IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received October 4, 2020, accepted October 22, 2020, date of publication November 3, 2020, date of current version November 17, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3035498

Prediction of Voluntary Motion Using Decomposition-and-Ensemble Framework With Deep Neural Networks

DOKYOON YOON[1,2], EUNCHAN KIM[1,3], INGU CHOI[1], SUNG WON HAN[2], AND SUNGWOOK YANG[1], (Member, IEEE)

[1]Center for Intelligent and Interactive Robotics, Korea Institute of Science and Technology, Seoul 02729, South Korea
[2]School of Industrial Management Engineering, Korea University, Seoul 02841, South Korea
[3]Division of Mechanical Engineering, Hanyang University, Seoul 04763, South Korea

Corresponding authors: Sung Won Han (swhan@korea.ac.kr) and Sungwook Yang (swyang@kist.re.kr)

This work was supported in part by the Bio and Medical Technology Development Program of the National Research Foundation (NRF) funded by the Ministry of Science and ICT under Grant NRF-2019M3A9E2061784, and in part by the Convergence Technology Development Program for Bionic Arm through the NRF funded by Ministry of Science, ICT and Future Planning (MSIP) under Grant 2014M3C1B2048419.

ABSTRACT It is essential for seamlessly delivering intended hand motion to surgical robots while actively suppressing undesired hand tremor during microsurgery. To achieve this goal, we propose a novel method for predicting voluntary motion based on deep learning with the signal decomposition and ensemble approach. This approach can thus deal with various forms of voluntary signals, such as either highly stationary or rather highly cyclic at a low range of frequencies. The proposed method comprises a series of signal blocks to decompose complex hand motion into multiple sub-signals using deep neural networks upon their signal characteristics. The signal block yields parameterized sub-signal and predicted voluntary motion. In addition, an ensemble layer allows for accurately predicting future voluntary motion by combining predicted motion from each signal block with the optimal weight. These signal blocks are connected by a decomposition flow in series and also by a forecast flow in parallel to ensemble the prediction output of each block. Given real data sets, we evaluated the prediction performance of the proposed algorithm compared to other data-driven deep learning models. The generalizability of the proposed algorithm was also investigated by applying the trained models to new data sets from new tasks and a different subject, which had not been involved in training procedures. As a result, the proposed algorithm outperforms the other baseline models in terms of prediction error and accuracy. Furthermore, we explored whether the proposed method could suppress tremor via spectral analysis, which shows substantial tremor attenuation more than -10 dB in a frequency of interest, 6—14 Hz. It is also found that the proposed method has the predictive power for dealing with inevitable control delays, where prediction error increased only by 2% per one-time sample, approximately 4.2 ms.

INDEX TERMS Ensemble deep learning, signal decomposition, motion prediction, hand tremor, active tremor compensation, surgical robots.

I. INTRODUCTION

Analyzing and interpreting human hand motion have attracted great interests from a variety of research areas, such as prosthesis [1]–[3], rehabilitation [4], and surgical robotics [5]–[10]. For example, the estimation of unintended hand motion, such as pathological tremor, is required to analyze the degree of patients' disorders or to suppress the tremor through prosthetic devices. The pathological tremor affecting everyday life, such as Parkinson's disease, is involuntary and pseudo-rhythmic movement from several age-related neurological disorders, which occurs in a broad frequency range of 3—14 Hz [11]. On the other hand, identifying voluntary motion and physiological tremor inherent from normal hand motion is a focus of research in surgical robotics. The voluntary motion in microsurgical operation is described as signals in frequencies below 2 Hz. In contrast, physiological tremor is originated from the combination of mechanical-reflex components and oscillation in a central nervous system [12], which is known to have an RMS (root-mean-square) amplitude on the order of 50-200 μm at a frequency commonly in the 6-14 Hz band [13].

The associate editor coordinating the review of this manuscript and approving it for publication was Shagufta Henna.

Specifically, teleoperated surgical robots, well-known as Da Vinci Surgical System (Intuitive Surgical, USA) [6], collect hand motion data from a master device, then deliver filtered and/or scaled-down motion to slave manipulators. Alternatively, handheld robots, known as Micron [9], [10], iTrem [14], first sense their own motion of the robots and then selectively filter out erroneous motion such as hand tremor. For such active tremor cancellation, the handheld robots use control signals as the counter-motion of tremor signal or the estimated voluntary motion at the end of the tool tip. Since the cancelation of hand tremor needs to be immediately done as soon as sensing its own motion, no latency in a control loop is preferred. However, inevitable delays in controlling the robots would degrade the performance of active tremor cancellation because the estimated motion at present becomes out of date at the instance of canceling. Therefore, the prediction of hand motion is required to seamlessly accomplish active tremor cancellation with a high level of accuracy, as compensating for the time delay in control.

Despite the need for hand motion prediction in active tremor compensation, no algorithm generalizable to various users and operation in microsurgery has been introduced yet. A common approach introduced in the literature is to estimate hand tremor via online learning. This approach regards spectro-temporal hand tremor as the combination of sinusoidal motions. The learning model either adaptively updates the time-varying frequencies [15] and amplitudes or update the amplitudes of a fixed set of multiple frequencies [16]. However, the proposed algorithm is prone to failure in reaching convergence because it is sensitive to initial parameters for searching the frequencies and amplitudes. A data-driven approach has recently been proposed to address the issues raised by the stringent modeling of hand tremor [17], [18]. Although such deep-learning-based methods offer a certain level of generalized performance, enormous training procedures and high computational power hinder the application of the algorithms to real-time active tremor cancellation.

To overcome those problems in estimation and prediction of hand motion, we propose a novel method for voluntary motion prediction with deep neural networks based on decomposition and ensemble learning. In the proposed model, we take advantage of signal decomposition that separates a raw signal into voluntary and tremor signals while mitigating the time delay, which is inherently yielded in a real-time lowpass filter. Hence, time-series data of hand motion is decomposed by a set of deep neural networks representing either voluntary motion or hand tremor. The prediction of future voluntary motion is then accomplished by the ensemble of prediction outputs from the decomposed signal blocks, as represented in Fig. 1. Given this end-to-end learning model, we evaluate the accuracy of the voluntary motion prediction on real data sets while comparing it with other data-driven machine learning algorithms. The generalized performance of the proposed algorithm is also investigated by applying the model to new tasks and different subjects. Finally, we explore how efficiently the proposed algorithm could suppress physiological hand tremor while maintaining the voluntary motion for active tremor cancellation.

II. RELATED WORK

The proposed model for voluntary motion prediction is closely related to two research topics: (II.A) hand tremor estimation and (II.B) decomposition and ensemble learning.

A. HAND TREMOR ESTIMATION

To accurately and effectively estimate hand tremor, various machine learning techniques have been introduced to this problem. The weighted-frequency Fourier linear combiner (WFLC) utilizes a multilayer perceptron (MLP) to learn time-varying frequencies and amplitudes of hand tremor [15]. However, the WFLC algorithm may suffer from distinguishing adjacent frequencies and assigning appropriate parameters for convergence [16]. The band-limited multiple Fourier linear combiner (BMFLC) was also proposed to overcome these issues by adapting linear combinations of harmonic signals of a fixed set of multiple frequencies [16]. The coefficients of the linear combiner can be found by a stochastic gradient descents algorithm or the Kalman filter, which offers an optimal solution to continuous Markov chain models assuming linear Gaussian noise models. However, computational load rises as more frequencies are combined to improve the accuracy of estimation. A support vector machine (SVM) was also adopted to estimate hand tremor using nonlinear kernels [19]. Although it outperforms the MLP algorithms, the heavy load in computation limits its application to real-time operation. Furthermore, these algorithms that estimate hand tremor primarily at the present time step are prone to failure in predicting future tremor because of the non-stationary nature of hand tremor. To overcome such the limitation, tremor prediction models have also been proposed, including autoregressive (AR) and autoregressive-moving-average (ARMA) models [20]. Tremor prediction methods based on least square-support vector machine (LS-SVM) [19], extreme learning machine [21] were also introduced. Recently, Shahtalebi *et al*. proposed a deep learning-based methodology to both estimate and predict pathological hand tremor [17]. However, the application of the proposed model was still less generalizable to various scenarios of motion. To address the limitation, a generalizable model adopting a deep recurrent model trained with a sizeable dataset was also proposed [18]. However, the proposed model entails a high computational load led by a complex learning architecture and enormous training.

