Explainable and Robust Deep Forests for EMG-Force Modeling

Explainable and Robust Deep Forests for EMG-Force Modeling

Xinyu Jiang, Kianoush Nazarpour*, *Senior Member, IEEE,* and Chenyun Dai*, *Member, IEEE,*

***Abstract*—Machine and deep learning techniques have received increasing attentions in estimating finger forces from high-density surface electromyography (HDsEMG), especially for neural interfacing. However, most machine learning models are normally employed as block-box modules. Additionally, most previous models suffer from performance degradation when dealing with noisy signals. In this work, we propose to employ a forest ensemble model for HDsEMG-force modeling. Our model is explainable and robust against noise. Additionally, we explored the effect of increasing the depth of forest models in EMG-force modeling problems. We evaluated the performance of deep forests with a finger force estimation task. Training and testing data were acquired 3—25 days apart, approximating realistic scenarios. Results showed that deep forests significantly outperformed other models. With artificial signal distortion in 20% channels, deep forests also showed a higher robustness, with the error reduced from that of the baseline by >50% compared with all other models. We provided explanations for the proposed model using the mean decrease impurity (MDI) metric, revealing a strong correspondence between the model and physiology.**

***Index Terms*—EMG-force modeling, electromyography, myoelectric control, deep learning, deep forest**

I. Introduction

SURFACE electromyography (sEMG)-force modeling is essential for a diversity of applications, such as neural interfacing, EMG-impedance mapping, ergonomic assessment, control of exoskeletons, and neuroprosthetics [1], [2]. Recent advancement in flexible electronics and wearable sensing [3], [4] promote the use of high-density sEMG (HDsEMG) electrode arrays in diverse EMG decoding applications [5]–[8]. By the use of HDsEMG, one can capture the neural code that drives motor function with a high spatial resolution, which in turn can contribute to the estimation of muscle-force mapping. Arguably, traditional least-squares-based polynomial models [2], [9] fall short of capturing the complex muscle-force relations when dealing with a large number of channels. Deep learning has proved a powerful computational model to extract useful information with multiple levels of abstraction from the training data [10]. In recent years, deep learning models have been widely applied in sEMG-force (kinematics) modeling tasks [11]–[14].

Most deep neural networks (DNNs), such as convolutional neural networks (CNNs), employ a model structure built upon parameterized differentiable non-linear modules, which can be trained by back-propagation [15]. However, most DNN structures as well as their training algorithms have many hyper-parameters and the overall performance is sensitive to the choice of the hyper-parameters. A deeper model is with a higher model complexity and hence may capture a more complex mapping functions between the input and output. However, with a small data size, models with higher complexity may overfit to certain biased patterns of the training data. Additionally, deep learning models are normally employed as block-box modules in a processing pipeline due to the lack of explainability. Meaningful explanations of the decision logic are recommended to ensure safety and uphold research ethics, especially in biomedical machine learning research [16].

The low robustness of EMG-force models against noises is another challenge that limits the real world applications. EMG signals can be contaminated by diverse noises. Previous studies [17], [18] found that if features extracted from the corrupted electrodes were not fixed by outlier smoothing filters, the performance of HDsEMG pattern recognition would be largely degraded. A series of previous studies [19], [20] also focused on how to identify or fix those electrodes corrupted by noises. However, most previous solutions require a separate module to monitor the real time signal quality in each electrode and fix the corrupted electrodes, increasing the computational burden in practical applications. Proposing a new model with inherent robustness against electrode corruption is therefore essential. Moreover, during the model training process, the model may overfit to those random patterns specific to a certain day (e.g., the patterns formed by a specific type of noise with a certain noise power in a group of electrodes, or the patterns formed by a certain physiological state, which may largely vary on a second day), leading to significant performance degradation in cross-day validation. Improving the robustness of models in cross-day applications is also essential, however challenging.

Decision tree is a transparent model with the logic of the decision making process clearly illustrated by the decision path from the root node to the leaf node. Random forest models built on multiple decision trees also retain the explainable property. With different decision trees focusing on different electrodes, those corrupted electrodes would only weaken the performance of a certain group of decision trees, with the overall robustness improved as well. However, no previous

Xinyu Jiang and Kianoush Nazarpour are with Edinburgh Neuroprosthetics Laboratory, School of Informatics, The University of Edinburgh, Edinburgh EH8 9AB, United Kingdom.

Chenyun Dai is with the Center for Biomedical Engineering, School of Information Science and Technology, Fudan University, Shanghai 200433, China.

*Corresponding authors: Kianoush Nazarpour (kianoush.nazarpour@ed.ac.uk), Chenyun Dai (chenyundai@fudan.edu.cn)

This work is supported in part by National Natural Science Foundation of China (Grant Number: 62001122) and in part by Natural Science Foundation of Shanghai (Grant No.20ZR1403400). The work of KN is supported by a grant from Engineering and Physical Sciences Research Council (EPSRC), UK (Grant No. EP/R004242/2).

Manuscript received XX XX, XXXX; revised XX XX, XXXX.

studies have provided a deep insight of forest models in EMG-force modeling problems, and the relation between the model explanations and the physiological basis of human neural systems remains unknown. Additionally, deep forest is a variant of forest model [21] which may be an alternative deep learning architecture built on explainable modules. Deep forest processes the input via a decision tree ensemble in each layer. Information could be processed layer-by-layer, forming a deep forest. Zhou et al. [21] have proved that, by cascading multiple random forest-based modules, the model performance showed significant improvement. According to the experiments of Zhou et al., deep forest is also quite robust with the hyperparameter selection. In the context of EMG-force modeling, deep forest is also a promising solution with explainability.

In this work, we provided explanations on forest-based models, by quantifying the importance of all electrodes for decision trees to make decisions via Mean Decrease Impurity (MDI) [22]. The effect of increasing the depth of forest models was also explored. Results demonstrated that increasing the depth of forest models can further refine the excellent performance of random forest models. Additionally, according to the model explanations, the important anatomical areas that forest models focused on are highly consistent with the spatial activation patterns of forearm motor neurons. Moreover, another important finding of our work is the extremely high robustness of forest models against channel corruption. Compared with other models, deep forest achieved excellent performance even when 20% channels were corrupted by noises. We evaluated the performance of the proposed model in estimation of the forces from each individual finger, which can be viewed as one of the most challenging tasks in sEMG-based motion decoding because our fingers contributed to the most subtle and dexterous activities in our daily life. In addition, our validations were performed on data collected over multiple days to take into account the cross-day variation of sEMG patterns. By comparing the performance of different models in both one-day and cross-day validations, we proved the better cross-day generalization of deep forests. Overall, the contributions of our work are as below:

We for the first time provide explanations for forest models in neurophysiological signal analyses and find the consistency between the model explanations and the physiological basis of human neural systems.

We explore the effect of increasing the depth of forest models in EMG-force regression tasks and show that a deeper forest model can further refine the excellent performance of a random forest model, outperforming benchmark models.

We demonstrate the high robustness of deep forest models against sensor noise, especially relevant when dealing with high-density sensors.

We report better cross-day generalization of deep forest models, which is an important consideration for practical applications.

II. Materials

A. Subjects

Twenty subjects (12 males, 8 females, 22-34 years old) took part in our experiment. The experiment protocol was reviewed and approved by the ethics committee of Fudan University (approval number: BE2035).

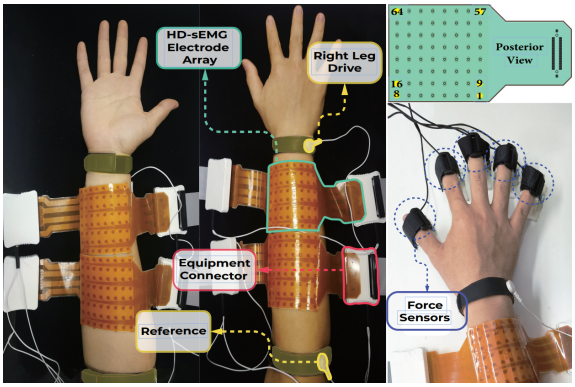


Fig. 1: Electrode placement.

