A Simultaneous Gesture Classification and Force Estimation Strategy Based on Wearable A-Mode Ultrasound and Cascade Model

Qing Zengyu, Lu Zongxing, Member, IEEE, Liu Zhoujie, Cai Yingjie, Cai Shaoxiong, He Baizheng, and Yao Ligang

Abstract — The existing Human-Machine Interfaces (HMI) based on gesture recognition using surface electromyography (sEMG) have made significant progress. However, the sEMG has inherent limitations as well as the gesture classification and force estimation have not been effectively combined. There are limitations in applications such as prosthetic control and clinical rehabilitation, etc. In this paper, a grasping gesture and force recognition strategy based on wearable A-mode ultrasound and two-stage cascade model is proposed, which can simultaneously estimate the force while classifying the grasping gesture. This paper experiments five grasping gestures and four force levels (5-50%MVC). The results demonstrate that the performance of the proposed model is significantly better than that of the traditional model both in classification and regression (p < 0.001). Additionally, the two-stage cascade regression model (TSCRM) used the Gaussian Process regression model (GPR) with the mean and standard deviation (MSD) feature obtains excellent results, with normalized root-mean-square error (nRMSE) and correlation coefficient (CC) of 0.1049 ± 0.0374 and 0.9461 ± 0.0354, respectively. Besides, the latency of the model meets the requirement of real-time recognition (T < 15ms). Therefore, the research outcomes prove the feasibility of the proposed recognition strategy and provide a reference for the field of prosthetic control, etc.

Index Terms — Gesture classification, force estimation, wearable A-mode ultrasound, cascade model.

I. INTRODUCTION

The Human-Machine Interface (HMI) based on gesture recognition allows more intuitive and convenient Human-Computer interaction, which is widely applied in prosthetic control, virtual reality, clinical rehabilitation, etc. Over the past few decades, researchers have explored a variety of HMI based on gesture recognition, including electroencephalography (EEG) [1], electrocorticography (ECoG) [2], mechanomyography (MMG) [3], computer vision, surface electromyography (sEMG) and ultrasound [4], [5]. Among them: Since sEMG has the characteristics of non-invasiveness, convenience, freedom of space constraints, and close correlation with forearm muscle contraction, researchers have conducted extensive exploration on sEMG-based gesture recognition [6], [7]. According to recent research reports, the data extracted from sEMG has been used in transfer learning [8], continuous gesture decoding [9], cross individual gesture recognition [10], fingertip and grip force estimation [11]–[13], and elbow force prediction [14], [15], which obtained desirable recognition accuracy.

However, sEMG has some inherent limitations, such as crosstalk, low signal-to-noise ratio, and inability to detect deep muscles, causing it is hard to observe the slight contraction and relaxation of deep muscles [16]. Therefore, it is difficult for sEMG to apply in a field that needs fine finger movements control. In contrast, ultrasound has been regarded as one of the alternative signals of sEMG owing to its ability of detecting both superficial and deep muscle deformation and sensitivity to muscle contraction changes [17].

Ultrasound is a type of sound wave belongs to the mechanical wave that travels by repeatedly compressing and expanding the medium. While in medium, the complex ratio of the sound pressure to the particle velocity at a certain point is called the acoustic impedance. The sound pressure is defined as the difference between pressure and static pressure in the medium when there is sound wave, and its specific size depends on the vibration form of the sound. Therefore, as the ultrasonic propagation speed is different in diverse medium, the acoustic impedance of ultrasound is also different between these medium. As a result, when the ultrasound travels across various tissues of the human body, the echo is reflected at the boundary between two mediums which have diverse acoustic impedances. Thus, through the echo, the structure of the human tissue is roughly described. Furthermore, the proportion of echo is proportional to the impedance mismatch between the two mediums. When ultrasound probes are used to send

Manuscript received 24 May 2022; revised 21 July 2022; accepted 2 August 2022. Date of publication 5 August 2022; date of current version 2 August 2022. This work was supported in part by the National Natural Science Foundation of China under Grant 61801122. (Qing Zengyu and Lu Zongxing contributed equally to this work.) (Corresponding authors: Lu Zongxing; Liu Zhoujie.)

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the Ethics Committee of School of Mechanical Engineering and Automation of Fuzhou University.

Qing Zengyu, Lu Zongxing, Cai Yingjie, Cai Shaoxiong, He Baizheng, and Yao Ligang are with the School of Mechanical Engineering and Automation, Fuzhou University, Fuzhou 350108, China (e-mail: qingzy9264@sina.com; luzongxing@fzu.edu.cn; cyj@fzu.edu.cn; caishaoxiong@sina.cn; hebai77@sina.com; ylgyao@fzu.edu.cn). Liu Zhoujie is with The First Affiliated Hospital, Fujian Medical University, Fuzhou 350004, China (e-mail: liuzhoujie_08@163.com).