B. DECOMPOSITION AND ENSEMBLE LEARNING

A hybrid approach embedding appropriate preprocessing models in deep learning architectures can improve learning performance because it eases to find learnable features [22]. Decomposition-and-ensemble is one of such hybrid models, of which main ideas are as follows. Firstly, it decomposes a raw signal into sub-signals with a decomposition method.

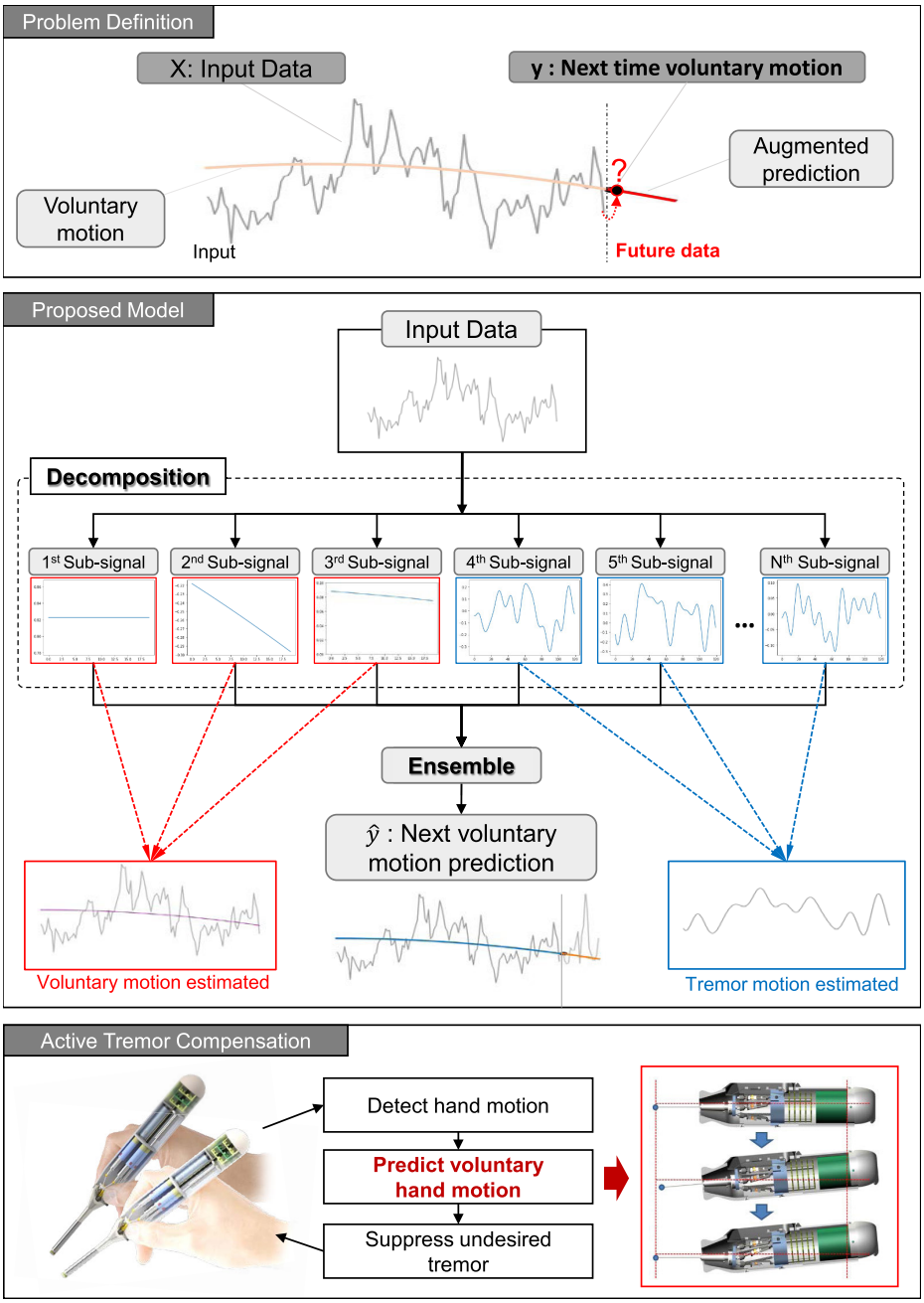


FIGURE 1. **Overview of predicting voluntary motion for active tremor compensation using deep learning model based on the signal decomposition and ensemble approach.**

Each sub-signal block then outputs a target signal to be predicted, respectively. Finally, it ensembles the multiple predictions from each processing block in order to obtain a final output.

A Fourier transform is one of the common decomposition algorithms used to analyze hand motion in time-series [23], [24]. It decomposes a raw signal into sinusoidal sub-signals corresponding to frequencies of interest. Empirical mode decomposition (EMD) is another decomposition method suitable for forecasting of timeseries data [25], which is a part of Hilbert-Huang transform (HHT). EMD also used in tremor decomposition to overcome limitations yielded by the prior assumption of Fourier transform-based methods: linearity and stationarity [26]. Although we can decompose signals using existing methods, the decomposed signals are not optimized for signal prediction. Recently, Wang *et al.* introduced a neural network-based decomposition adopting a wavelet decomposition network (WDN) [27]. It can thus decompose timeseries data into a group of sub-signals in the form of frequencies, which is crucial for taking into account frequency factors in prediction. Consequently, neural network-based decomposition can take advantage of both timeseries decomposition and the learning ability of neural networks.

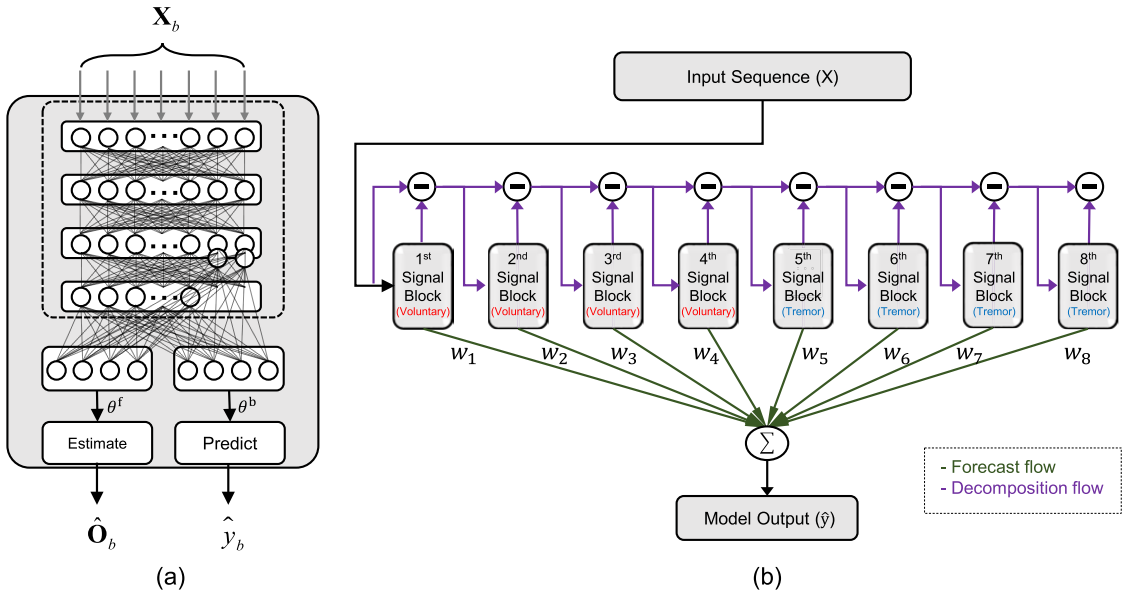


FIGURE 2. The architecture of deep neural networks with signal decomposition and ensemble for voluntary motion prediction: (a) signal block and (b) entire model with sub-signal blocks and ensemble.