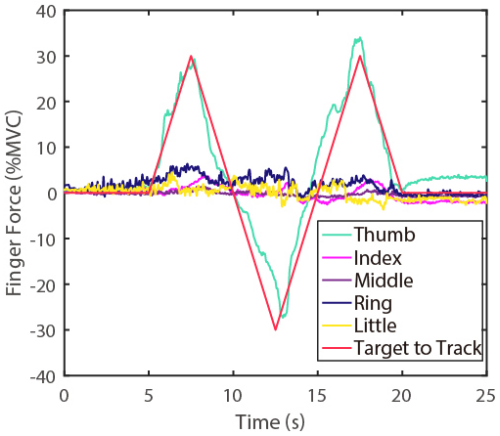


Fig. 2: Visual instruction of force-tracking tasks.

B. Data Acquisition

Data underpinning this paper was published in our previous work [23] and are available at the website (https://doi.org/10.13026/ym7v-bh53). A brief description of the dataset is presented below.

1) Setup:

Four HDsEMG electrode arrays, each of an 8×8 layout with 10 mm inter-electrode distance, were placed on the forearm. A Quattrocento amplifier (OT Bioelettronica, Torino, Italy) acquired 256-channel forearm sEMG signals, as shown in Fig. 1. The signals were amplified with a gain of 150, converted to digital with a 16-bit resolution, and sampled at 2048 Hz. The finger force data was collected using five sensor-amplifier pairs for five fingers (sensor: SAS, Huatran, Shenzhen, China; amplifier: HSGA, Huatran, Shenzhen, China), and registered at 100 Hz. Data of each subject was acquired on two separate days (3 to 25 days apart, 8.5 $\pm$ 6 days on average).

2) Maximum Voluntary Contraction (MVC) Measurement:

Before each experiment, we measured the forces at MVC in an isometric condition for each finger. Subjects performed the flexion and extension of each finger at MVC successively. Subjects were allowed to perform MVC at any time within a 10

s window. The measured MVC forces were used to normalize the ground truth of finger force.

3) Force-Tracking Tasks:

Subjects performed isometric contraction of each individual finger within a 25-s trial duration. In each trial, subjects activated a specific finger with varying forces. The real time measured force was presented on a screen and used as a feedback to track a target triangle force trajectory, as shown in Fig. 2. This target force trajectory ranged from -30% MVC to 30% MVC. A positive sign denotes finger extension. We recorded three trials per finger.

III. Methods

A. Preprocessing of HDsEMG Signals

The acquired HDsEMG signals were filtered with a 10-Hz Butterworth high-pass filter and a 500-Hz Butterworth low-pass filter. All filters were zero-phase digital filters processing the input bidirectionally, with 8^{th} order each direction. The powerline interference at 50 Hz and its harmonics up to 400 Hz were removed using a combination of notch filters. The ground truth force data was low-pass filtered to 10 Hz (zero-phase digital fingers processing the input bidirectionally, with 8^{th} order each direction).

B. Feature Extraction

In line with [2], [9], four features namely root mean square (RMS), waveform length (WL), slope sign change (SSC), and zero crossing (ZC) were extracted from each 40 continuous samples (19.5 ms) in each EMG channel. Previous studies [24] have investigated the effect of window length in sEMG-based hand gesture classification tasks. In this work, we select this window length similar with previous studies [9], [23]. Such selection of window length can provide a real time update of estimated finger forces in practical applications. The feature vector corresponding to each 19.5 ms time window consists of 1024 (256 channels × 4 features) elements. Similar to previous studies, e.g. [18], [25], to reduce the impact of outliers, a feature value larger than 3× standard deviation away from the mean of all 64 elements corresponding to each feature in each electrode array was replaced by the average of features in its 4 neighbor channels (2 or 3 neighbor channels for vertex and edge channels). The average force in each 19.5 ms time window was also calculated as the ground truth force for this time window. We removed the preprocessed data within the first and last 2 s of each 25 s task, to account for filter startup and tail transients (in practical scenarios, all processing steps can be performed via causal filters, with no need to remove the startup and tail transients).

C. Least-Squared-based Polynomial Models

Polynomial models were applied to provide one of the baseline force estimation results for comparison, described as the following:

$$y[i] = \sum_{v=1}^{V} \sum_{q=0}^{Q} \sum_{f=1}^{F} \sum_{c=1}^{C} \theta_{v,q,f,c} \cdot x_{f,c}^{v}[i-q] \tag{1}$$

where x is the input feature examples, y is the output finger force, i denotes the index of input examples, $V = 1$ and $V = 2$ refer to linear and quadratic models, respectively. $Q = 20$ denotes the number of time lags (the total number of windows is $Q + 1$ with both Q time lags and 1 window at the moment). $F = 4$ denotes the 4 types of features. $C = 256$ represents the number of channels. And θ is the model parameters. The above polynomial model can also be rewritten in a matrix form $X\Theta = Y$, or:

$$\begin{bmatrix} x_{1,1}^{1}[Q+1] & \cdots & x_{f,c}^{v}[Q+1-q] & \cdots & x_{F,C}^{V}[1] \\ \vdots & & \ddots & & \vdots \\ x_{1,1}^{1}[i] & \cdots & x_{f,c}^{v}[i-q] & \cdots & x_{F,C}^{V}[i-Q] \\ \vdots & & \ddots & & \vdots \\ x_{1,1}^{1}[N] & \cdots & x_{f,c}^{v}[N-q] & \cdots & x_{F,C}^{V}[N-Q] \end{bmatrix} \begin{bmatrix} \theta_{1,0,1,1} \\ \vdots \\ \theta_{v,q,f,c} \\ \vdots \\ \theta_{V,Q,F,C} \end{bmatrix} = \begin{bmatrix} y[Q+1] \\ \vdots \\ y[i] \\ \vdots \\ y[N] \end{bmatrix} \tag{2}$$

where each row of X represents a specific example, and each column of X represents a specific feature. The value of N was determined by the duration of training signals. The total number of rows in matrix X, or the total number of training examples is denoted by $N_{examples} = N - Q$.

The least squares (LS)-based model estimation algorithm was employed to train the polynomial model. The Moore-Penrose pseudoinverse regularization method [26] was applied to avoid overfitting, by first performing singular value decomposition on data matrix X and then discarding those small singular values, the ratio of which to the largest singular value were lower than a tolerance parameter. A previous study [27] investigated the effect of the tolerance parameter on finger force estimation from HDsEMG, and found that the regression error gradually converged and reached the minimum with the tolerance parameter set between 0.08 to 0.12 for cross-day validations. In this work, the tolerance parameter was selected as 0.1. The above polynomial models have been widely applied in estimating muscle forces from sEMG [23], and can also be applied in applications in estimating other kinematics information (e.g., finger positions/angles) from sEMG [28], [29].

For cross-day validation, data acquired on the first and the second day were used as the training and testing set, respectively. We also performed one-day validation for comparison, to evaluate the performance degradation in cross-day application scenarios. For one-day validation, the leave-one-trial-out strategy was applied. For each finger, each one of the three trials acquired on a specific day was used as testing data, with the other two trials on the same day used as training data. To ensure that all testing samples were exactly the same as that in cross-day validation (where data on the second day were used as testing set), only data in the second day were used in one-day validation. For least-square based polynomial models, a tolerance of 0.03 was selected in one-day validation (the same value suggested by the previous study [27] which reported that the optimal tolerance parameter in one-day applications were generally lower than that in cross-day applications because the effect of overfitting in cross-day validations is higher). Trial-and-error on data from the first day (which are independent with testing dataset) also demonstrated that the selected parameter can yield an excellent performance

compared with other choices. All other settings in one-day validation were the same as the cross-day validation.

D. Least Absolute Shrinkage and Selection Operator Models

The least absolute shrinkage and selection operator (Lasso) model is another widely applied model in estimating motion information from EMG data [30], [31]. In Lasso regression, the objective function is similar to the least-squares-based linear model but an $l1$-norm of the model parameters was added as a regularization term, denoted in equation (3),

$$\underset{\Theta}{min}\ \frac{1}{2N_{examples}}\|X\Theta - Y\|_2^2 + \alpha\|\Theta\|_1 \tag{3}$$

where α is a parameter tuning the effect of model regularization and $\|\Theta\|_1$ is the $l1$-norm. In this work, $\alpha = 0.005$ was selected via trial-and-error on data from the first day (the testing dataset on the second day were independent with parameter selection). Allocation strategies of training and testing data were completely the same as the counterpart of polynomial models.