Digital Object Identifier 10.1109/TNSRE.2022.3196926

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/
out a series of ultrasonic pulses to the forearm, the muscle movement and shape information can be obtained by detecting and analyzing the echo signal [4], [18].

Ultrasound has different modes (A-mode, B-mode, M-mode, etc.). B-mode ultrasound received the first attention as it can visualize the signal as a two-dimensional image, intuitively displaying muscle deformation. Zheng et al. [19]–[21] first evaluated the relationship between wrist angle and B-mode ultrasound and demonstrated the feasibility of characterizing muscle fatigue using B-mode ultrasound. Castellini et al. [22], [23] placed B-mode ultrasound probes on the wrist and predicted the relations between signal, finger position, and fingertip force. In addition, researchers had also used B-mode ultrasound to conduct grip strength assessment [24], real-time gesture classification [25], and lower extremity motion analysis [26], etc., and achieved desirable accuracy. However, B-mode ultrasound is expensive, bulky, and unsuitable for wearing, which restricts its application in HMI based on gesture recognition [27].

Unlike ultrasound images provided by B-mode probes, A-mode ultrasound signals can be provided by independent A-mode probes. On the one hand, as a kind of one-dimensional ultrasound, A-mode ultrasound can only observe the muscle deformation in a predefined direction. On the other hand, the probes of A-mode ultrasound can be smaller and cheaper than B-mode ultrasound. Meanwhile, the computation cost can be further reduced to meet real-time recognition requirements. Guo et al. [28], [29] predicted the relationship between A-mode ultrasound signal and wrist angle by detecting the dynamic thickness change of skeletal muscle during its contraction. In 2017, Sun et al. [30] used a single-channel ultrasound system to explore muscle fatigue in biceps brachii. The following year, Yang et al. [31] evaluated finger movement performance both offline and online via multi-channel A-mode ultrasound and verified the feasibility of wearable HMI. Guo et al. [18] proposed a method of setting a threshold in 2022, which can be applied to stabilize A-mode ultrasound while its signals’ amplitude has changed during the period of recognition.

However, these researches above only classified discrete gestures or estimated the force under one single gesture. More importantly, these researches did not effectively combine gesture classification and force estimation to realize more natural gesture control, which is the goal of this paper. Yang et al. [32] explored simultaneous gesture classification and muscle strength evaluation in 2020, which they called “proportional pattern recognition.” Nevertheless, it is worth noting that this paper only used statistical features to realize robust gesture classification under different muscle contraction degrees and accurate force estimation on a given gesture, which cannot be regarded as simultaneous muscle gesture classification and force evaluation.

In general, if different gestures and force levels are needed to be classified simultaneously, each force level of each gesture can be considered as one class for classification. However, as the number of classes increases, it usually leads to the more complex classifier, the longer training process, and reduced classification accuracy. Moreover, this method cannot be applied to continuous force estimation, which is infinitely separable. According to papers [32]–[34], this paper proposed a two-stage cascade model recognition strategy to classify grasping gestures and estimate corresponding force levels simultaneously based on A-mode ultrasound. The strategy used the first stage gesture model to determine gesture class. Then the second stage force model estimates force under the specific gesture class to realize the grasping gesture classification and its corresponding force estimation.

The rest of this paper is organized as follows: The experiment subjects and apparatus are described in Section II-A to B. Section II-C provides the experiment procedure and signal acquisition. The data preprocessing and feature extraction are given in Section II-D and E. Meanwhile, the classification and regression methods are provided in Section II-F and G, respectively. The cascade model frame and estimation metrics are reported in Section II-H, I and J. Additionally, the performances of cascade models are presented in Section III. Finally, Section IV concludes the paper.

II. METHODOLOGY

A. Experiment Subjects

The experiments were performed by nine non-disabled subjects (six males and three females, age = 23 ± 2) with their dominant hand. Seven subjects were right-hand dominant, and the other two subjects were left-hand dominant. Moreover, four of them had the experience of grasping gestures before the experiments. Additionally, none of them had a history of neuromuscular or joint diseases. Before the experiment, all subjects were informed about the experiment and provided informed consent. The testing procedure followed the declaration of Helsinki and was approved by the Ethics Committee of School of Mechanical Engineering and Automation of Fuzhou University.

B. Experiment Apparatus

The ultrasound signals were acquired by a 4-channel A-mode ultrasound instrument (US-HYS-Q2302A, Elonxi Ltd., Hangzhou, China), which facilitated 4 A-mode
ultrasound probes (height: 11mm, diameter: 9mm). As shown in Fig. 1(a): Each probe’s operating frequency is 2.25MHz, and probing depth is 3.94cm. In addition, the sampling frequency of the ultrasound system is 20MHz. The frame rate of each channel is 10Hz, which means it grabs 10 frames of data per second with 1000 sample points per frame. Consequently, the acquisition data of the A-mode ultrasound instrument per second is a $40 \times 1000$ matrix. In the acquisition process, 4 ultrasound probes were placed around the forearm with a custom belt approximately 3/5 of the forearm length away from the wrist. At the same time, the ultrasonic coupling agent should be applied between the probe and skin to weaken the influence of air and the external environment to keep signal quality.

