MVC ($F(2,12) = 16.15, p < 0.01$) and 15% MVC ($F(2,12) = 37.61, p < 0.01$). However, at 5% MVC, no significant difference was observed ($F(2,12) = 0.43, p > 0.05$).

C. Exploring Minimal Training Data

To further enhance the training efficiency, we determined the minimum amount of data required to train the models effectively. We progressively reduced the training data from 100% to 75%, 50%, and 25% using a subset of the single-finger data (corresponding to 1, 2, and 3 out of the 4 total available trials). This exploration allowed us to identify the optimal volume training data that can yield satisfactory performance (Fig.8). Two-way repeated measures ANOVA (*method* $\times$ *training size*) revealed a significant effect of training size on both RMSE ($F(3,18) = 16.80, p < 0.01$) and $R^2 (F(3,18) = 11.04, p < 0.01)$ with no interaction with the method. We observed that the DF method consistently demonstrated the lowest RMSE and the highest R^2 regardless of the training data size. Remarkably, using merely 25% of the training data, the force estimation reached an RMSE of $9.3 \pm 0.4\%$MVC and an R^2 value of 0.78 ± 0.03. Upon *post-hoc* analysis, both the DF and CNN methods were seen to substantially surpass the AMP method under every condition. However, a significant difference in RMSE between CNN and AMP was only evident when utilizing 75% and 100% data ($p < 0.05$). The difference in R^2 between CNN and

DF was only significant when utilizing 50% and 100% data. Overall, these findings underscore the DF method's proficiency in harnessing a limited training dataset to effectively decode finger force to some extent.

D. Robustness to Noise

As depicted in Fig.9, a declining trend in performance was noticeable for all methods as SNR decreased. The two-way repeated measures ANOVA (*method* $\times$ *SNR*) showed an interaction effect between these two factors for both RMSE ($F(6,36) = 25.92, p < 0.01$) and R^2 ($F(6,36) = 4.78, p < 0.01$). Under all conditions, both DF and CNN show significantly outperformed the AMP method. Significant differences in RMSE between DF and CNN were observed when $SNR > 10$ ($p = 0.02$ for $SNR = 15$ and $p = 0.015$ for $SNR = 20$).Similarly, significant differences in R^2 between DF and CNN were observed when $SNR > 10$ ($p = 0.04$ for $SNR = 15$ and $p = 0.03$ for $SNR = 20$). Despite the observable performance variation, the DF method was able to maintain the best performance across most scenarios.

TABLE I: The training and testing times (mean ± standard error) are provided for the three methods. The testing time indicates the time taken to infer one data segment, defined as an EMG signal with a 500-ms window and a 100-ms moving step. Both training and testing times are averaged across all subjects.

Method	Training time (second)	Testing time (second)
AMP	$6 \times 10^{-4} \pm 1.6 \times 10^{-4}$	$2.98 \times 10^{-5} \pm 5.3 \times 10^{-7}$
CNN	119.67 ± 4.18	$1.2 \times 10^{-3} \pm 1.4 \times 10^{-4}$
DF	18.20 ± 0.92	$3.3 \times 10^{-2} \pm 4.9 \times 10^{-3}$

E. Training Time Efficiency

The average layer depth for DF across experimental runs is 2.28 ± 0.17, with a maximum depth of 5 and a minimum of 2. In table I, we documented the training time for the AMP ($6 \times 10^{-4} \pm 1.6 \times 10^{-4}$ seconds), CNN (119.67 ± 4.18 seconds), and DF (18.20 ± 0.92 seconds) methods. The EMG-amplitude method showed the shortest training time, and the DF method exhibited approximately 5 times faster training time than that of CNN. The average time for inferring a single EMG input segment (with a 500-ms window and a 100-ms updating interval) was also estimated by three methods. For the AMP method, the inference time is $2.98\times10^{-5}\pm5.3\times10^{-7}$ seconds. The CNN, with a time of $1.2 \times 10^{-3} \pm 1.4 \times 10^{-4}$ seconds, was more efficient in inference compared to the DF, which took $3.3 \times 10^{-2} \pm 4.9 \times 10^{-3}$ seconds.