E. Support Vector Regression Models

Support vector regression (SVR) has been proposed as an effective model for EMG decoding [32]. SVR models aim to find a function which takes the same form as the linear model described by equation (1). However, the optimal model parameters were not obtained based on the least squares criterion. Instead, the objective function is to minimize the $l2$-norm of the model parameters, denoted by the first term in the first line of equation (4). The regression error was then handled in the constraint of model optimization, i.e., the second line of equation (4), where the absolute error was set less than a margin value, i.e. the maximum error ϵ. The terms involving ξ introduced slack variables into the model optimization process. For an example that falls outside the maximum error ϵ, its deviation from the margin was denoted as ξ.

$$\begin{aligned} &\underset{\Theta}{min}\ \frac{1}{2}\|\Theta\|_2^2 + C\sum_{i=Q+1}^{N}|\xi_i| \\ &s.t.\ |\widehat{y}[i] - y[i]| < \epsilon + |\xi_i| \end{aligned} \tag{4}$$

In this work, $C = 1$ and $\epsilon = 0.01$ were selected via trial-and-error on data from the first day (the testing dataset on the second day were independent with parameter selection). The above SVR model can be further extended with kernel functions, similar to kernel support vector machine for classification tasks. In our work, the radial basis function was selected as the kernel function, the same choice as the previous study [32]. Allocation strategies of training and testing data were completely the same as the counterpart of polynomial models.

F. Bayesian Ridge Regression Models

Ridge regression and its variants have also been widely employed in EMG decoding [33], [34]. Bayesian ridge regression (BRR) [35], which estimates a probabilistic model of the regression problem and preserves the advantages of both ridge regression and Bayesian regression, was employed as another baseline model. In a general Bayesian regression problem, the output Y is assumed to be Gaussian distributed around $X\Theta$, denoted by equation (5),

$$p(Y|X,\Theta,\alpha) = \mathcal{N}(Y|X\Theta,\alpha) \tag{5}$$

where α is a random variable to be estimated from the given data. For BRR models in particular, the prior probability distribution of model parameters also follows a spherical Gaussian distribution, as presented by equation (6),

$$p(\Theta|\lambda) = \mathcal{N}(\Theta|0,\lambda^{-1}\mathbf{I}_d) \tag{6}$$

where λ is the precision parameter in the Gaussian distribution and $\mathbf{I}_d$ is the identity matrix with matrix rank the same as the length of feature vectors (d). The model parameters Θ, λ and α were estimated jointly during the model fitting process.

G. CNN Models

CNN models were also applied to provide baseline force estimation results for comparison. The architecture of CNN models is presented in Fig. 3. CNN models with 2, 3, 4, and 5 convolutional layers were applied to avoid a large performance variation with different model depths. For these models, the input data representations were $D \times Nr \times Nc$ feature tensors without vectorization, where $D = F\times(Q+1) = 4\times(20+1) = 84$ is the number of feature elements corresponding to each EMG electrode, and $Nr = Nc = 16$ are size of feature maps constructed by four 8×8 HDsEMG electrode arrays. Mean square error (MSE) was used as the objective function. Adam optimizer [36] was used to optimize the model parameters. Batch normalization was employed to speed up convergence and avoid gradient vanishing, with the learning rate α set to 0.001, exponential decay rates for the moment estimates ($\beta1$, $\beta2$) set to (0.9, 0.999). Other detailed parameters on the structure of the employed CNN model (e.g. the kernel size of the convolutional filters) can be found at Fig. 3. The CNN parameters were set via trial-and-error on data from the first day (the testing dataset on the second day were independent with parameter selection).

Allocation of training and testing data was the same as the counterpart of polynomial models, except that only 80% training data were used to update model parameters via backpropagation, with the other 20% allocated to the validation set in all cases for CNN. During the model training, the model performance on the validation set was evaluated after each epoch. A total of 50 training epochs were performed. The CNN model with the optimal performance on the validation set was saved as the final model.

H. Deep Forest Models

The architecture of deep forest is presented in Fig. 4. The forest ensemble module in each layer contains two estimators to obtain the primary label estimation in leaf nodes in each layer. Each estimator contains a random forest and a completely random forest (distinguished by different colors in Fig. 4). In a random forest model, within the available feature set for each decision tree, the feature contributing to the highest

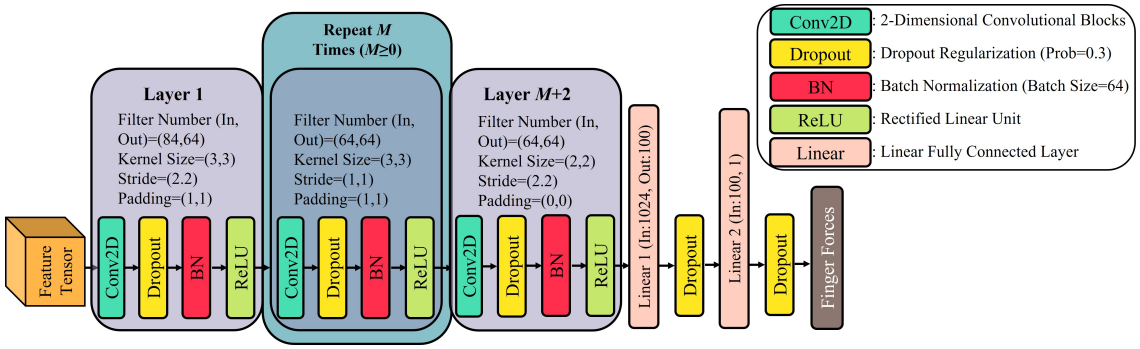


Fig. 3: CNN model architecture.

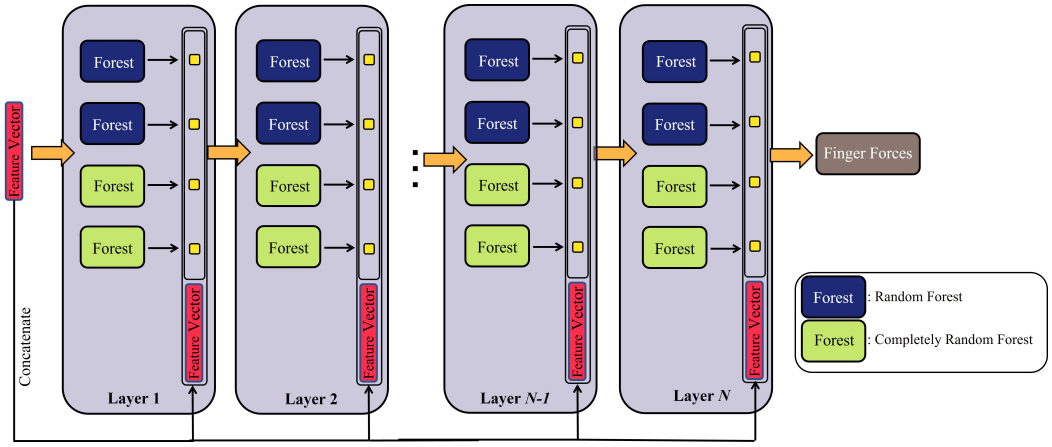


Fig. 4: Deep forest model architecture.

impurity decrease at each node would be picked to split examples in the current node into two branches. However, in a completely random forest model, the split in each node was determined in a random way instead of exhaustively evaluating the quality of a split on all possible features and split values of features then picking the one with the highest impurity decrease. The combination of different types of random forests in each estimator can encourage diversity, as demonstrated in the previous study [21]. Specifically, for a machine learning model based on individual learner ensembles, the overall model performance would be excellent if all individual learners are accurate and diverse, so that their predictions are reliable and can complement each other. By constructing both random forests and completely random forests in each module, different forests constructed by different rules would increase the diversity of decision trees. After processing in each layer, the estimation output of all forest modules can serve as a new feature vector. The original input feature vectors (4 types of features $\times$ 256 channels $\times$ (20+1) time windows = 21504 elements) were then concatenated with the estimation output to form an augmented feature vector, and then fed to the next layer. Such simple feature augmentation is beneficial to information learning in a layer-by-layer way [21]. To train a deep forest model, the module in the first layer was first trained and then fixed. With the fixed module in the first layer, the second layer was then trained. The same training strategy was repeated to train modules in all layers one by one. When the training procedure of a new layer was finished, the model performance on the validation dataset was evaluated. If increasing a new layer did not contribute to a decrease of regression error by a threshold (set as 10^{-5}), the new layer was discarded and the training procedure terminated. In this work, each random forest generates 100 decision trees. The number of trees in each forest module was set via trial-and-error on data from the first day (the testing dataset on the second day were independent with parameter selection). According to [21], the model performance is also highly robust with different parameters, and the same set of parameters can contribute to excellent performance in different tasks.