The grasping signals were collected using a force sensor (AT8602, Autoda Ltd., Suzhou China) with a sampling rate of 100Hz. As shown in Fig. 1(b), the sensor’s rated load, repeatability, and excitation voltage are 100N, 0.1%FS, and 5-12V, respectively.

C. Experiment Procedures

Before the experiment, the maximum voluntary contraction (MVC) was tested. For each gesture, the subjects were asked to use the maximum grip strength as much as possible to keep the gesture stable for 10s and record the force value. This process was repeated three times, and the MVC force was the maximum of them. Between the two MVC test processes, there was a rest of 5 min to avoid muscle fatigue.

For all manipulation tasks, each subject was asked to sit upright with their forearm rested on a custom-made stand (Fig. 1). Cutkosky [35] divided the grasping gestures into two categories: Power and Precision. We divided the grasping gestures into five classes combining Mark’s classification and actual needs of the palm during grasping. As shown in Fig. 2, namely:

1) Later Grasp (LG): with thumb and index finger closing on a flat object, like a book.

2) Tripod Grasp (TG): with thumb, index, and middle fingers closing on a spherical object or slender cylinder such as an egg or a pen.

3) Four Finger Grasp (FFG): with all fingers except the little finger, usually for picking up objects whose mass center is far from the grip point, for example, a cup with a handle.

4) All Finger Grasp (AFG): used when holding a sturdy cylinder.

5) Power Grasp (PG): used when lifting a heavy object with all fingers except the thumb.

The grasping force of each gesture was divided into four force levels: 5%, 20%, 35%, and 50%MVC. Each force level lasted for 5s, and the four force levels lasted for 20s as a trial. Fig. 3 shows the diagram of acquisition time of one trial and the typical A-mode ultrasound signal. This procedure was repeated for four trials for each gesture, and there was a rest of 5s between two continuous trials. Besides, to prevent mutual interference between different force levels, the valid time of each force level was 2s in the middle.

According to above, each gesture had four trials, and each trial had four force levels with 2s valid time. Moreover, the acquisition data of 4 channels’ ultrasound per second was a $40 \times 1000$ matrix. Hence, the data of each gesture was a $4 \times 4 \times 20 \times 1000 = 1280 \times 1000$ matrix. The total data of each subject was a 6400 x 1000 matrix. The ultrasound signals of 4 channels acquired simultaneously were put together and regarded as one ultrasound sample data. So the sample data of each gesture was a $320 \times 4000$ matrix, while that of each subject was a $1600 \times 4000$ matrix.
E. Feature Extraction

As for ultrasound feature extraction, Yang et al. [36] compared conventional time-domain features with a frequency-domain feature. The results showed that the position estimation performance of the frequency-domain feature was inferior to the time-domain feature, while the computation cost was higher than the latter. In this paper, feature comparison was not the focus point. Hence, this paper only selected the time-domain features that have been tested with outstanding performance [31], [32], namely: linear fitting (KB)(We choose the two important parameters K and B as abbreviation), mean and standard deviation (MSD), and their combination (KBMSD).

As mentioned above, the number of sampling points for each channel is 1000. In order to exclude the influence of invalid information from the skin surface and deep layer, 20 points at the head and tail of each frame of each channel have been cut off; only 960 points in the middle have been reserved. During feature extraction, every 20 points were divided into a segment for feature extraction, so 960 points were cut into 48 parts; thus, the dimension of KB and MSD were $48 \times 2 \times 4 = 384$ while that of KBMSD were $48 \times 4 \times 4 = 768$. In order to avoid the dimension disaster and reduce the computational expense, principal component analysis (PCA) was used to reduce the dimension of different features to 40 dimensions, in which contribution rates were greater than 99%.

There was only one dimension for the extraction of force data, and no data preprocessing was required. However, as the sampling frequency of the force sensor was 100Hz, it needed to be sampled down to 10Hz to match the sampling frame rate of the ultrasound system.

F. Classification Model

For the selection of classification models, LDA and SVM models were widely applied in HMI based on gesture recognition of their simple structure, high computational efficiency, and robustness [36]–[38]. This paper chose LDA and SVM (polynomial kernel) to classify different gestures and force levels.

G. Regression Model

In order to successfully access the continuous estimation of grasping force, the regression problem needs to be solved, which is to map the features extracted from ultrasound to actual force hand exerting. We chose two regression models: Gaussian Process Regression (GPR) and Random Forest Regression (RF).

1) Gaussian Process Regression: GPR model is the non-parametric kernel-based probabilistic model with a finite collection of random variables with a multivariate distribution [32], [39]. Every linear combination is evenly distributed.