III. PROPOSED METHOD

A. OVERALL ARCHITECTURE

We propose a deep learning model incorporating neural network structures that decompose complex signals into sub-signals and ensemble them to accurately and efficiently predict future voluntary motion. The overall structure of the model comprises sequential signal blocks corresponding to decomposed sub-signals. Each signal block estimates sub-signal and predicts future voluntary motion. The signal blocks used in our model can be classified as a voluntary signal block and a tremor signal block, depending on the characteristics of signals to be estimated. Signal blocks are connected by a decomposition flow in series and also by a forecast flow in parallel to ensemble the prediction output of each block. The overall architecture is represented in Fig. 2. The algorithm details are as follows.

B. SIGNAL BLOCK

Each signal block is responsible for estimating a specific model of sub-signal and predicting voluntary motion that would occur the next time step. For a specific size of historical time window from the present time, a block input is described by $\mathbf{X}_b$, and the dual outputs of the block are by $\hat{y}_b$ and $\hat{\mathbf{O}}_b$ from the bth block. We define the input $\mathbf{X}_b$ as in (1).

$$\widehat{\mathbf{X}}_b = \left[s_{b,t_{i-F_s/2+1}}, \quad \cdots, \quad s_{b,t_i} \right], \tag{1}$$

where s_{b,t_i} is the input signal of the bth block at the ith timestamp t_i, and the size of historical window is set by 0.5 s in our model. The output $\hat{y}_b$ represents the prediction of voluntary motion at the next time step t_{i+1}. In addition, another output $\hat{\mathbf{O}}_b$ is sub-signal subject to decomposition from the bth signal block. The input to the next block $\mathbf{X}_{b+1}$ is given by subtracting $\hat{\mathbf{O}}_b$ from $\mathbf{X}_b$ and it is repeated until the end of the signal block. The details of the decomposition flow are described in III.C.

Depending on the type of sub-signal subject to decomposition, the signal block is modeled differently while taking its signal characteristics into account. Herein, we introduce two types of the signal blocks to decompose raw signal into voluntary motion and hand tremor. For the voluntary signal block, we approximate the signal as a low-order polynomial with a small degree of P since voluntary motion appears as fairly monotonic or cyclic with frequencies below 2 Hz within the specified time window. The predicted voluntary signal is then obtained by extrapolation of the polynomial as in (2).

$$\hat{y}_b = \sum_{p=0}^{P} \theta_{b,p}^F (t_{i+1} - t_i)^p, \tag{2}$$

where $\theta_{b,p}^F$ is the pth coefficient of the polynomial for forecast. The output of the signal block $\hat{\mathbf{O}}_b$, which describes the voluntary motion in the given time window, is also obtained by a parameterized function in (3) as we assume the predicted voluntary motion as a low-order polynomial.

$$o_{b,k} = \sum_{p=0}^{P} \theta_{b,p}^D (t_k - t_i)^p, \tag{3}$$

where $\theta_{b,p}^D$ is the pth coefficient of the polynomial for decomposition and $k = [i - F_s/2 + 1, \ldots, i]$ with sampling frequency F_s. The output $\hat{\mathbf{O}}_b$ is then constructed as in (4) using (3).

$$\hat{\mathbf{O}}_b = \left[o_{b,t_{i-F_s/2+1}}, \quad \cdots, \quad o_{b,t_i} \right] \tag{4}$$

On the other hand, the outputs of the tremor signal block $\hat{y}_b$ and $\hat{\mathbf{O}}_b$ are approximated by the sum of sinusoidal signals while considering its rhythmical and oscillatory characteristics. Accordingly, the output $\hat{y}_b$ from the tremor block is

described by the Fourier series as in (5).

$$\widehat{y}_b = \sum_{m=0}^{N-1} \left[\theta^F_{b,2m} \cos\left(2\pi f_m t_{i+1}\right) + \theta^F_{b,2m+1} \sin\left(2\pi f_m t_{i+1}\right) \right], \tag{5}$$

where $\theta^F_{b,2m}$ and $\theta^F_{b,2m+1}$ are Fourier coefficients for a set of multiple frequencies, $f_m = \left[1 \quad 2, \quad \cdots, \quad F_s/2\right]$. $\hat{\mathbf{O}}_b$ corresponding to the decomposed tremor signal is then constructed using (4) and (6) in the same manner.

$$o_{b,k} = \sum_{m=0}^{N-1} \left[\theta^D_{b,2m} \cos\left(2\pi f_m t_k\right) + \theta^D_{b,2m+1} \sin\left(2\pi f_m t_k\right) \right], \tag{6}$$

where $\theta^D_{b,2m}$ and $\theta^D_{b,2m+1}$ are Fourier coefficients to describe decomposed tremor signal. We obtain those parameters from each signal block using a neural network with four hidden layers. Each layer has an activation function for nonlinear mapping (a rectified linear unit, ReLU [28]), and the last layer consists of a fully connected layer without an activation function.

C. FORECAST AND DECOMPOSITION FLOWS

The two types of outputs $\hat{y}_b$ and $\hat{\mathbf{O}}_b$ from each signal block are used for a forecast flow and a decomposition flow, respectively.

In the forecast flow, the ensemble of the prediction results from the multiple signal blocks allows accurately predicting voluntary motion at the next time step. The formula for the forecast flow is as follows:

$$\widehat{y} = \sum_{b=1}^{n} w_b \hat{y}_b \tag{7}$$

where n is the number of blocks used and w_b is the weight of prediction outcome from the bth block. Consequently, the final prediction $\hat{y}$ is attained by the weighted sum of prediction results from the multiple signal blocks.

The decomposition flow serves to decompose the entire signal into multiple sub-signals while proceeding in the cascaded model. Each block estimates sub-signal modeled by a specific type of the signal block. The formula in the decomposition flow is as follows:

$$\mathbf{X}_{b+1} = \mathbf{X}_b - \hat{\mathbf{O}}_b \quad \text{and} \tag{8}$$

$$\mathbf{X}_b = \mathbf{X} - \sum_{i=1}^{b-1} \hat{\mathbf{O}}_i, \tag{9}$$

where $\mathbf{X}$ is a raw signal input and $\mathbf{X}_1$ is equal to $\mathbf{X}$ at the first signal block. As noticed in (8), the input of each block is connected to the output of the previous block with residual connection [29]. Therefore, subtracting the output of each block from the given input signal leads the sub-signal found in the current block to be excluded in the next estimation. By proceeding with this process, the entire signal is decomposed into each sub-signal.

IV. EXPERIMENTAL RESULTS

We evaluated the proposed prediction algorithm on real hand motion data by comparing its prediction error and accuracy with those of other machine learning algorithms: 1) artificial neural networks (ANN) [30], 2) recurrent neural network (RNN) [31], 3) long short term memory model (LSTM) [32], and 4) PHTNet that employs bidirectional gated recurrent units (bi-GRU) [18]. Given trained models, the first step was to validate the prediction performance of the proposed algorithm on trained data. Next, we tested the algorithm on new data sets that were not involved during training procedures in order to investigate the generalizability of the prediction algorithm: for new tasks and different subjects.

A. HAND MOTION DATASET

We collected hand motion data using a magnetic tracker (Liberty^{TM} with Micro Sensor 1.8^{TM}, Polhemus Ltd., USA), which provides six-degrees-of-freedom (6-DOF) motion with a micron level of precision at a sampling frequency of 240 Hz; we thus aimed to predict voluntary motion at the next time step, approximately 4.167 ms ahead. To collect hand motion data, which would potentially occur during surgical operation, the sensor was fixed on one end of a hand-piece. In addition, a 23G hypodermic needle was attached to the other end of the hand-piece to mimic a surgical tool. The position of the tool tip was then retrieved by a homogenous transformation given the 6-DOF motion of the sensor.