For both one-day and cross-day validations, allocation of training and testing data were completely the same as the counterpart of polynomial models, but in all cases for deep forest models, the training data were further divided into two parts, a growing set (80%) and an estimating (validation) set (20%). The growing set was used to grow a new cascade and the estimating set was used to estimate the model performance each time a new cascade was built to determine if the model depth should be further increased.

I. Explainability Analysis of Deep Forest via MDI

Decision tree is a highly explainable machine learning algorithm. The path from the root node to all leaf nodes of the decision tree explicitly shows the logic of the decision making procedure. Formed by multiple decision trees via both tree

ensembling (forming a random forest) and forest cascading (increasing the depth), deep forest shows its superiority in explainability. In this work, we calculated the importance of all input features via MDI [22]. MDI is the criterion that a decision tree employed to split an existing node into new nodes. The decrease in impurity of samples from an existing node to new nodes should be maximized. Sum of squared residuals (SSR) is a widely used measures of sample impurity for regression tasks, defined in equation (7).

$$\begin{aligned} SSR &= \sum_{i \in node} (\widehat{y}_{node} - y_i)^2 \\ \widehat{y}_{node} &= \frac{1}{N_{node}} \sum_{i \in node} y_i \end{aligned} \tag{7}$$

where N_{node} is the number of samples in a specific node. By calculating the MDI in the node corresponding to each feature, we can obtain the importance of different features of all decision trees. Because the MDI values of different features were with a large variance, we first logarithmically transformed the MDI data (via $x = log_{10}(x)$) as the feature importance, to adapt the gaps between extremely high and low MDI values. The average of importance values obtained from all decision trees was calculated. Considering multiple features were extracted from each electrode, we can further obtain the global electrode importance by calculating the average importance (with equal weights) of features extracted from each electrode. In this way, we can intuitively see which electrodes the deep forest model picked in estimating the forces of each finger. Because deep forest models for different subjects were with different numbers of layers (which were automatically determined during model training in a data-dependent way), the importance of features in the first layer was analyzed.

After obtaining the electrode importance, we further quantified the locations of those important electrodes related to each finger. We first normalized the electrode importance of each finger and each subject to a range of [0,1]. An electrode was defined as important if its normalized importance value was higher than a threshold of 0.8. With a low threshold we cannot select those electrodes which are really important, but with a too high threshold, many important electrodes would be ignored in our analyses. The locations of important electrodes were defined as the weighted average of coordinate values of all important electrodes. Normalized electrode importance values were used as location weights. The quantified locations of important electrodes related to each finger can provide guidance on electrode optimization for future studies.

J. Robustness Analysis with Different Levels of Signal Distortion

To evaluate the robustness of each model against noises, different noise sources were artificially added to the acquired HDsEMG signals. Following [19], Gaussian white noise and power line interference were considered. These two types of noises are the most common ones in sEMG recordings [37], [38]. The amplitude of Gaussian white noise added to each channel is set to 10% of the amplitude of sEMG in that channel. Simulated power line interference was generated as the summation of components at 50 Hz and its harmonics up to 400 Hz. The amplitude of power line interference was set to 5 times of the amplitude of sEMG signal in each channel, with the phase randomly selected within [0, 2π]. To simulate different levels of signal distortion, we randomly added noise to each channel with a certain probability (parameter $prob$). Probability was set from 0 to 0.2 with an increment of 0.04. We assumed that a probability of 0.2 (equivalent to 256×0.2=51.2 corrupted channels) would lead to an extremely high level of signal distortion. Note that we only artificially distorted the signals in testing set because signal quality during testing phase cannot be controlled but acquisition of training data can be well-supervised to avoid extremely low-quality signals.

K. Evaluation Metrics

Root mean square error (RMSE), Pearson correlation coefficient (COR), and coefficient of determination (also known as R^2) were used as the performance evaluation metrics, defined as follows,

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y[i] - \widehat{y}[i])^2} \tag{8}$$

$$COR = \frac{\sum_{i=1}^{N}(y[i] - \bar{y})(\widehat{y}[i] - \bar{\widehat{y}})}{\sqrt{\sum_{i=1}^{N}(y[i] - \bar{y})^2}\sqrt{\sum_{i=1}^{N}(\widehat{y}[i] - \bar{\widehat{y}})^2}} \tag{9}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(y[i] - \widehat{y}[i])^2}{\sum_{i=1}^{N}(y[i] - \bar{y})^2} \tag{10}$$

where N is the number of testing examples.

L. Statistical Analysis

The Shapiro-Wilk test was first performed to verify the normality of data. Because the result data obtained by different models did not consistently follow a normal distribution, we performed the Friedman test, a non-parametric alternative to the repeated measures analysis of variance, to test the overall significance of the performance difference among all models. Then we further compared the performance of each pair of models. If the metric data obtained by the two models followed the normal distribution, the paired-sample t-test was performed. Otherwise, the non-parametric Wilcoxon signed-rank test was performed. Bonferroni correction was also performed (if applicable) to avoid the multi-comparison error. Significant difference was claimed if $p < 0.05$ was observed.

IV. RESULTS

A. Comparison of Different Models in One-Day Validation

Results yielded by different models in one-day validation were presented in Table I. RMSE, COR, and R^2 metrics were all reported. For LS-based polynomial models, linear models significantly outperformed quadratic models evaluated by all three metrics ($p < 0.01$ for all metrics). In all following analyses, the performance of linear model was selected to represent the performance LS-based polynomial models. When

TABLE I: Results of different models in one-day validation. Red color represents the optimal results.

Models		RMSE (%MVC)	COR (%)	R^2 (%)
LS-Polynomial	Linear	4.6±1.9	94.8±5.7	83.1±14.1
	Quadratic	6.5±1.9	89.6±7.3	73.9±13.6
Lasso		4.7±1.6	95.1±4.6	85.2±10.9
SVR		4.6±1.6	95.1±4.8	86.2±10.6
Bayesian		4.5±2.0	95.6±5.5	84.8±14.8
CNN	2 Layers	4.9±1.8	94.9±5.3	84.0±13.3
	3 Layers	5.0±1.8	94.5±5.7	83.4±13.7
	4 Layers	5.0±1.9	94.4±5.8	83.4±14.3
	5 Layers	5.1±2.0	93.8±6.4	82.0±15.2
Deep Forest (2.9±0.4 Layers)		4.6±1.7	95.2±5.1	86.6±10.5

TABLE II: Results of different models in cross-day validation. Red color represents the optimal results

Models		RMSE (%MVC)	COR (%)	R^2 (%)
LS-Polynomial	Linear	10.0±2.9	87.1±10.4	25.7±40.3
	Quadratic	21.0±17.6	64.9±22.0	-661±1588
Lasso		8.7±2.1	89.8±9.0	54.2±19.9
SVR		9.9±2.8	82.4±11.3	45.0±20.1
Bayesian		10.6±3.5	83.2±13.6	28.3±38.3
CNN	2 Layers	8.6±2.6	89.8±9.1	52.9±25.2
	3 Layers	8.8±3.0	88.6±9.2	51.9±27.0
	4 Layers	8.9±3.0	88.3±9.5	49.8±32.1
	5 Layers	9.0±2.9	87.4±9.7	48.8±29.2
Deep Forest (2.6±0.7 Layers)		8.0±2.3	90.0±10.1	63.1±17.2

comparing the performance of CNN models with different numbers of layers, statistical significance on RMSE and R^2 metrics were found only in comparison between 2-layer and 5-layer CNN ($p = 0.028$ for RMSE and $p < 0.01$ for R^2), with 2-layer CNN achieving better performance. For the COR metric, significant differences were found for comparison between 2-layer CNN and CNN with all other numbers of layers ($p = 0.014$, $p = 0.032$ and $p < 0.01$ compared with 3-layer, 4-layer and 5-layer CNN, respectively). In all following analyses, the performance of 2-layer CNN was selected to represent the performance of CNN models. The Friedman test also demonstrated an overall significance when comparing the performance of different models ($p < 0.01$ for all of RMSE, COR and R^2 metrics). Comparisons between deep forest and each of the other models were also performed. Specifically, if evaluated by RMSE and R^2, significantly difference was found only in comparison with CNN models ($p = 0.047$ and $p = 0.047$ for RMSE and R^2 metrics, respectively), with deep forest models achieved better performance. If measured by COR, significantly difference was found only in comparison between deep forests and Bayesian models ($p = 0.014$), with Bayesian models achieved better performance. Overall, deep forest and other models with the optimal settings all contribute to relative low regression error in one-day validations.