Firstly, the ultrasound training dataset can be expressed as $D = \{(x_1, y_1), i = 1, 2, \ldots, N\}$, where $x_i = [x_{i1}, x_{i2}, \ldots, x_{im}] \in \mathbb{R}^m$ represents each sample of ultrasound features, and $y_i$ is the force label corresponding to $x_i$. Here, $X \in \mathbb{R}^{N \times m}$ is used to denote ultrasound features, and $y \in \mathbb{R}^N$ is applied to denote force label. Thus the training dataset can be rewritten as $D = \{X, y\}$. Our goal is to infer the mapping function $f(\cdot)$ from the labeled training dataset as follows:

$$y = f(X) + \epsilon$$

where: $f(X) \sim N(0, K(X, X))$, $\epsilon \sim N(0, \sigma^2 I)$ represents random noise. All stand for Gaussian distribution. $K(X, X)$ is the kernel matrix.
is the covariance matrix obtained by calculating all training samples by the covariance (kernel) function \( k(x, x') \), which can be expressed as follows:

\[
\begin{align*}
K(X, X) &= \begin{bmatrix}
k(x_1, x_1) & k(x_1, x_2) & \cdots & k(x_1, x_N) \\
k(x_2, x_1) & k(x_2, x_2) & \cdots & k(x_2, x_N) \\
\vdots & \vdots & \ddots & \vdots \\
k(x_N, x_1) & k(x_N, x_2) & \cdots & k(x_N, x_N)
\end{bmatrix}
\end{align*}
\]  

(4)

Selecting “Squared Exponential” as the core of kernel function \( k(x, x') \) because its robustness. In addition, random noise can be added into kernel function:

\[
k(x, x') = \sigma_f^2 \exp\left[-\frac{(x-x')^T(x-x')}{2l^2}\right] + \sigma_n^2 (x, x')
\]

(5)

Bring the testing dataset \((x_s, y_s)\) into the GPR model, the joint distribution of the training dataset (observed value) and the testing dataset (desired value) can be expressed as:

\[
\begin{bmatrix} y \\ y_s \end{bmatrix} = N \left( 0, \begin{bmatrix} K(X, X) & K^* \\ (K^*)^T & K_{ss} \end{bmatrix} \right)
\]

(6)

where: \( K^* \) indicates the covariance matrix of training dataset \( X \) and testing dataset \( x_s \), and \( K_{ss} \) is the covariance matrix between the testing dataset.

Since the joint distribution is Gaussian, it can be shown that the \( y \) and \( y_s \) are also Gaussian. The Gaussian distribution of the probability \( p(y_s|y) \) is the following expression:

\[
p(y_s|y) = N(\bar{y}_s, var(y_s))
\]

(7)

where:

\[
\bar{y}_s = (K^*)^T[K(X, X)]^{-1}y
\]

(8)

\[
var(y_s) = K_{ss} - (K^*)^T[K(X, X)]^{-1}K^*
\]

(9)

The equations above are the key to Gaussian process regression. Among them: Mean value \( \bar{y}_s \) represents the best-estimated force label considering the training dataset, and its value depends on the covariance matrix \( K(X, X) \) and \( K^* \), etc. Variance value \( var(y_s) \) represents the confidence of the GPR model associated with output estimate.

2) Random Forest Regression: RF is an ensemble algorithm of the decision (i.e., it consists of multiple decision trees) trees, which was originally proposed by Ho using the random subspace method [40]. In 2001, Leo [41] extended this algorithm and combined the “Bagging” idea and Ho’s “feature selection” idea.

On the basis of the Bagging ensemble using decision trees as base learners, RF further introduces random feature selection in the training process of decision trees (Fig. 5). In general, if we collect data according to bootstrap sampling, approximately 2/3 of the training dataset (in-bag sample set) is used to train the decision tree model. For each node of this decision tree model, randomly choose a subset including “k” features out of total “m” features from the node’s feature set. Then select an optimal feature for division-the recommended number of \( k = \log_2 m \). Once the tree was trained, the remaining 1/3 of the training dataset (out-of-bag sample set) is used to test the tree model. The performance of each tree is evaluated, and the tree with the best performance is selected.

This process is repeated “T” times to grow “T” trees. Here the number of \( T \) is 100.

The output of the RF regression model is the average prediction of all basic decision trees, which makes the predictions less prone to overfitting. In addition, the RF regression model is efficient with small and large databases, with a fast prediction rate, and solves multi-dimensional problems efficiently.

H. Two-Stage Cascade Model Framework

The two-stage cascade model was composed of two types of sequential classification or regression models, which were divided into two stages, as shown in Fig. 6: The model of the first stage was gesture classification model (GCM), which was utilized to classify 5 different gestures with all force levels through ultrasound. Notably, different force levels in the same gesture were established as the same label (gesture dataset).