Two types of tasks were designed for the experiments: point/line and circle tracing. These tasks are regarded as relatively static and dynamic, respectively. For the static task, the subject was instructed to hold the tool tip above a printed target surface during data collection. The task was performed under four different scenarios: blind, bare eye, low magnification (1.5X), and high magnification (6X) using a stereo-microscope (SZX7, Olympus Corp., Japan). For the dynamic task, the subject was instructed to trace various sizes of circles above the printed target surface, where the sizes of the circles were 1, 2, and 10 mm. Each task under the various settings was repeated five times. A single trial was logged for 30 seconds, which resulted in 7080 data sets for training and testing from 7200 samples, regarding a historical time window of 120 samples. Consequently, we collected about 310,000 data sets in total: 9 of 10-tasks sets for training a model and the rest of the data for testing.

B. MODEL TRAINING

We adopted the zero-phase filter of the seventh's order Butterworth with a cut-off frequency of 2 Hz on raw motion signal to provide the ground-truth of predicted voluntary motion. In addition, mean squared error (MSE) between the predicted motion and the ground-truth was used as a loss function for training the proposed model. The model was trained for 1000 epochs using a learning rate of 10^{-3}. To prevent overfitting, the learning rate decayed by 0.5 times if no decrement in the loss was found for five epochs. An early stopping criterion

was invoked to prevent overfitting if the loss would not be decreased for 40 epochs.

We set the total eight signal blocks for the model: the first four blocks for the decomposition of voluntary signal and the rest of four blocks for tremor signal. Two of the four voluntary blocks were parametrized as linear functions for estimating voluntary motion, and the other two blocks were defined as cubic functions. The four tremor blocks were modeled as the combination of cyclic functions. For example, the DC offsets of the input signal were taken from the first two voluntary blocks. Then, the next sub-signals were fitted into cubic functions for the estimation of voluntary signal, while approximating voluntary motion with a frequency of 2 Hz as the third-order polynomial within a time series window of 0.5 s.

Learning procedures were accomplished by applying a backpropagation algorithm to minimize root-mean-square-error (RMSE) in prediction. We also used the adaptive moment estimation (Adam) as an optimizer. We performed the 9-fold cross-validation to test the performance of the proposed model. The learning settings were set for the other baseline models, ANN, RNN, LSTM, and PHTNet, in the same manner.

C. DESCRIPTION OF BASELINE MODELS

In this section, we introduced the four baseline models in detail. The ANN model consists of two hidden layers and a single output layer. The first layer of the ANN is a dense layer with 480 hidden units. In this layer, we used ReLU as an activation function for nonlinear mapping. The first layer is followed by another dense layer with the same number of hidden units and the activation function. The second layer is followed by the output layer that contains a single unit for prediction output without any activation function. The RNN model consists of an RNN layer and an output layer. The RNN layer includes eight hidden units, in which a hyperbolic tangent (*tanh*) function was used as an activation function. The cell output of the RNN layer is connected to the single unit of the output layer without any activation function. Similarly, the LSTM model incorporates an LSTM layer and an output layer. The LSTM layer also consists of eight hidden units with the *tanh* activation function. Finally, the cell output of the LSTM layer is connected to the output layer that has a single unit for prediction output without any activation function. The architecture of the PHTNet comprises the four layers of bi-directional GRU [30] and one output layer. The first layer of the PHTNet is bi-direction GRU (bi-GRU) layer with four hidden units. In this layer, we used ReLU as an activation function. The first layer is followed by another bi-GRU layer with the same setting, which is repeated for the third and fourth bi-GRU layers. The fourth layer is then followed by an output layer that has one unit without any activation function.

D. PERFORMANCE MEASURE

For quantitative analysis of the performance, we measured the root mean square error (RMSE) between actual voluntary motion signal and predicted signal as in (10). In addition, accuracy was also investigated whether an algorithm accurately predicts target signal with respect to the ground truth.

$$RMSE = \sqrt{\frac{1}{T}\sum_{t=1}^{T}(y(t) - \hat{y}(t))^2}, \tag{10}$$

where T is the total length of signal, $y(t)$ is actual voluntary motion, and $\hat{y}(t)$ is the predicted voluntary motion signal at the tth timestamp. Accuracy is defined as below:

$$Accuracy = \frac{RMS(s_x) - RMSE}{RMS(s_x)} \times 100(\%). \tag{11}$$

Herein, s_x is equal to $y(t)$ and $RMS(s_x)$ is the root-mean-square amplitude of the voluntary signal.

E. VOLUNTARY MOTION PREDICTION

We first validated our trained models by comparing them with the other algorithms. For the comparison of prediction performance, we trained two types of models. One was trained from nine task sets out of the total 10 task sets, excluding one of the static tasks: Model I. The other model was trained from nine task sets, excluding one of the dynamic tasks: Model II. As a result, we obtained the smallest RMSEs with our proposed method in both models among the five algorithms. In addition, the highest accuracies were also achieved with our algorithm. Interestingly, PHTNet shows considerably lower performance than those of the proposed method in Model I (static), but the performance was improved in Model II (dynamic), showing a similar level of performance to our algorithm. On the other hand, the LSTM model shows a similar performance to our algorithm with Model I, but the model resulted in the worst performance with Model II. Fig. 3 shows the resulting trajectories of predicted voluntary motion from the five algorithms: ANN, RNN, LSTM, PHTNet, and the proposed method. The RMSE and accuracy are summarized in Table I.

F. GENERALIZED PERFORMANCE

We also evaluated whether the proposed model would be generalizable to data sets with new distributions by testing the prediction model on new tasks and a different subject. The first experiment was to verify generalized performance on the new tasks that had not been involved in the training procedures. As summarized in Table II, the generalized performance of the proposed method is comparable with PHTNet. The resulting trajectories are also shown in Fig. 4. The second test was to see the generalizability of the algorithm to new data from different subjects. In this experiment, the proposed method sill shows its generalizability on the new data set, of which results are similar to those obtained from PHTNet. Fig 5 presents the resulting trajectories in voluntary motion prediction given a trained model by another subject. All results are summarized in Table III. In this experiment, it is noted that the different subject shows a smaller level of hand tremor than the subject participated in model training, which