B. Comparison of Different Models in Cross-Day Validation

Results yielded by different models in cross-day validation were presented in Table II. For LS-based polynomial models, linear models significantly outperformed quadratic models evaluated by all three metrics ($p < 0.01$ for all metrics). For CNN models, we found no significant performance differences between 2-layer CNN and all other CNN models with different numbers of layers evaluated by RMSE and R^2. When evaluated by COR, significant performance differences were found in comparison between 2-layer CNN and 4-layer CNN, and in comparison between 2-layer CNN and 5-layer CNN ($p = 0.015$ and $p < 0.01$, respectively). In all following analyses, the performance of linear model was selected to represent the performance LS-based polynomial models, and the performance of 2-layer CNN was selected to represent the performance of CNN models. The Friedman test likewise demonstrated an overall significance when comparing the performance of different models ($p < 0.01$ for all of RMSE, COR and R^2 metrics). Comparisons between deep forest and each of the other models were also performed. Specifically, when evaluated by RMSE, significant differences were found when comparing deep forest models with LS-linear models, Lasso models, SVR models and Bayesian models ($p < 0.01$, $p = 0.026$, $p < 0.01$, $p < 0.01$, respectively). When evaluated by COR, we found significant differences when comparing the performance of deep forest models with LS-linear models and SVR models ($p = 0.032$ and $p < 0.01$, respectively), and found a near significant difference when comparing the performance of deep forest models with Bayesian models ($p = 0.060$). When evaluated by R^2, deep forest models outperformed all other models ($p < 0.01$, $p = 0.013$, $p < 0.01$,$p < 0.01$ and $p = 0.045$ compared with LS-linear models, Lasso models, SVR models, Bayesian models and 2-layer CNN models, respectively). In all, the relatively comparable performance of different models in one-day validation and the overall better performance of deep forest models in cross-day validation demonstrated the high cross-day generalization ability of deep forest on sEMG-force modeling task.

C. Comparison of Deep Forest Models and 1-Layer Forest Ensemble Models.

To prove the necessity of using deep models, we compared the performance of deep forest models with fewer-layer forest models. The number of layers of the deep forest model for each subject was determined automatically during model training in a data-dependent way, and an average of 2.9 layers and 2.6 layers across all subjects were finally determined in one-day and cross-day validations, respectively. Considering the trained deep forest models for several subjects were only with 2 layers, to investigate the effect of model depth, here we only compared the performance of the final deep forests and the counterpart with the number of layers reduced to 1. Quantitative results of 1-layer forest ensemble and deep forest models in one-day and cross-day validations are presented in Table III and Table IV, respectively. When increasing the number of layers, models for all subjects achieved improved R^2 in one-day validations, and models for 15 out of 20 subjects achieved improved R^2 in cross-day validations, with an overall statistical significance observed in both one-day and cross-day validations and evaluated by all of RMSE, COR and R^2 metrics.

TABLE III: Comparison between 1-layer forest ensemble and deep forest models in one-day validation.

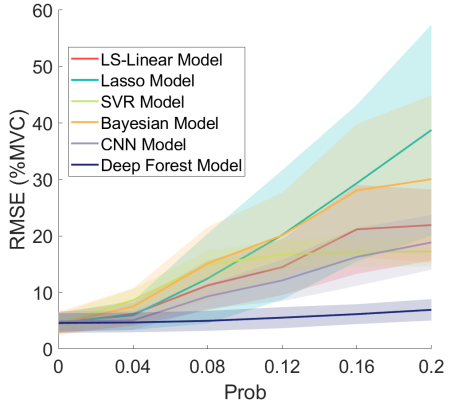
Models	RMSE (%MVC)	COR (%)	R^2 (%)
1-Layer Forest Ensemble	4.7±1.8	95.0±5.3	86.0±10.7
Deep Forest (2.9±0.4 Layers)	4.6±1.7	95.2±5.1	86.6±10.5

TABLE IV: Comparison between 1-layer forest ensemble and deep forest models in cross-day validation.

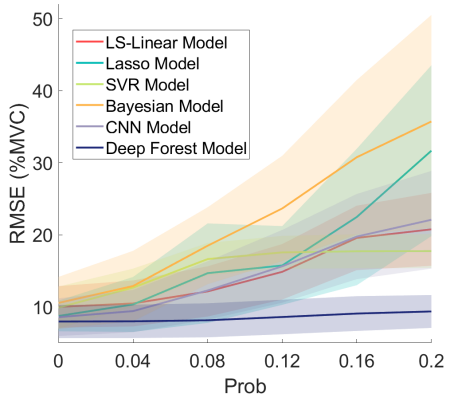
Models	RMSE (%MVC)	COR (%)	R^2 (%)
1-Layer Forest Ensemble	8.1±2.3	89.8±10.1	62.5±17.1
Deep Forest (2.6±0.7 Layers)	8.0±2.3	90.0±10.1	63.1±17.2

D. Comparison of Different Models with Artificial Signal Distortion

Regression results with increasing levels of signal distortion (different values of parameter $prob$) in both one-day and cross-day validations were presented in Fig. 5. In one-day validations, although all models achieved excellent performance with a good signal quality, the regression errors of all models except deep forest models dramatically increased with a higher level of signal distortion, as demonstrated in Fig. 5a. We calculated the increase in regression errors from $prob = 0$ to $prob = 0.2$. The increase in regression errors of deep forest models was significantly lower than those of all other models ($p < 0.01$ in all cases) in one-day validations. In cross-day validations, deep forest models likewise achieved significantly lower error increments with signal distortion (from $prob = 0$ to $prob = 0.2$) compared with all other models ($p < 0.01$ in all cases). Overall, deep forest models showed a highly robust performance with increasing levels of signal distortion.



(**a**) Results of one-day validation.



(**b**) Results of cross-day validation.