The second stage model consisted of 5 classification models or 5 regression models to estimate grip force. Each of 5 models corresponded to the specific gesture class in GCM and used the ultrasound dataset corresponding to the specific gesture class to train model (force dataset). The label of the classification model was a discrete value artificially assigned to each ultrasound sample according to different force levels. In contrast, the label of the regression model was the continuous force value that had a functional relationship with every sample of ultrasound.

In brief, the gesture class was obtained in GCM, then the second stage model corresponding to the specific gesture class was selected to perform force classification or regression. We called the second stage cascade model for force levels classification as two-stage cascade classification model (TSCCM) and the second stage cascade model for force levels regression as two-stage cascade regression model (TSCR).
Fig. 6. Two-stage cascade model framework. LG, TG, FFG, AFG, and PG are different gestures. FCM means force classification model, and FRM means force regression model. The model of first stage is classification model, which is used to classify different gestures with all force levels. The second stage model consists of 5 classification models or 5 regression models to estimate grip force. When gestures are classified in first stage model, the second stage model could estimate force under specific gesture class immediately.

gesture, two trials of data were randomly selected as the training dataset. The remaining two trials of data were used as the testing dataset to evaluate the performance of each cascade model. Furthermore, every result obtained was a mean value of fifty times of the model running. The purpose of randomly selecting was to simulate the actual application scenario and avoid an accidental situation.

Classification accuracy (CA) was used as the metric to compare the performance of TSCCM and the traditional classification model (TCM). It should be noted that the traditional model classified each force level of each gesture as one class. Therefore, TCM had 20 classes because there were 5 gestures and 4 force levels. Moreover, the CA of TSCCM was not simply the gesture CA in the first stage multiplied by the force CA in the second stage. The correct CA of TSCCM was the number of samples correctly classified into the corresponding gesture and force after twice classifications (SCCTCGF) divided by the total number of the testing dataset (TNTD), and can be expressed as:

\[
CA = \frac{SCCTCGF}{TNTD} \tag{10}
\]

For the regression model, the normalized root-mean-square error (nRMSE) and the correlation coefficient (CC) were used to quantify the performance of gesture and force assessment, which were defined as:

\[
nRMSE = \sqrt{\frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{n}} / (y_{\text{max}} - y_{\text{min}}) \tag{11}
\]

\[
CC = \sqrt{\frac{\sum_{i=1}^{n} (y_i - \bar{y})(\hat{y}_i - \bar{y})}{\sqrt{\sum_{i=1}^{n} (y_i - \bar{y})^2 \sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2}}} \tag{12}
\]

where: \( n \) was the total number of testing samples, \( y_i \) was the observed force value of the \( i \)th sample in the testing dataset, and \( \hat{y}_i \) was the predicted value corresponding to the \( i \)th sample, which was outputted by the regression model. The \( y_{\text{max}}, y_{\text{min}} \) and \( \bar{y} \) were the maximum, minimum, and average force values of the testing sample, respectively.

J. Statistical Analysis

In this paper, for CA in classification as well as nRMSE and CC in regression, the two independent sample t-test was used to assess the different models, the paired sample t-test was applied to estimate the difference between the cascade models and traditional models, and one-way Analysis of Variance (ANOVA) test was employed to value the divergence between features. The significance was set to \( p < 0.05 \).

III. RESULTS

A. CA of Gesture and Force

This section compared the CA of gesture and force between TSCCM and TCM. As shown in Fig. 7: Every accuracy is the averaged CA of each gesture and force of 9 subjects. It can be seen that the CA of TSCCM (average of LDA and SVM were 86% and 88%, respectively) is significantly higher than that of the TCM (average of LDA and SVM were 82% and 85%, respectively) (\( p < 0.001 \)). Meanwhile, the two independent sample t-test is used to compare the performance of the LDA and SVM classification model under the same ultrasound feature. The results show that SVM had more outstanding performance than LDA whatever in TSCCM or TCM (\( p < 0.032 \)). Moreover, the effect of different features on CA was compared using the one-way ANOVA test. The feature KB shows the worst result in the LDA classification model (\( p < 0.05 \)), while feature KBMSD shows the best result in the
B. Performance of Gesture and Force Regression

This section used the regression models in the second stage of two-stage cascade model (also called TSCRM) to evaluate force regression performance when simultaneously classifying gestures. As mentioned above, the performance of the SVM classification model was better than that of the LDA classification model in classifying the gesture and force. Due to the paper’s space limitations, this section only selected the SVM classification model as GCM. It was notable that although the GCM has been achieved desirable CA in all subjects, there were still a small number of gesture classification errors, as shown in Fig. 8: For pursuing the actual usage scenario, samples that predicted what gesture in GCM were regarded as the gesture’s testing samples to assess the force regression model training with specific real gesture samples. At the same time, two trial samples of each gesture were randomly selected and put together as a training dataset to train the traditional regression model (TRM). The other half were picked as testing dataset to test the performance of the model.