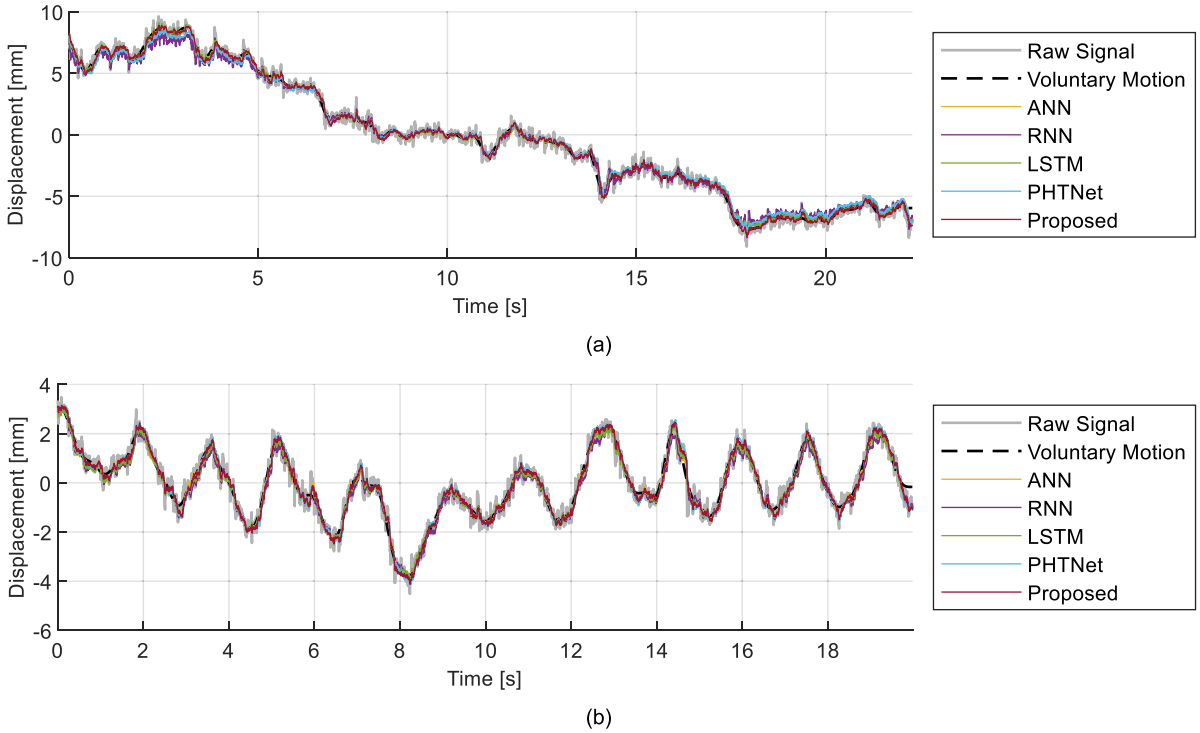


FIGURE 3. Voluntary motion prediction resulted from model validation on static and dynamic tasks with four baseline models: ANN, RNN, LSTM and PHTNet. Raw signal (grey) and voluntary signal (black dashed) are depicted for reference. (a) Static and (b) dynamic tasks.

TABLE 1. Model validation result.

Model	Static		Dynamic	
	RMSE (*mm*)	Accuracy (%)	RMSE (*mm*)	Accuracy (%)
ANN [30]	0.287 (0.084)	74.89 (7.36)	0.370 (0.080)	79.43 (4.43)
RNN [31]	0.263 (0.030)	76.30 (2.72)	0.397 (0.124)	77.90 (6.87)
LSTM [32]	0.219 (0.017)	80.29 (1.52)	0.347 (0.042)	80.71 (2.34)
PHTNet [18]	0.254 (0.038)	77.82 (3.30)	0.283 (0.010)	84.28 (0.57)
Proposed Method	**0.193** (0.016)	**83.14** (11.88)	**0.260** (0.044)	**85.56** (2.43)

The number in parenthesis is standard deviation.

TABLE 2. Model test on new task.

Model	Static		Dynamic	
	RMSE (*mm*)	Accuracy (%)	RMSE (*mm*)	Accuracy (%)
ANN [30]	0.286 (0.056)	87.13 2.54	0.182 (0.020)	84.07 (1.72)
RNN [31]	0.494 (0.313)	77..73 (14.09)	0.238 (0.047)	77.30 (4.53)
LSTM [32]	0.482 (0.617)	78.28 (27.82)	0.187 (0.018)	82.18 (1.73)
PHTNet [18]	0.259 (0.014)	88.33 (0.63)	0.149 (0.007)	86.95 (0.58)
Proposed Method	**0.259** (0.010)	**88.33** 0.45	**0.143** (0.008)	**87.52** (0.73)

The number in parenthesis is standard deviation.

may result in smaller RMSE and higher accuracy overall. Even for various tasks and subjects, the proposed model shows consistent performance overall, whereas the performance of the RNN and the LSTM models varied significantly, depending on the test sets.

G. SPECTRAL ANALYSIS

Our ultimate goal of the voluntary motion prediction is to actively cancel undesired hand tremor while seamlessly maintaining voluntary motion without time delay in control, specifically for robotic microsurgery. Therefore, we examined how much hand tremor would be suppressed by applying the future voluntary motion to robots for control. Hence, we conducted the spectrum analysis on raw signal and predicted voluntary motion using short-time Fourier transform over the time of operation: as an indication of active tremor compensation. As shown in Fig. 6, hand tremor drastically was suppressed within a frequency range of 6—14 Hz over the time of operation. Furthermore, we ran the analysis of power spectrum density on the predicted signals. As a result, it is found that hand tremor is decreased by -10.9 and -13.1 dB for a frequency of interest, 6—14 Hz, while applying voluntary signals predicted in the new task and the different subject to control, respectively.

H. PREDICTABILITY

The proposed model can predict voluntary motion one-time-step ahead as well as further voluntary motion over time horizon because the motion is parameterized by a time sequence. Hence, we explored the prediction capability on voluntary motion multiple-time-steps ahead. The RMSE at each future time step is depicted in Fig. 8. With the advancement of the

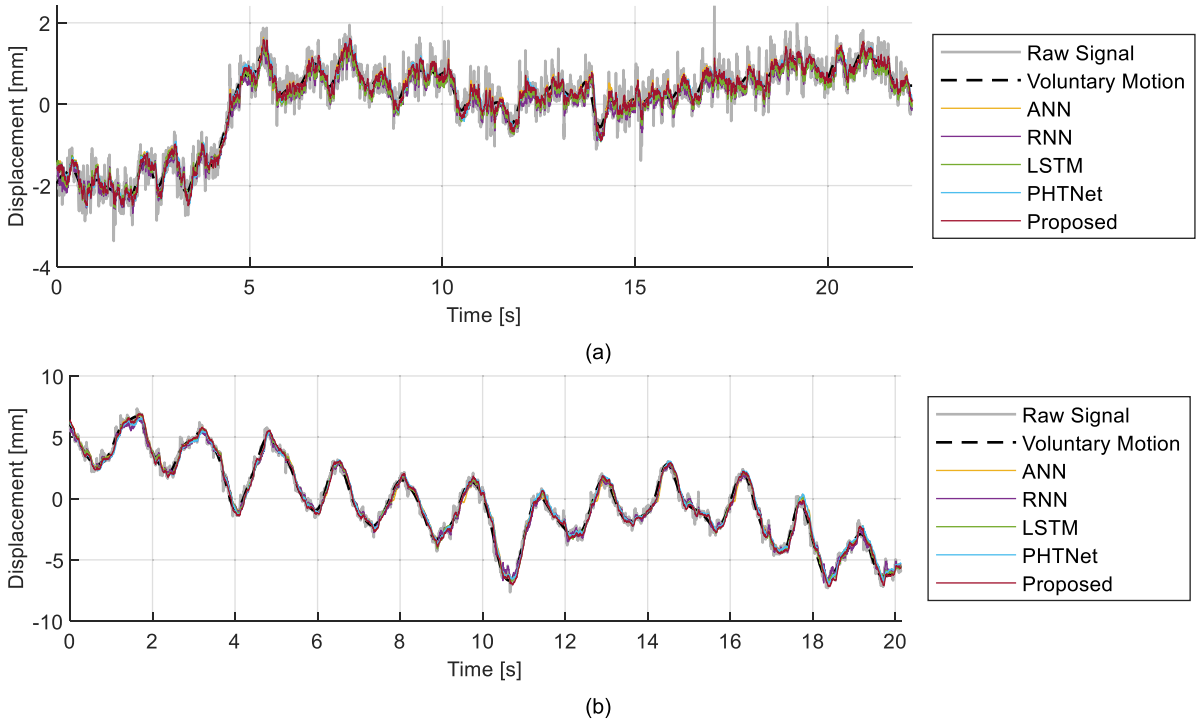


FIGURE 4. Voluntary motion prediction resulted from test on untrained new task data with four baseline models: ANN, RNN, LSTM and PHTNet. Raw signal (grey) and voluntary signal (black dashed) are depicted for reference. (a) Static and (b) dynamic tasks.