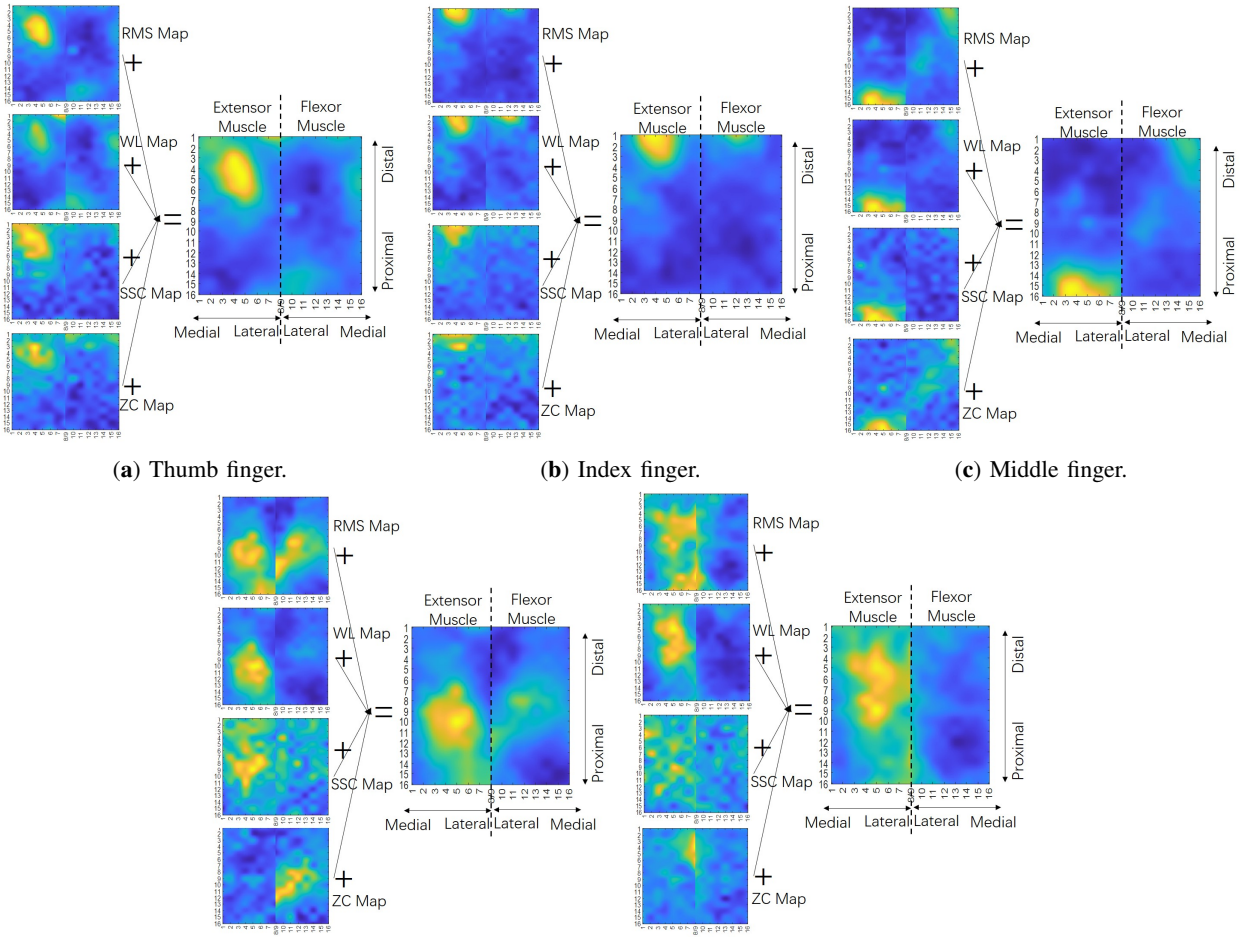
Fig. 5: Performance degradation with increasing levels of signal distortion. CNN models were with 2 convolutional layers.

E. Explainability of Deep Forest via MDI

The electrode importance maps for all fingers were presented in Fig. 6. According to Fig. 6, for each finger, the electrode importance maps of RMS, WL, SSC and ZC features shared a relatively similar pattern. The global important electrodes clustered together within a local area. When estimating the forces of different fingers, deep forest focused on electrodes distributed in different locations of subjects' forearm.

Quantification results on the locations of important electrodes were presented in Fig. 7. When comparing Fig. 7a with Fig. 7b, we can intuitively find that the location difference of important electrodes was more obvious in proximal–distal direction (Fig. 7b). Statistical analyses also support this claim. Specifically, the locations of important electrodes of each pair of fingers showed no significant difference in the medial–lateral direction. By contrast, significant differences between locations of important electrodes in the proximal–distal direction were found in comparisons between thumb-middle fingers ($p = 1.9 \times 10^{-5}$), thumb-ring fingers ($p = 3.9 \times 10^{-5}$), index-middle fingers ($p = 3.8 \times 10^{-5}$), index-ring fingers ($p = 1.9 \times 10^{-5}$), middle-ring fingers ($p = 0.0085$), middle-little fingers ($p = 3.8 \times 10^{-5}$) and ring-little fingers ($p = 2.2 \times 10^{-5}$). The important electrodes of middle, ring, little, thumb and index fingers were successively distributed along the direction from the proximal end to the distal end of the forearm. These findings are in line with the physiological basics of spatial activation locations of motor units in forearm, which will be discussed in the following section.

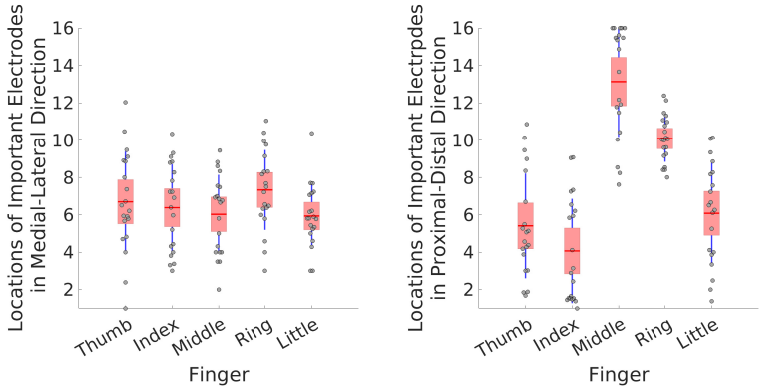
According to Fig. 6, the important electrodes were mainly located at extensor muscles. We then evaluated the finger force estimation performance of models using 128 electrodes from only extensor or flexor muscles separately, with results presented in Table V. Using electrodes from only extensor muscles contributed to a significantly better performance compared with the counterpart of flexor muscles ($p < 0.01$). The results of comparison between EMG amplitudes in extensor and flexor muscles when subjects performing extension and flexion forces were presented in Fig. 8. According to Fig. 8, the amplitudes of EMG from extensor muscles are constantly high during finger extension and flexion, while the amplitudes of EMG from flexor muscles are relatively low during finger extension. Statistical analysis showed that the amplitudes of EMG from flexor muscles under finger extension ("flexor-extension" in Fig. 8) are significantly lower ($p < 0.01$) than the other three cases (the other three cases did not show significant differences with each other). This finding may contribute to future human-machine interfaces concurrently estimating extension and flexion forces using fewer electrodes



(**a**) Thumb finger. (**b**) Index finger. (**c**) Middle finger.

(**d**) Ring finger. (**e**) Little finger.

Fig. 6: Electrode importance maps in estimating forces of five fingers. Note that all maps are presenting importance values instead of feature values. Bright pixels refer to important electrodes. The original 16×16 importance maps were up-sampled to 160×160 via bicubic interpolation for better visualization.



(**a**) Locations of important electrodes in medial–lateral direction. (**b**) Locations of important electrodes in proximal–distal direction.

Fig. 7: Quantification results of electrode importance maps.

from only one muscle (extensor).

V. Discussion

In this work, we proposed a deep forest model for HDsEMG-based finger force estimation. Compared with other benchmark models, deep forest models contributed to a lower error and a higher robustness against noises. The success of DNN models has proved the power of layer-by-layer information processing mechanism, which has been believed as a crucial factor for

TABLE V: Comparison between deep forest models using electrodes from only extensor or flexor muscles (in cross-day validation).

Models	RMSE (%MVC)	COR (%)	R^2 (%)
Electrodes from Extensor Muscles	8.2±2.3	85.6±12.1	57.3±19.3
Electrodes from Flexor Muscles	10.0±2.3	75.5±13.6	40.0±20.1

*Note that in these analyses, we first calculated the number of extension and flexion samples. The number of extension and flexion samples was kelp completely the same by randomly down-sampling the samples with higher sample number. Samples were defined as extension or flexion samples if the force values were >3% MVC or <-3% MVC, respectively. In this way, the comparison between electrodes from extensor and flexor muscles would not be biased by different sample numbers.

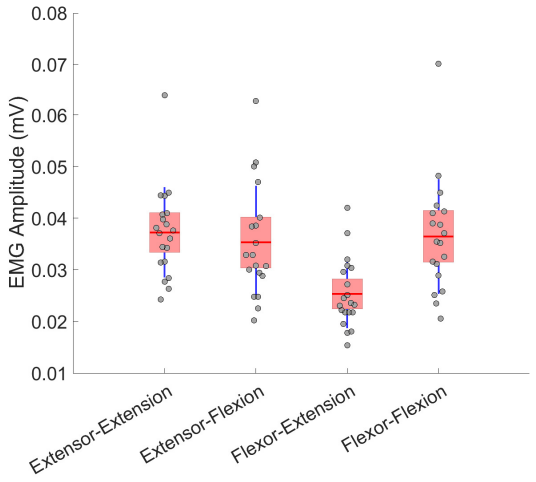


Fig. 8: EMG amplitudes in extensor and flexor muscles corresponding to extension and flexion samples. A total of four muscle-direction pairs were compared. For example, "Extensor-Flexion" refers to the amplitudes of EMG collected from the extensor muscle side of forearm during finger flexion.

the high ability of representation learning for deep learning techniques [21]. Highly abstract and informative features gradually emerge when the models go deeper. In our work, we provide another evidence that such layer-by-layer processing mechanism is also powerful for non-differentiable module (forest-based module) in the context of EMG-based finger force estimation.