Fig. 9 and Fig.10 demonstrate the performance of TSCRM and TRM in gesture and force estimation considering different features and regression models across all subjects. According to Fig. 9(a) and Fig. 10(a), it can be seen that the nRMSE and CC of TSCRM are significantly lower and higher than that of TRM, respectively ($p < 0.001$), no matter of different features. Additionally, the two independent sample t-test was applied to count the statistical differences between regression models with the same feature. The results reveal that the GPR model outperformed the RF regression model as the base regression model of TSCRM and TRM ($p < 0.001$). Moreover, the results of the one-way ANOVA test in Fig. 9(b) and Fig. 10(b) present that when using GPR as the base model of TRM, the performances of different features are comparable ($p > 0.19$), while employing GPR as base model of TSCRM, the KB feature demonstrates worse outcome than the other two features ($p < 0.05$). These statistical differences are more evident when RF was the base model of TSCRM and TRM ($p < 0.02$). Nevertheless, there is no significant difference between MSD and KBMSD features.

Taken overall, it can be concluded that in TSCRM the nRMSE for the GPR model with MSD feature outperform other features and models with $0.1049 \pm 0.0374$. And the CC for the GPR model with MSD feature have the best results with $0.9461 \pm 0.0354$. On the contrary, the worst results of nRMSE and CC are obtained using the RF model and KB feature, with $0.1174 \pm 0.0237$ and $0.9385 \pm 0.0266$, respectively.

According to the conclusions above, a representative example of comparing normalized real force and predicted force using TSCRM (combing GPR model and MSD feature) is...
provided (nRMSE = 0.0772, CC = 0.9807). As shown in Fig. 11: Among different gestures, the fitting degree of force level 1 is highest while force level 2 and level 3 have slight fluctuations while significant fluctuations in force level 4.

Correspondingly, the TSCRM combing GPR model and MSD feature were also utilized to compare the normalized real force and the predicted force of the two more complex force regression trials (RT1 and RT2). As shown in Fig. 12: It can be concluded that with the increase of force level, the prediction error of the TSCRM becomes bigger. Moreover, the model cannot predict well during transient change moments.

Additionally, Fig. 12 and Table. 1 present that the estimation accuracy of “LG” was not desirable. It is assumed that when using LG gesture for grasping objects, the mainly contracted muscles were Abductor pollicis brevis and Flexor pollicis brevis of hand, causing the ultrasound signal on the forearm to be relatively weak and resulting in poor performance of TSCRM. This poor performance of TSCRM was more significant when continuously changing the force level.

**TABLE I**

| Gesture | nRMSE  | CC    | nRMSE  | CC    |
|---------|--------|-------|--------|-------|
| LG      | 0.2671 | 0.6107| 0.2029 | 0.3705|
| TG      | 0.1746 | 0.7932| 0.1421 | 0.8838|
| FFG     | 0.1533 | 0.8812| 0.0753 | 0.9422|
| AFG     | 0.1544 | 0.8366| 0.1198 | 0.8417|
| PG      | 0.1427 | 0.8436| 0.0794 | 0.9542|

RT1 and RT2 meant the two complex force regression trials.
Finally, we summarized the data processing and prediction time of 1600 samples for all five gestures in TSCRM (Table II). All samples were used in data collecting, preprocessing, feature extraction, and PCA dimensionally reduction. While in classification and regression, half samples were applied to train the model, and the other half were employed to test the trained model.

From the Table II, it can be observed that the feature extraction time of the MSD feature is the shortest (2.0823), while that of the KBMSD feature is the longest (6.9031). When training the classification and regression model, the difference in time consumption between features is minor. However, it should be noted that the training and predicting time of the LDA model is about 7 times longer than that of the SVM model. Besides, the RF model’s time consumption is 3.5 times longer than the GPR model. Although the latency of ultrasound data is a little longer, the result reveals that the single sample’s processing and predicting time is less than 15ms in all features, which met the real-time recognition.

**IV. DISCUSSION AND CONCLUSION**

**A. Discussion**

In this paper, we proposed a recognition strategy based on wearable A-mode ultrasound and two-stage cascade model to classify grasping gestures and estimate grip force simultaneously. To this end, a wearable A-mode ultrasound system was used to collect the ultrasound signal of the forearm. At the same time, the force sensor was employed to acquire the force signal during grasping. Additionally, time-domain features with outstanding performance, such as KB, MSD, and KBMSD, were selected as ultrasound features. Meanwhile, LDA and SVM classification models were chosen to classify gesture and force. GPR and RF regression models were applied to assess gesture and force.

This paper was done in 2 parts to compare the performance of the two-stage cascade model and the traditional classification/regression model. The performance of classifying gesture and force between TSCCM and TCM was provided in the first part. The second part presented the TSCRM and TRM’s comparison results in estimating gesture and force.