IV. Discussion

Our study addressed the intricate challenge of decoding concurrent multi-finger motion intention from EMG signals for robotic hand control. Due to the numerous combinations of multi-finger movement, capturing exhaustive data for every possible scenario is not feasible. Consequently, we sought to harness single-finger data to predict multi-finger movements, which could enhance training efficiency and practicality in controlling advanced prosthetics.

Although AMP [11], [13] and CNN methods [23], [29], [30] have been widely explored for multi-finger force prediction, achieving satisfactory performance using only single-finger data remains challenging. A key reason might be that the AMP method lacks the necessary complexity to effectively manage the finger-specific nonlinear mapping between HD-EMG features and forces. The CNN method, while capable of modeling complex finger-force relationships, often requires large data volumes for effective training to avoid overfitting. The DF method potentially bridges this gap between model complexity and data requirements. Our findings suggest that DF significantly outperforms both AMP and CNN methods. With its training efficiency and data efficacy, DF emerges as a particularly promising technique for facilitating fine motor control in advanced prosthetic hands

Our initial results suggested that all three methods (EMG amplitude, CNN, and DF) were capable of predicting multi-finger forces with high accuracy when the multi-finger data were utilized during training. The DF and CNN methods, in particular, exhibited superior predictive performance, indicating that complex non-linear modeling is essential to capture the EMG-force relation. However, when we only used single-finger data to train the model, an inevitable decrease in performance was observed across all methods, with the EMG amplitude method demonstrating notably poor performance. It was noticeable that the activation patterns for the flexion/extension of different fingers display distinct localized activations on the 2D electrode grid. Consequently, conventional amplitude methods, which use information from all channels, may not be optimal for force prediction at the individual finger level, often resulting in overestimated or underestimated forces. Channel refinement is critical for identifying finger-specific channel pools and contributes to accurate force predictions for each finger [11], [13]. In our experiments, a significant increase in prediction error was observed when multi-finger data was unavailable (which means that the application of channel refinement is inapplicable) for the AMP method.

By contrast, both CNN and DF had the model complexity necessary to establish the complex relationship between EMG features and multi-finger forces, even when only single-finger is available for training. In our study, the DF method demonstrated notable performance advantages over CNN, particularly for target fingers. Specifically, DF benefits from its inherent ability to perform implicit feature selection and transformation, allowing it to focus on and effectively utilize the most significant features [27]. This is especially beneficial when HD-EMG features from each channel did not uniformly contribute to the forces of different fingers. While CNNs excel at processing complex data structures and identifying intricate patterns, their lack of an explicit feature selection mechanism means that they extract hierarchical representations from all EMG channels. This approach may result in overestimated force values compared to DF in certain scenarios (Fig.6). The different mechanisms in feature selection might primarily account for the observed difference in performance between

CNN and DF methods.

Both the CNN and DF methods exhibited significantly reduced estimation errors on unintended fingers compared to the AMP method. Instead of comparing against a zero value, we evaluated the force of unintended fingers against the actual recorded force. This approach is crucial as accurately estimating the force of unintended fingers contributes to the natural control of prosthetics. Muscle contractions were classified into two states: rest and active. We assessed the ability of the decoder to separate fingers using metrics such as the false activation rate, false rest rate, and overall state detection accuracy. The DF method demonstrated the lowest false activation rate and the highest detection accuracy across various thresholds, likely due to its explicit feature selection capability that effectively distinguishes force differences across fingers. However, this advantage becomes less pronounced in detecting lower force levels. For example, the false rest rates of CNN and DF are notably similar when the thresholds are set at 5% and 10% MVC.

The scarcity of training data for EMG-force modeling remains a significant challenge, largely due to the high costs of data acquisition. Our observations indicated that the DF method consistently outperforms other methods in most scenarios, even with training data progressively reduced to only one trial per finger. This superior performance is likely due to the inherent ability of DF to adaptively adjust its complexity based on the available training data, thereby effectively utilizing limited datasets. In contrast, for the CNN, we implemented a lightweight architecture with approximately 4.2×10^4 parameters for optimization. However, the complexity of this neural network model is fixed before training and did not adapt to the volume of data available. This static nature of the CNN model limited its effectiveness when dealing with with a limited amount of training data.