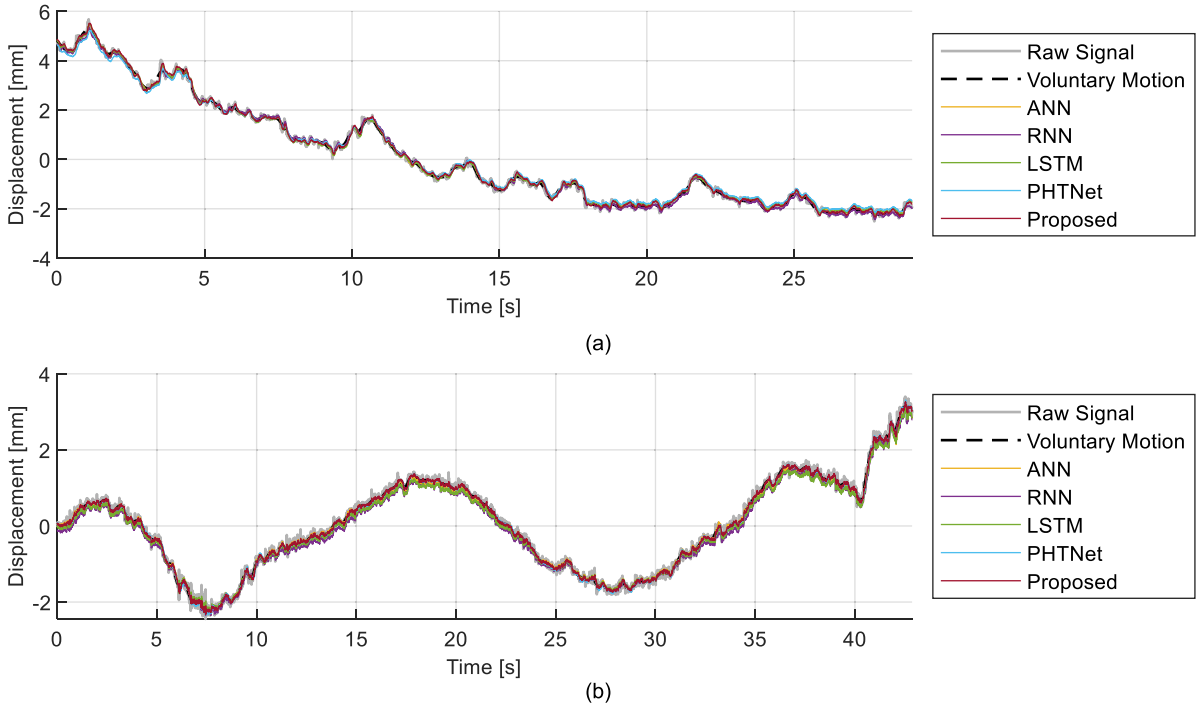


FIGURE 5. Voluntary motion prediction resulted from test on different user data with four baseline models: ANN, RNN, LSTM and PHTNet. Raw signal (grey) and voluntary signal (black dashed) are depicted for reference. (a) Static and (b) dynamic tasks.

time step, the RMSEs also gradually increased. In particular, the prediction error at the 10-time step ahead from the present time step increases by 25%, compared to the error obtained at the time one step ahead.

V. DISCUSSION

Compared with the previous studies to predict voluntary motion [17], [18], [31], the proposed algorithm addresses the issues raised in the existing models [18] by adopting a hybrid

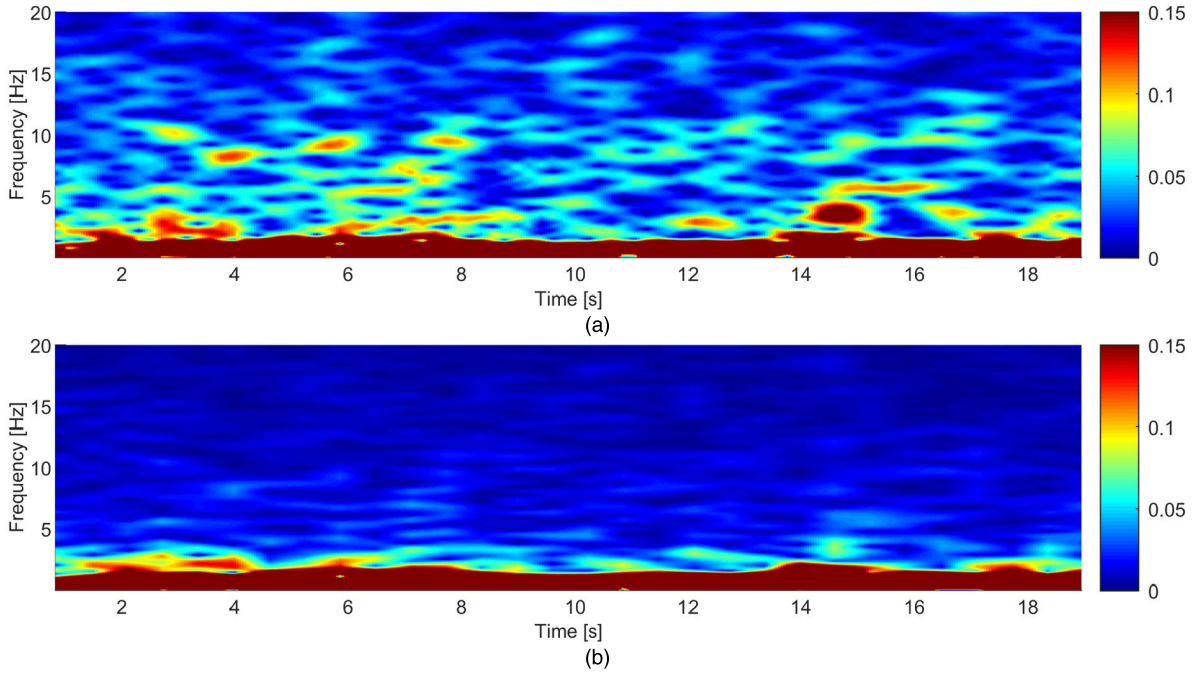


FIGURE 6. Short-time Fourier transform (STFT) of hand motion for 20 s: (a) raw signal and (b) voluntary signal predicted by the proposed method.

TABLE 3. Model test on different subject.

Model	Static		Dynamic	
	RMSE (*mm*)	Accuracy (%)	RMSE (*mm*)	Accuracy (%)
ANN [30]	0.090 (0.300)	92.14 2.62	0.068 (0.020)	93.96 (1.82)
RNN [31]	0.078 (0.020)	94.10 (1.85)	0.071 (0.023)	92.28 (2.12)
LSTM [32]	0.060 (0.013)	94.72 (1.20)	0.054 (0.006)	95.06 (0.57)
PHTNet [18]	0.068 (0.022)	94.06 (1.88)	**0.052** (0.005)	**95.20** (0.44)
Proposed Method	**0.060** (0.007)	**94.72** 0.64	0.055 (0.005)	94.94 (0.50)

The number in parenthesis is standard deviation.

model capable of interpreting such complex hand motion as a composition of sub-signals. Moreover, the ensemble learning approach enables accurately predict nonstationary voluntary signals. As a result, the proposed algorithm outperforms all of the baseline models, including a general (ANN) and the state-the-of-art (PHTNet) algorithms overall in terms of prediction error, accuracy. Specifically, the proposed algorithm is also superior in predicting voluntary signals because the algorithm can deal with various circumstances. For example, the decomposition and ensemble approach can adapt to various forms of voluntary signals: either highly stationary (nearly DC motion) or rather highly cyclic at a range of low frequencies. Consequently, our model can seamlessly accommodate various hand motion signals, which allows us to attain the lower RMSE as well as the lower standard deviation in prediction compared to the baseline models used, which may imply a lower probability of yielding excessive error.

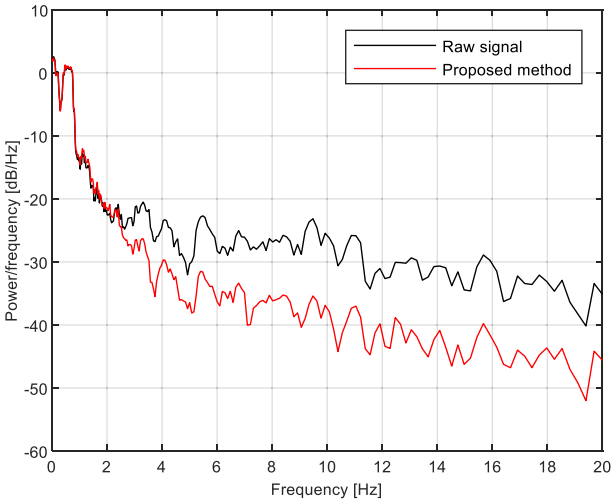


FIGURE 7. Power spectral density of raw signal versus voluntary signal predicted by the proposed method.