From section III-A, it can be concluded that the CA of TSCCM for gesture and force was significantly higher than that of TCM for about 3-5% whether LDA or SVM was used as the base classification model ($p < 0.001$). The reason for
this can be attributed to the increase of classes that needed to be classified. The increased classes led to a more complex model and decreased its CA. In contrast, the TSCCM only needed to classify 5 gestures at the first stage and 4 force levels corresponding to specific gestures at the second stage. The base classification was simple and maintained a high CA (Fig. 7(a)). At the same time, the effects of different features on the classification model were explored. The results revealed that MSD and KBMAD features showed better performance than KB feature, which may be because the MSD and KBMSD features were more sensitive to the change of A-mode ultrasound (Fig. 7(b)).

Section III-B presented the performance between TSCRM and TRM for gesture and force evaluation. Remarkably, the samples that predicted to what gesture in GCM were regarded as the gesture’s testing samples to assess the force regression model training with specific real gesture samples for pursuing the actual usage scenario. Surprisingly, the results demonstrated that the nRMSE and CC of TSCRM were significantly lower and higher than that of TRM, respectively (p < 0.001), which means the fitting of the force regression trial of TSCRM was remarkably better than that of TRM, even putting the wrong samples into the second stage regression model. It was speculated that the fitting of the force regression trial of the TSCRM for each gesture was better than the corresponding TRM (similar to the classification model), so that the impact of wrong samples on the model was offset. It was also possible that the ultrasound was not sensitive to force change, especially high-level force. It was causing that the fitting result of wrong samples in TSCRM was not too bad compared to TRM. The results of Fig. 11 and Fig. 12 also confirmed this speculation; with the increased force level, the prediction error of the TSCRM based on ultrasound became bigger. Besides, the latency of TSCRM has been evaluated, and the results demonstrated that the strategy could be used in real-time recognition (T < 15ms for a single sample).

B. Limitations and Future Work

Although the theoretical and experimental results were provided in this paper, there were still many problems to be further discussed. Firstly, this paper only discussed off-line classification and regression accuracy of grasping gestures and forces. Although the actual application scenarios were considered, the accuracy of classification and regression will be reduced when online real-time recognition is carried out, and the model needs to be further improved. Secondly, this paper only discussed recognizing of different grasping gestures and force levels when fixing the forearm, which was not applicable in practical application. Exploring the classification and regression of various gestures under different forearm movements was necessary to be closer to the real-life scene. Thirdly, this paper only explored the feasibility of simultaneous gesture classification and force regression in healthy subjects based on ultrasound. No amputee has been recruited to participate in this experiment, which made the feasibility of this strategy in amputees are unknown.

In addition, A-mode ultrasound is a one-dimensional ultrasound, which cannot provide visual muscle and tendon images of the forearm like B-mode ultrasound, and the small displacement of A-mode ultrasound probe led to significant signal changes, so the robustness was relatively poor. Meanwhile, because the echo signal of A-mode ultrasound detected muscle movement as the morphological response of movement intention, compared with the sEMG signal, the rest state of hand was regarded as a kind of movement in ultrasound. In order to recognize the rest action at a high quality, subjects were required to keep consistent gestures during rest. This paper only discussed several grip gestures and the corresponding force levels and did not involve the rest gesture.

Therefore, our future work mainly starts from the following four points. Firstly, realize online recognition of grasping gesture and force cause the actual application scenarios needs real-time feedback. Secondly, consider multiple gesture and force recognition scenarios, not just a few gestures with fixed body and forearm. Thirdly, collect amputation patients’ ultrasound signals to explore the feasibility of the proposed strategy on disabled people. Finally, the sEMG and ultrasound have complementary in gesture and force recognition, it may be combined with the sEMG signal in the later work.