Our study revealed that both CNN and DF exhibit a certain degree of resistance to noise. This resilience underscores their potential for implementation in real-world applications, where encountering signal interference is common. Specifically, in testing signals with an SNR above 15 dB, the DF method consistently outperformed the CNN. However, as the noise in the testing signal increased and the SNR decreased, the performance gap between DF and CNN narrowed. This observation suggests that while the DF method may be preferable in environments with relatively clean signals, both models demonstrate comparable performance in high-noise scenarios.

Before training, the CNN underwent an extensive grid search for the best filter size and layer depth. Conversely, the DF method showed its efficiency by necessitating minimal hyperparameter tuning. Ideally, a system for prosthesis control should need short and easy training [31]. In our study, the AMP method completed training in just 0.6 ms. While training the CNN took an average of approximately 2 minutes, the DF method was notably more efficient, requiring only 18.2 seconds. The longer training duration of CNN was expected given the inherent complexity of neural networks, which involve multiple layers of interconnected neurons to be optimized via back-propagation. By contrast, the DF method exhibited a balance between the two methods, with a training time of around 18.20 seconds, highlighting its ability to offer sufficient complexity without compromising efficiency to some extent. Regarding testing time, all methods could complete the inference within the input update time (100 ms), indicating that all methods were suitable for real-time applications. As expected, the AMP method exhibited the shortest inference time, while the CNN method demonstrated a faster inference speed than DF. This can be attributed to the efficient implementation of the deep neural network framework, which was enhanced by numerous acceleration techniques using GPU computations (we also used PyTorch in this work). On the other hand, as deep forest is a relatively new technique, its implementation was performed using CPUs. Nonetheless, the current implementation still meets real-time requirements. As a burgeoning trend in the field, there are implementations like "ForestLayer" [32] that are considerably faster than current methods. We foresee significant advancements and refinements in the inference efficiency of DF as this emerging technique progresses.

In comparison with previous studies on force prediction at the individual finger level, we found that the DF method surpasses the conventional regression-based amplitude method and the state-of-the-art neural drive method [11], [13], including CNN models trained on the neural drive information [33], [34]. Clearly, global features are more efficient to extract than neural drive information, with the latter requiring computationally intensive motor unit decomposition. Furthermore, the neural drive method [11], [13] required data from multi-finger trials to refine the motor unit pools for accurate force prediction. In contrast, our approach predicted multi-finger movement forces solely based on data from single-finger trials. Overall, our approach introduces new and efficient solutions for multi-finger motor decoding, expanding upon existing methods.

Despite the promising results, several limitations should be addressed in future works. Firstly, the study was conducted with seven subjects, and the applicability of the proposed method to a larger population should be involved to validate the findings. Secondly, the current study only involved intact subjects. The effectiveness of the proposed methods has not been validated on amputees. The muscle activation patterns in amputees can undergo significant alterations, combined with the potential reduction in available EMG recording channels due to amputation. Future studies should aim to address these complexities and evaluate the adaptability of these methods to the unique EMG characteristics of amputees. Third, our focus was on isometric contractions. In reality, hand movements are more complex, often involving a range of dynamic motions, which plays a crucial role in fine motor tasks our hands perform daily. Therefore, future work should consider incorporating dynamic finger joint estimation to potentially allow for finer motor control in practical settings.

V. Conclusion

This study offered insights into enhancing the training efficiency of neural decoders potentially viable for fine control

of advanced prostheses. Our findings demonstrate that the DF method can achieve superior performance in concurrently predicting the flexion and extension force of individual fingers, even when training exclusively using single-finger data. With its superior performance and capacity to effectively utilize limited training data, the DF method can enable efficient EMG-based control systems for dexterous finger control in robotic applications.

REFERENCES

[1] C. Castellini, "Upper limb active prosthetic systems—overview," *Wearable Robotics*, pp. 365–376, 2020.