Moreover, the proposed algorithm allows for predicting future voluntary motion over multiple time steps ahead via the parametric modeling of voluntary motion rather than simply taking a black-box model. Hence, the proposed model can also deal with the motion prediction at any time step requested potentially by inconsistent time delay in control. For instance, the prediction error was increase only by 2.2% per time step (4.17 ms) in the experiments.

Lastly, the proposed algorithm is substantially time-efficient in learning models since it mitigates the issues raised by the high complexity of deep learning-based approaches. For comparison, the proposed algorithm took 40 seconds to reach one epoch while it was about 5 minutes for PHTNet.

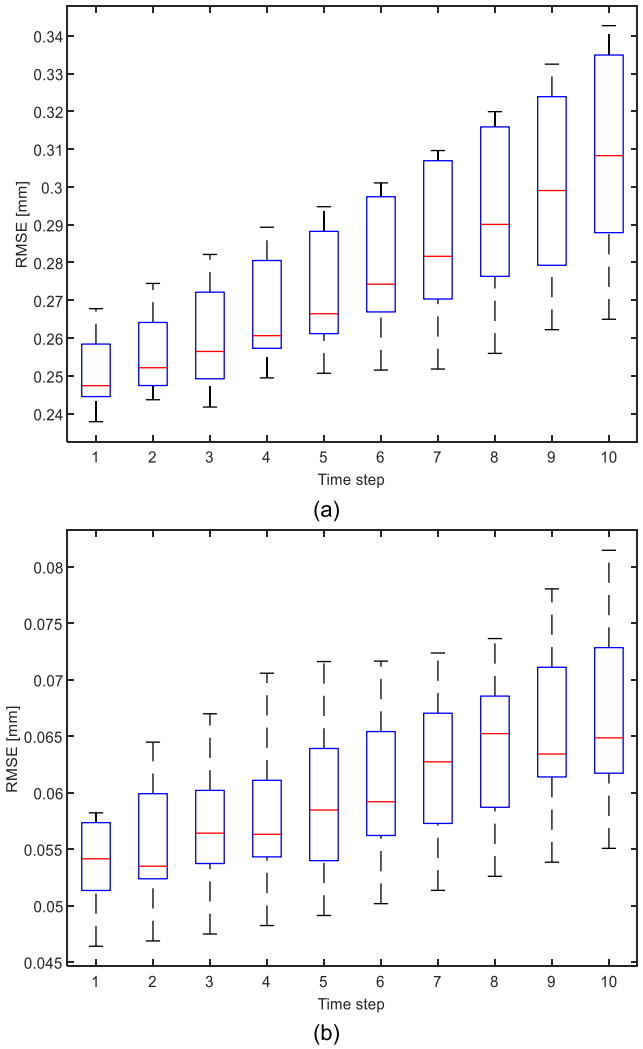


FIGURE 8. Predictive performance for future time steps. (a) Test on the new task, (b) test on the different user.

VI. CONCLUSION AND FUTURE WORK

In summary, this study aims to develop a new deep learning model that can accurately predict voluntary motion from complex and nonstationary hand motion. It is essential for actively suppressing undesired hand tremor and seamlessly delivering voluntary motion to surgical robots during microsurgery. To achieve this goal, we proposed the deep neural network with the decomposition and ensemble approach that can predict future voluntary motion with minimal error. This model thus decomposes complex hand motion into multiple sub-signals and ensembles the parameterized outputs for predicting future voluntary motion. As a result, the proposed model outperforms the other algorithms, which were taken as a baseline for comparison in terms of prediction error (RMSE) and accuracy. Moreover, the algorithm could also be generalizable to new data from the new tasks and also by different subjects. Finally, the proposed algorithm was tested for active tremor suppression, which resulted in substantial tremor attenuation in a frequency of interest, 6—14 Hz. Furthermore, the proposed algorithm is capable of predicting future voluntary motion within a certain time window with low error increment since the output is parameterized with respect to the time step.

To improve the proposed algorithm, future work involves the optimization of the preset historical time window and the number and types of signal blocks. In addition, reinforcement learning to evolve a trained model according to new data would be preferable to enhance overall performance as well as adaptability to various circumstances.

We also plan to apply this algorithm to a surgical robot platform to explore its capability of active tremor compensation. We thus believe the proposed decomposition and ensemble method also suggests a direction to address the time delay caused by linear filters commonly used for active tremor compensation. Finally, the approach and structure of the proposed algorithm would also be utilized in other research areas, such as estimation of pathological tremor for rehabilitation.

REFERENCES

[1] C. Cipriani, M. Controzzi, and M. C. Carrozza, "The smarthand transradial prosthesis," *J. Neuroeng. Rehabil.*, vol. 8, no. 29, pp. 1–13, 2011.

[2] R. Weir, M. Mitchell, S. Clark, G. Puchhammer, K. Kelley, M. Haslinger, N. Kumar, R. Hofbauer, P. Kuschnigg, V. Cornelius, M. Eder, and R. Grausenburger, "New multifunctional prosthetic arm and hand systems," in *Proc. 29th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, Lyon, France, Aug. 2007, pp. 4359–4360.

[3] S. A. Dalley, T. E. Wiste, T. J. Withrow, and M. Goldfarb, "Design of a multifunctional anthropomorphic prosthetic hand with extrinsic actuation," *IEEE/ASME Trans. Mechatronics*, vol. 14, no. 6, pp. 699–706, Dec. 2009.

[4] P. Heo, G. M. Gu, S.-J. Lee, K. Rhee, and J. Kim, "Current hand exoskeleton technologies for rehabilitation and assistive engineering," *Int. J. Precis. Eng. Manuf.*, vol. 13, no. 5, pp. 807–824, May 2012.

[5] G. Dogangil, B. L. Davies, and F. Y. R. Baena, "A review of medical robotics for minimally invasive soft tissue surgery," *Proc. Inst. Mech. Eng. H*, vol. 224, no. 5, pp. 653–679, 2010.

[6] G. H. Ballantyne and F. Moll, "The da Vinci telerobotic surgical system: The virtual operative field and telepresence surgery.," *Surg. Clin. North Amer.*, vol. 83, no. 6, pp. 1293–1304, Dec. 2003.

[7] B. Mitchell, J. Koo, I. Iordachita, P. Kazanzides, A. Kapoor, J. Handa, G. Hager, and R. Taylor, "Development and application of a new steady-hand manipulator for retinal surgery," in *Proc. IEEE Int. Conf. Robot. Autom.*, Apr. 2007, pp. 623–629.

[8] R. Taylor, P. Jensen, L. Whitcomb, A. Barnes, R. Kumar, D. Stoianovici, P. Gupta, Z. Wang, E. deJuan, and L. Kavoussi, "A steady-hand robotic system for microsurgical augmentation," in *Medical Image Computing and Computer-Assisted Intervention—MICCAI* (Lecture Notes in Computer Science), vol. 1679. Berlin, Germany: Springer, 1999.

[9] R. A. MacLachlan, B. C. Becker, J. C. Tabares, G. W. Podnar, L. A. Lobes, and C. N. Riviere, "Micron: An actively stabilized handheld tool for microsurgery," *IEEE Trans. Robot.*, vol. 28, no. 1, pp. 195–212, Feb. 2012.

[10] S. Yang, R. A. MacLachlan, and C. N. Riviere, "Manipulator design and operation of a six-degree-of-freedom handheld tremor-canceling microsurgical instrument," *IEEE/ASME Trans. Mechatronics*, vol. 20, no. 2, pp. 761–772, Apr. 2015.

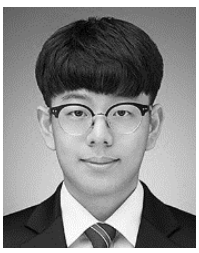
[11] S. Wang, Y. Gao, J. Zhao, and H. Cai, "Adaptive sliding bandlimited multiple Fourier linear combiner for estimation of pathological tremor," *Biomed. Signal Process. Control*, vol. 10, pp. 260–274, Mar. 2014.