Most DNN models are known to be black-box models that lack explainability. However, such DNN (e.g. CNN) models have achieved a remarkable success in many areas, especially in computer vision applications [39]. A series of variants of DNN and CNN have also showed excellent performance [40], [41]. Therefore, researchers all over the world have proposed diverse useful tools to provide explanations on the decision making process of deep learning models, to enhance users' trust on such powerful models in real world applications. For example, Li et al. proposed a nonnegative matrix factorization (NMF)-based algorithm to learn image representations with physical meaning and theoretical interpretations [42] and a weakly supervised deep matrix factorization algorithm via a DNN architecture to enhance social image understanding [43]. Gu et al. recently proposed a comprehensive attention-based CNN to provide explanations in medical image analyses [44]. In the area of EMG decoding, a very recent study [45] proposed an encoder-decoder DNN with an attention mechanism which is explainable by analyzing the attention matrix. The above previous studies made tremendous contributions to open the black-box of the decision making process in diverse areas. However, the ultimate goal of explainable machine learning is to provide a completely transparent decision making process and maintain an excellent model performance at the same time. According to [16], existing recognized interpretable models mainly include decision tree, rules, and linear models [46]–[48]. Specifically, decision tree-based models are inherently explainable because the flowchart-like structure of each decision tree clearly explains which feature was taken into consideration to make the current decision in a branch. Forest models built on multiple decision trees are also completely transparent. The frequency of occurrence of different features in different paths, together with their contributions to the information gain provide the global importance of different features. Moreover, by analyzing the specific decision path each example passes, the decision making logic for a single example can also be analyzed. The decision path from the root node to a leaf node clearly demonstrates the logic in making the final decision.

Moreover, in our work, we quantified the importance map of all electrodes for each finger, and found the locations of important electrodes are highly consistent with the spatial activation areas of motor neurons in forearm muscles reported in a previous study [49]. In our explainability analyses, we found that spatial distributions of important electrodes for different fingers showed no significant differences in medial–lateral direction, but was highly distinguishable in proximal–distal direction. These findings are in line with the conclusions of the previous study [49] that the activation locations of motor units related to different fingers were more distinguishable in proximal-distal direction. The previous study [49] also reported that, the activation locations of motor units related to middle, ring, little, and index fingers were distributed in order along the direction from the proximal end to the distal end of forearm (the thumb finger was not investigated in the previous study [49]), with motor units related to the middle finger activated the most proximal end and motor units related to the index finger activated the most distal end of the forearm. The physiological basis of motor unit activation is highly consistent with the locations of important electrodes for different fingers (as presented in Fig. 7b). All these results showed that, the importance maps of electrodes for different fingers obtained via explainability analyses demonstrated the underlying physiological basis. Overall, when estimating the forces of different fingers, deep forest models focused more on those electrodes located at the activation areas of motor units which are highly relevant to each finger.

Our results also demonstrated that, EMG signals from extensor muscles contain useful information (high amplitudes) during both finger extension and finger flexion. By contrast, EMG signals from flexor muscles are only relevant to finger flexion. Using EMG from extensor muscles alone achieved significantly better performance compared with flexor muscles. When applying finger force estimation models in

real world human-machine interface applications with limited computational or electrode resources, more electrodes should be put on the side of extensor muscle. More specifically, the electrode positions can be further optimized by the explanation results presented in Fig.6.

The high robustness of deep forest against channel corruption may be due to the fact that, the module in each layer is built upon the ensemble of decision trees. Each corrupted channel only influences a small proportion of decision trees instead of the whole model. With a large number of decision trees in all layers to give the final decision together, the negative biased influence of corrupted channels was greatly reduced by the statistical power of all decision trees. By contrast, other benchmark models performed worse with respect to the robustness against noises, because each corrupted channel influences the optimal estimation of all other parameters which are dependent with each other, increasing the destructive power of noises on model performance.

Improving the robustness of EMG-force models against noises is a longstanding challenge, especially with HDsEMG. Previous studies mainly focused on designing a descriptor to measure the quality of acquired EMG signals and then removing or smoothing the corrupted channels [17], [19], [50]–[52]. In this work, we provide an alternative perspective to address the issue of channels corrupted by noises. By using deep forest models, the obtained model has inherently gained the robustness against noise during model training due to the decision tree ensemble architecture in each layer.

Human-machine interfacing has been an important research direction for decades. However, several main problems limit the applications of current models. First, most currently applied proportional force control models focus on estimation of grip force [53]. Proportional estimation of finger forces might be one of the most challenging tasks in EMG-force modeling because the control of fingers can be viewed as most dexterous control of the human body, which contributes to hand dexterity during manipulation of objects in our daily life. Studies on proportional estimation of finger forces are relatively few compared with grip forces. Even fewer studies have ever validated the performance of finger force estimation models in a more realistic cross-day validation protocol. Models trained and tested on the same day can be viewed as single-use models. However, in practical use, such sEMG-force models are required to be highly robust when tested on a second day. In this case, the risks of model overfitting to training data may lead to a greatly degraded model performance in cross-day validation because training and testing data acquired on separate days normally show largely different data (feature) distributions. In this work, we validated our deep forest models on finger force estimation tasks in a cross-day validation, taking realistic issues into consideration. The deep forest models have been proved to be more powerful than currently widely applied models in such realistic and rigorous validation protocol.

VI. Conclusions

In this work, we proposed a deep forest model for the estimation of finger forces from HDsEMG measurements. The complexity of deep forest model can be determined automatically during model training, minimizing the efforts in tuning hyperparameters. We evaluated the excellent performance and many interesting properties of deep forest models in finger-force estimation tasks under a realistic and rigorous cross-day validation without any model calibration on the second day. Through our validations and above discussions, we demonstrated that deep forest is a promising new solution for EMG-force modeling with inherent explainability, robustness against sensor noises, and better cross-day generalization ability, outperforming other benchmark models. Our work can broaden the horizons of methodologies in human-machine interfacing fields as well as more general deep learning related applications.

References

[1] Y. Chen *et al.*, "New Channel Merging Methods for Multi-DoF Force Prediction of Finger Contractions," in *2019 IEEE Biomedical Circuits and Systems Conference (BioCAS)*. Nara, Japan: IEEE, Oct. 2019, pp. 1–4.

[2] C. Dai *et al.*, "Comparison of Constant-Posture Force-Varying EMG-Force Dynamic Models About the Elbow," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 25, no. 9, pp. 1529–1538, Sep. 2017.

[3] G. L. Cerone *et al.*, "A modular, smart, and wearable system for high density semg detection," *IEEE Trans. Biomed. Eng.*, vol. 66, no. 12, pp. 3371–3380, 2019.

[4] Y. Guo *et al.*, "A review of wearable and unobtrusive sensing technologies for chronic disease management," *Comput. Biol. Med.*, vol. 129, p. 104163, 2021.

[5] D. Staudenmann *et al.*, "Improving EMG-based muscle force estimation by using a high-density EMG grid and principal component analysis," *IEEE Trans. Biomed. Eng.*, vol. 53, no. 4, pp. 712–719, Apr. 2006.

[6] L. Meng *et al.*, "User-tailored hand gesture recognition system for wearable prosthesis and armband based on surface electromyogram," *IEEE Trans. Instrum. Meas.*, vol. 71, pp. 1–16, 2022.

[7] L. Meng *et al.*, "Evaluation of decomposition parameters for high-density surface electromyogram using fast independent component analysis algorithm," *Biomed. Signal Process. Control*, vol. 75, p. 103615, 2022.

[8] X. Jiang *et al.*, "High-density surface electromyogram-based biometrics for personal identification," in *2020 42nd Annual International Conference of the IEEE Engineering in Medicine Biology Society (EMBC)*, 2020, pp. 728–731.

[9] X. Jiang *et al.*, "Data Management for Transfer Learning Approaches to Elbow EMG-Torque Modeling," *IEEE Trans. Biomed. Eng.*, vol. 68, no. 8, pp. 2592–2601, 2021.