[2] M. S. Johannes, E. L. Faulring, K. D. Katyal, M. P. Para, J. B. Helder, A. Makhlin, T. Moyer, D. Wahl, J. Solberg, S. Clark *et al.*, "The modular prosthetic limb," in *Wearable Robotics*. Elsevier, 2020, pp. 393–444.

[3] T. Worsnopp, M. Peshkin, J. Colgate, and D. Kamper, "An actuated finger exoskeleton for hand rehabilitation following stroke," in *2007 IEEE 10th international conference on rehabilitation robotics*. IEEE, 2007, pp. 896–901.

[4] K. Tong, S. Ho, P. Pang, X. Hu, W. Tam, K. Fung, X. Wei, P. Chen, and M. Chen, "An intention driven hand functions task training robotic system," in *2010 Annual International Conference of the IEEE Engineering in Medicine and Biology*. IEEE, 2010, pp. 3406–3409.

[5] G. Li, A. E. Schultz, and T. A. Kuiken, "Quantifying pattern recognition—based myoelectric control of multifunctional transradial prostheses," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 18, no. 2, pp. 185–192, 2010.

[6] X. Jiang, X. Liu, J. Fan, X. Ye, C. Dai, E. A. Clancy, M. Akay, and W. Chen, "Open access dataset, toolbox and benchmark processing results of high-density surface electromyogram recordings," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 29, pp. 1035–1046, 2021.

[7] R. N. Khushaba, S. Kodagoda, M. Takruri, and G. Dissanayake, "Toward improved control of prosthetic fingers using surface electromyogram (emg) signals," *Expert Systems with Applications*, vol. 39, no. 12, pp. 10 731–10 738, 2012.

[8] 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 Transactions on neural systems and rehabilitation engineering*, vol. 20, no. 5, pp. 663–677, 2012.

[9] J. G. Ngeo, T. Tamei, and T. Shibata, "Continuous and simultaneous estimation of finger kinematics using inputs from an emg-to-muscle activation model," *Journal of neuroengineering and rehabilitation*, vol. 11, no. 1, pp. 1–14, 2014.

[10] C. Dai, Z. Zhu, C. Martinez-Luna, T. R. Hunt, T. R. Farrell, and E. A. Clancy, "Two degrees of freedom, dynamic, hand-wrist emg-force using a minimum number of electrodes," *Journal of Electromyography and Kinesiology*, vol. 47, pp. 10–18, 2019.

[11] Y. Zheng and X. Hu, "Concurrent prediction of finger forces based on source separation and classification of neuron discharge information," *International journal of neural systems*, vol. 31, no. 06, p. 2150010, 2021.

[12] R. Roy, Y. Zheng, D. G. Kamper, and X. Hu, "Concurrent and continuous prediction of finger kinetics and kinematics via motoneuron activities," *IEEE Transactions on Biomedical Engineering*, 2022.

[13] Y. Zheng and X. Hu, "Concurrent estimation of finger flexion and extension forces using motoneuron discharge information," *IEEE transactions on biomedical engineering*, vol. 68, no. 5, pp. 1638–1645, 2021.

[14] E. A. Clancy, E. L. Morin, G. Hajian, and R. Merletti, "Tutorial. surface electromyogram (semg) amplitude estimation: Best practices," *Journal of Electromyography and Kinesiology*, vol. 72, p. 102807, 2023.

[15] P. Liu, D. R. Brown, F. Martel, D. Rancourt, and E. A. Clancy, "Emg-to-force modeling for multiple fingers," in *2011 IEEE 37th Annual Northeast Bioengineering Conference (NEBEC)*. IEEE, 2011, pp. 1–2.

[16] E. Clancy and N. Hogan, "Single site electromyograph amplitude estimation," *IEEE Transactions on Biomedical Engineering*, vol. 41, no. 2, pp. 159–167, 1994.

[17] X. Hu, N. L. Suresh, C. Xue, and W. Z. Rymer, "Extracting extensor digitorum communis activation patterns using high-density surface electromyography," *Frontiers in physiology*, vol. 6, p. 279, 2015.