[12] J. H. McAuley, "Physiological and pathological tremors and rhythmic central motor control," *Brain*, vol. 123, no. 8, pp. 1545–1567, Aug. 2000.

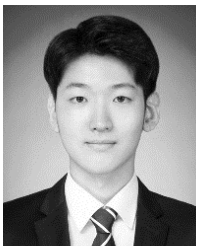
[13] S. Tatinati, K. C. Veluvolu, and W. T. Ang, "Multistep prediction of physiological tremor based on machine learning for robotics assisted microsurgery," *IEEE Trans. Cybern.*, vol. 45, no. 2, pp. 328–339, Feb. 2015.

[14] W. T. Latt, U.-X. Tan, C. Y. Shee, C. N. Riviere, and W. T. Ang, "Compact sensing design of a handheld active tremor compensation instrument," *IEEE Sensors J.*, vol. 9, no. 12, pp. 1864–1871, Dec. 2009.

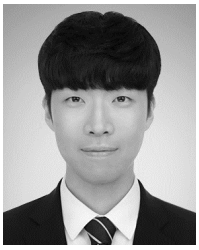
[15] C. N. Riviere, R. Scott Rader, and N. V. Thakor, "Adaptive canceling of physiological tremor for improved precision in microsurgery," *IEEE Trans. Biomed. Eng.*, vol. 45, no. 7, pp. 839–845, 1998.
[16] K. C. Veluvolu and W. T. Ang, "Estimation of physiological tremor from accelerometers for real-time applications," *Sensors*, vol. 11, no. 3, pp. 3020–3036, Mar. 2011.
[17] S. Shahtalebi, S. F. Atashzar, R. V. Patel, and A. Mohammadi, "HMFP-DBRNN: Real-time hand motion filtering and prediction via deep bidirectional RNN," *IEEE Robot. Autom. Lett.*, vol. 4, no. 2, pp. 1061–1068, Apr. 2019.
[18] S. Shahtalebi, S. F. Atashzar, O. Samotus, R. V. Patel, M. S. Jog, and A. Mohammadi, "PHTNet: Characterization and deep mining of involuntary pathological hand tremor using recurrent neural network models," *Sci. Rep.*, vol. 10, no. 1, pp. 1–19, Dec. 2020.
[19] Z. Liu, Q. Wu, Y. Zhang, and C. L. Philip Chen, "Adaptive least squares support vector machines filter for hand tremor canceling in microsurgery," *Int. J. Mach. Learn. Cybern.*, vol. 2, no. 1, pp. 37–47, Mar. 2011.
[20] B. C. Becker, H. Tummala, and C. N. Riviere, "Autoregressive modeling of physiological tremor under microsurgical conditions," in *Proc. 30th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, Aug. 2008, pp. 1948–1951.
[21] S. Tatinati, K. Nazarpour, W. T. Ang, and K. C. Veluvolu, "Multidimensional modeling of physiological tremor for active compensation in handheld surgical robotics," *IEEE Trans. Ind. Electron.*, vol. 64, no. 2, pp. 1645–1655, Feb. 2017.
[22] Y. Xiao, J. Xiao, and S. Wang, "A hybrid model for time series forecasting," *Hum. Syst. Manage.*, vol. 31, no. 2, pp. 133–143, 2012.
[23] J. G. Gonzalez, E. A. Heredia, T. Rahman, K. E. Barner, and G. R. Arce, "Optimal digital filtering for tremor suppression," *IEEE Trans. Biomed. Eng.*, vol. 47, no. 5, pp. 664–673, May 2000.
[24] J. Zhang and F. Chu, "Real-time modeling and prediction of physiological hand tremor," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Philadelphia, PA, USA, 2005, pp. v/645–v/648.
[25] A.-O. Boudraa and J.-C. Cexus, "EMD-based signal filtering," *IEEE Trans. Instrum. Meas.*, vol. 56, no. 6, pp. 2196–2202, Dec. 2007.
[26] A. Lee and E. Altenmüller, "Detecting position dependent tremor with the empirical mode decomposition," *J. Clin. Movement Disorders*, vol. 2, no. 1, pp. 2–7, Dec. 2015.
[27] J. Wang, Z. Wang, J. Li, and J. Wu, "Multilevel wavelet decomposition network for interpretable time series analysis," in *Proc. 24th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Jul. 2018, pp. 2437–2446.
[28] V. Nair and G. E. Hinton, "Rectified linear units improve restricted Boltzmann machines," in *Proc. Int. Conf. Mach. Learn.*, 2010, pp. 807–814.
[29] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.
[30] W. S. McCulloch and W. Pitts, "A logical calculus of the ideas immanent in nervous activity," *Bull. Math. Biophys.*, vol. 5, no. 4, pp. 115–133, Dec. 1943.
[31] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by back-propagating errors," *Nature*, vol. 323, no. 6088, pp. 533–536, Oct. 1986.
[32] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.
[33] K. Cho, B. van Merrienboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio, "Learning phrase representations using RNN encoder–decoder for statistical machine translation," in *Proc. Conf. Empirical Methods Natural Lang. Process. (EMNLP)*, 2014, pp. 1724–1734.
[34] J. Gallego, E. Rocon, J. O. Roa, J. Moreno, and J. L. Pons, "Real-time estimation of pathological tremor parameters from gyroscope data," *Sensors*, vol. 10, no. 3, pp. 2129–2149, Mar. 2010.



DOKYOON YOON received the B.S. degree in industrial engineering from Korea University, Seoul, South Korea, in 2018, where he is currently pursuing the M.S. degree in industrial engineering. His research interests include the development of active compensation algorithm, sensor signal processing, time series modeling using statistical methods, and machine learning-based approaches.



EUNCHAN KIM received the B.S. degree in automotive engineering from Hanyang University, Seoul, South Korea, in 2019, where he is currently pursuing the M.S. degree in mechanical convergence engineering. His research interests include computer vision, signal processing, modern control, and design optimization.



INGU CHOI received the B.S. degree from the Division of Robotics Engineering, Kwangwoon University, Seoul, South Korea, in 2020. He is currently a Research Intern with the Center for Intelligent and Interactive Robotics, Korea Institute of Science and Technology (KIST), Seoul. His research interests include the development of surgical robotics and mechatronics systems.

SUNG WON HAN received the M.S. degrees in operations research, in statistics, and in mathematics from the Georgia Institute of Technology, in 2006, 2007, and 2010, respectively, and the Ph.D. degree from the School of Industrial and Systems Engineering (Statistics), Georgia Institute of Technology. He was a Senior Research Scientist with the Division of Biostatistics, School of Medicine, New York University, and a Postdoctoral Researcher with the Department of Biostatistics and Epidemiology/Center for Clinical Epidemiology and Biostatistics, School of Medicine, University of Pennsylvania. He is currently an Associate Professor with the School of Industrial Management Engineering, Korea University. His research interests include probabilistic graphical model, network analysis, machine learning, and deep learning.

SUNGWOOK YANG (Member, IEEE) received the B.S. and M.S. degrees in mechanical and aerospace engineering from Seoul National University, Seoul, South Korea, in 2004 and 2006, respectively, and the Ph.D. degree in robotics from Carnegie Mellon University, Pittsburgh, PA, USA, in 2015.

From 2006 to 2015, he was a Research Scientist with the Center for BioMicrosystem, Korea Institute of Science and Technology (KIST), Seoul. Since 2016, he has been a Senior Research Scientist with the Center for Intelligent and Interactive Robotics, KIST. His research interests include surgical robotics, microrobotics, neural engineering, and mechatronics systems for biomedical applications. He was a recipient of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) Best Application Paper Award, in 2014, and the IEEE/ASME TRANSACTIONS ON MECHATRONICS Best Paper Award, in 2015. He is currently an Editorial Director of the Korea Society of Mechanical Engineers and the Bio Engineering Division.