[10] Y. LeCun *et al.*, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, May 2015.

[11] D. Yang and H. Liu, "An EMG-based Deep Learning Approach for Multi-DOF Wrist Movement Decoding," *IRE Trans. Ind. Electron.*, pp. 1–1, 2021.

[12] Y. Chen *et al.*, "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.

[13] A. Ameri *et al.*, "Regression convolutional neural network for improved simultaneous EMG control," *J. Neural Eng.*, vol. 16, no. 3, p. 036015, Apr. 2019.

[14] C. Ma *et al.*, "A Novel and Efficient Feature Extraction Method for Deep Learning Based Continuous Estimation," *IEEE Robot. Autom. Lett.*, vol. 6, no. 4, pp. 7341–7348, 2021.

[15] D. E. Rumelhart *et al.*, "Learning representations by back-propagating errors," *Nature*, vol. 323, no. 6088, pp. 533–536, Oct. 1986.

[16] R. Guidotti *et al.*, "A survey of methods for explaining black box models," *ACM Comput. Surv.*, vol. 51, no. 5, aug 2018.

[17] X. Jiang *et al.*, "Optimization of hd-semg-based cross-day hand gesture classification by optimal feature extraction and data augmentation," *IEEE Trans. Hum.-Mach. Syst.*, pp. 1–11, 2022.

[18] X. Jiang *et al.*, "Cancelable HD-sEMG-Based Biometrics for Cross-Application Discrepant Personal Identification," *IEEE J. Biomed. Health Inform.*, vol. 25, no. 4, pp. 1070–1079, Apr. 2021.

[19] M. Stachaczyk *et al.*, "Adaptive Spatial Filtering of High-Density EMG for Reducing the Influence of Noise and Artefacts in Myoelectric Control," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 28, no. 7, pp. 1511–1517, Jul. 2020.

[20] P. Ghaderi and H. R. Marateb, "Muscle activity map reconstruction from high density surface emg signals with missing channels using image inpainting and surface reconstruction methods," *IEEE Trans. Biomed. Eng.*, vol. 64, no. 7, pp. 1513–1523, 2017.

[21] Z.-H. Zhou and J. Feng, "Deep forest," *Natl. Sci. Rev.*, vol. 6, no. 1, pp. 74–86, Jan. 2019.

[22] E. Scornet, "Trees, forests, and impurity-based variable importance," *arXiv*, Jan. 2020.

[23] X. Jiang *et al.*, "Open access dataset, toolbox and benchmark processing results of high-density surface electromyogram recordings," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 29, pp. 1035–1046, 2021.

[24] R. N. Khushaba and K. Nazarpour, "Decoding HD-EMG Signals for Myoelectric Control - How Small Can the Analysis Window Size be?" *IEEE Robot. Autom. Lett.*, vol. 6, no. 4, pp. 8569–8574, 2021.

[25] X. Jiang *et al.*, "Measuring neuromuscular electrophysiological activities to decode hd-semg biometrics for cross-application discrepant personal identification with unknown identities," *IEEE Trans. Instrum. Meas.*, vol. 71, pp. 1–15, 2022.

[26] J. C. A. Barata and M. S. Hussein, "The Moore–Penrose Pseudoinverse: A Tutorial Review of the Theory," *Braz. J. Phys.*, vol. 42, no. 1-2, pp. 146–165, Apr. 2012.

[27] X. Jiang *et al.*, "Random channel masks for regularization of least squares-based finger emg-force modeling to improve cross-day performance," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 30, pp. 2157–2167, 2022.

[28] A. Krasoulis *et al.*, "Effect of User Practice on Prosthetic Finger Control With an Intuitive Myoelectric Decoder," *Frontiers in Neuroscience*, vol. 13, 2019.

[29] D. Blana *et al.*, "Model-Based Control of Individual Finger Movements for Prosthetic Hand Function," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 28, no. 3, pp. 612–620, Mar. 2020.

[30] H.-J. Hwang *et al.*, "Channel selection for simultaneous and proportional myoelectric prosthesis control of multiple degrees-of-freedom," *J. Neural Eng.*, vol. 11, no. 5, p. 056008, aug 2014.

[31] E. S. Altinkaynak and D. J. Braun, "A phase-invariant linear torque-angle-velocity relation hidden in human walking data," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 27, no. 4, pp. 702–711, 2019.

[32] A. Ameri *et al.*, "Support vector regression for improved real-time, simultaneous myoelectric control," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 22, no. 6, pp. 1198–1209, 2014.

[33] M. Sierotowicz *et al.*, "Emg-driven machine learning control of a soft glove for grasping assistance and rehabilitation," *IEEE Robot. Autom. Lett.*, vol. 7, no. 2, pp. 1566–1573, 2022.

[34] M. Barsotti *et al.*, "Online finger control using high-density emg and minimal training data for robotic applications," *IEEE Robot. Autom. Lett.*, vol. 4, no. 2, pp. 217–223, 2019.

[35] M. Tipping, "Sparse bayesian learning and the relevance vector machine," *J. Mach. Learn. Res.*, vol. 1, no. 3, pp. 211–244, SUM 2001.

[36] D. P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," in *ICLR*, Jan. 2015.

[37] E. Huigen *et al.*, "Investigation into the origin of the noise of surface electrodes," *Med. Biol. Eng. Comput.*, vol. 40, no. 3, pp. 332–338, May 2002.

[38] M. Chimene and R. Pallas-Areny, "A comprehensive model for power line interference in biopotential measurements," *IEEE Trans. Instrum. Meas.*, vol. 49, no. 3, pp. 535–540, Jun. 2000.

[39] A. Krizhevsky *et al.*, "Imagenet classification with deep convolutional neural networks," *Commun. ACM*, vol. 60, no. 6, p. 84–90, may 2017.

[40] Z. Peng *et al.*, "Few-shot image recognition with knowledge transfer," in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, October 2019.

[41] Z. Liu *et al.*, "A convnet for the 2020s," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2022, pp. 11 976–11 986.

[42] Z. Li *et al.*, "Robust structured nonnegative matrix factorization for image representation," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 5, pp. 1947–1960, 2018.

[43] Z. Li and J. Tang, "Weakly supervised deep matrix factorization for social image understanding," *IEEE Trans. Image Process.*, vol. 26, no. 1, pp. 276–288, 2017.

[44] R. Gu *et al.*, "Ca-net: Comprehensive attention convolutional neural networks for explainable medical image segmentation," *IEEE Trans. Med. Imaging*, vol. 40, no. 2, pp. 699–711, 2021.

[45] H. Lee *et al.*, "Explainable deep learning model for emg-based finger angle estimation using attention," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 30, pp. 1877–1886, 2022.

[46] A. A. Freitas, "Comprehensible classification models: A position paper," *SIGKDD Explor. Newsl.*, vol. 15, no. 1, p. 1–10, mar 2014.

[47] J. Huysmans *et al.*, "An empirical evaluation of the comprehensibility of decision table, tree and rule based predictive models," *Decis. Support Syst.*, vol. 51, no. 1, pp. 141–154, 2011.

[48] M. T. Ribeiro *et al.*, ""why should i trust you?": Explaining the predictions of any classifier," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ser. KDD '16. New York, NY, USA: Association for Computing Machinery, 2016, p. 1135–1144.

[49] X. Jiang *et al.*, "Quantifying spatial activation patterns of motor units in finger extensor muscles," *IEEE J. Biomed. Health Inform.*, vol. 25, no. 3, pp. 647–655, 2021.

[50] H. R. Marateb *et al.*, "Outlier detection in high-density surface electromyographic signals," *Med. Biol. Eng. Comput.*, vol. 50, no. 1, pp. 79–89, Jan. 2012.

[51] C. Grönlund *et al.*, "On-line signal quality estimation of multichannel surface electromyograms," *Med. Biol. Eng. Comput.*, vol. 43, no. 3, pp. 357–364, Jun. 2005.

[52] P. McCool *et al.*, "Identification of Contaminant Type in Surface Electromyography (EMG) Signals," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 22, no. 4, pp. 774–783, Jul. 2014.

[53] R. Barański and A. Kozupa, "Hand Grip-EMG Muscle Response," *Acta Physica Polonica, A.*, vol. 125, no. 4A, pp. A–7, Apr. 2014.