[18] N. van Beek, D. F. Stegeman, J. C. van den Noort, D. H. Veeger, and H. Maas, "Activity patterns of extrinsic finger flexors and extensors during movements of instructed and non-instructed fingers," *Journal of electromyography and kinesiology*, vol. 38, pp. 187–196, 2018.

[19] C. Dai and X. Hu, "Extracting and classifying spatial muscle activation patterns in forearm flexor muscles using high-density electromyogram recordings," *International journal of neural systems*, vol. 29, no. 01, p. 1850025, 2019.

[20] T. Bao, S. Q. Xie, P. Yang, P. Zhou, and Z.-Q. Zhang, "Toward robust, adaptiveand reliable upper-limb motion estimation using machine learning and deep learning–a survey in myoelectric control," *IEEE journal of biomedical and health informatics*, vol. 26, no. 8, pp. 3822–3835, 2022.

[21] Y. Chen, C. Dai, and W. Chen, "Cross-comparison of emg-to-force methods for multi-dof finger force prediction using one-dof training," *IEEE Access*, vol. 8, pp. 13 958–13 968, 2020.

[22] G. Hajian, A. Etemad, and E. Morin, "Generalized emg-based isometric contact force estimation using a deep learning approach," *Biomedical Signal Processing and Control*, vol. 70, p. 103012, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1746809421006091

[23] A. Ameri, M. A. Akhaee, E. Scheme, and K. Englehart, "Regression convolutional neural network for improved simultaneous emg control," *Journal of neural engineering*, vol. 16, no. 3, p. 036015, 2019.

[24] X. Li, X. Zhang, L. Zhang, X. Chen, and P. Zhou, "A transformer-based multi-task learning framework for myoelectric pattern recognition supporting muscle force estimation," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 2023.

[25] R. Hu, X. Chen, H. Zhang, X. Zhang, and X. Chen, "A novel myoelectric control scheme supporting synchronous gesture recognition and muscle force estimation," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 30, pp. 1127–1137, 2022.

[26] C. Lin, X. Chen, W. Guo, N. Jiang, D. Farina, and J. Su, "A bert based method for continuous estimation of cross-subject hand kinematics from surface electromyographic signals," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 31, pp. 87–96, 2022.

[27] Z.-H. Zhou and J. Feng, "Deep forest," *National science review*, vol. 6, no. 1, pp. 74–86, 2019.

[28] V. Zatsiorsky, Z. Li, and M. Latash, "Enslaving effects in multi-finger force production," *Experimental Brain Research*, vol. 131, no. 2, pp. 187–195, 2000.

[29] D. Yang and H. Liu, "An emg-based deep learning approach for multi-dof wrist movement decoding," *IEEE Transactions on Industrial Electronics*, vol. 69, no. 7, pp. 7099–7108, 2021.

[30] C. Ma, W. Guo, H. Zhang, O. W. Samuel, X. Ji, L. Xu, and G. Li, "A novel and efficient feature extraction method for deep learning based continuous estimation," *IEEE Robotics and Automation Letters*, vol. 6, no. 4, pp. 7341–7348, 2021.

[31] D. Farina, N. Jiang, H. Rehbaum, A. Holobar, B. Graimann, H. Dietl, and O. C. Aszmann, "The extraction of neural information from the surface emg for the control of upper-limb prostheses: emerging avenues and challenges," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 22, no. 4, pp. 797–809, 2014.

[32] G. Zhu, Q. Hu, R. Gu, C. Yuan, and Y. Huang, "Forestlayer: Efficient training of deep forests on distributed task-parallel platforms," *Journal of Parallel and Distributed Computing*, vol. 132, pp. 113–126, 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0743731518305392

[33] R. Roy, F. Xu, D. G. Kamper, and X. Hu, "A generic neural network model to estimate populational neural activity for robust neural decoding," *Computers in Biology and Medicine*, vol. 144, p. 105359, 2022.

[34] J. Fan, L. Vargas, D. G. Kamper, and X. Hu, "Robust neural decoding for dexterous control of robotic hand kinematics," *Computers in Biology and Medicine*, p. 107139, 2023.