REFERENCES

[1] A. K. Mohamed, T. Marwala, and L. R. John, “Single-trial EEG discrimination between wrist and finger movement imagery and execution in a sensorimotor BCI,” in Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc., Boston, MA, USA, Aug. 2011, pp. 6289–6293.
[2] T. Yanagisawa et al., “Electrocorticographic control of a prosthetic arm in paralyzed patients,” Ann. Neurol., vol. 71, no. 3, pp. 353–361, Mar. 2012.
[3] H.-B. Xie, Y.-P. Zheng, and J.-Y. Guo, “Classification of the mechanomyogram signal using a wavelet packet transform and singular value decomposition for multifunction prosthesis control,” Physiol. Meas., vol. 30, no. 5, pp. 441–457, Mar. 2009.
[4] L. Guo, Z. Lu, and L. Yao, “Human-machine interaction sensing technology based on hand gesture recognition: A review,” IEEE Trans. Human-Mach. Syst., vol. 51, no. 4, pp. 300–309, Aug. 2021.
[5] S. Cai, Z. Lu, L. Guo, Z. Qing, and L. Yao, “The LET procedure for gesture recognition with multiple forearm angles,” IEEE Sensors J., vol. 22, no. 13, pp. 13226–13233, Jul. 2022.
[6] M. Simão, N. Mendes, O. Gibaru, and P. Neto, “A review on electromyography decoding and pattern recognition for human-machine interaction,” IEEE Access, vol. 7, pp. 39564–39582, 2019.
[7] A. Jaramillo-Yánez, M. E. Benalcázár, and E. Mena-Maldonado, “Real-time hand gesture recognition using surface electromyography and machine learning: A systematic literature review,” Sensors, vol. 20, no. 9, p. 2467, May 2020.
[8] U. Cote-Allard et al., “Deep learning for electromyographic hand gesture signal classification using transfer learning,” IEEE Trans. Neural Syst. Rehabil. Eng., vol. 27, no. 4, pp. 760–771, Apr. 2019.
[9] A. Dwivedi, “A learning scheme for EMG based decoding of dexterous, in-hand manipulation motions,” IEEE Trans. Neural Syst. Rehabil. Eng., vol. 27, no. 10, pp. 2205–2215, Oct. 2019.
[10] H. Min et al., “Cross-individual gesture recognition based on long short-term memory networks,” Sci. Program., vol. 2021, pp. 1–11, Jul. 2021.
[11] S. Zhang et al., “Muscle strength assessment system using sEMG-based force prediction method for wrist joint,” J. Med. Biol. Eng., vol. 36, no. 1, pp. 121–131, Feb. 2016.
[12] H. Cao, S. Sun, and K. Zhang, “Modified EMG-based hand grip force prediction using extreme learning machine,” Soft Comput., vol. 21, no. 2, pp. 491–500, Jan. 2017.
[13] R. Ma, L. Zhang, G. Li, D. Jiang, S. Xu, and D. Chen, “Grasping force prediction based on sEMG signals,” Alexandria Eng. J., vol. 59, no. 3, pp. 1135–1147, Jun. 2020.
[14] C. Huang, X. Chen, S. Cao, B. Qiu, and X. Zhang, “An isometric muscle force estimation framework based on a high-density surface EMG array and an NMF algorithm,” J. Neural Eng., vol. 14, no. 4, May 2017, Art. no. 046005.
[15] L. Xu, X. Chen, S. Cao, X. Zhang, and X. Chen, “Feasibility study of advanced neural networks applied to sEMG-based force estimation,” *Sensors*, vol. 18, no. 10, p. 3226, Oct. 2018.

[16] K. Fujikawa, S. Shirafuji, B. Su, E. Piovaneli, and J. Ota, “Estimation of fingertip forces using high-density surface electromyography,” in *Proc. Int. Symp. Micro-NanoMechatron. Hum. Sci. (MHS)*, Nagoya, Japan, Dec. 2017, pp. 1–5.

[17] J.-Y. Guo, Y.-P. Zheng, Q.-H. Huang, X. Chen, J.-F. He, and H. L.-W. Chan, “Performances of one-dimensional sonomyography and surface electromyography in tracking guided patterns of wrist extension,” *Ultrasound Med. Biol.*, vol. 35, no. 6, pp. 894–902, Jun. 2009.

[18] L. Guo, Z. Lu, L. Yao, and S. Cai, “A gesture recognition strategy based on A-mode ultrasound for identifying known and unknown gestures,” *IEEE Sensors J.*, vol. 22, no. 11, pp. 10730–10739, Jun. 2022.

[19] Y. P. Zheng, “Sonomyography: Monitoring morphological changes of forearm muscles in actions with the feasibility for the control of powered prosthesis,” *Med. Eng. Phys.*, vol. 28, no. 5, pp. 405–415, Jun. 2006.

[20] J. Shi, Y.-P. Zheng, X. Chen, and Q.-H. Huang, “Assessment of muscle fatigue using sonomyography: Muscle thickness change detected from ultrasound images,” *Med. Eng. Phys.*, vol. 29, no. 4, pp. 472–479, May 2007.

[21] H.-B. Xie, Y.-P. Zheng, J.-Y. Guo, X. Chen, and J. Shi, “Estimation of wrist angle from sonomyography using support vector machine and artificial neural network models,” *Med. Eng. Phys.*, vol. 31, no. 3, pp. 384–391, Apr. 2009.

[22] C. Castellini, G. Passig, and E. Zarka, “Using ultrasound images of the forearm to predict finger positions,” *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 20, no. 6, pp. 788–797, Nov. 2012.

[23] D. S. González and C. Castellini, “A realistic implementation of ultrasound imaging as a human-machine interface for upper-limb amputees,” *Frontiers Neurorobot.*, vol. 7, p. 17, Oct. 2013.

[24] T. Abe, B. R. Counts, B. E. Barnett, S. J. Dankel, K. Lee, and et al., “Measurement of wrist angle from sonomyography using support vector machine and artificial neural network models,” *Frontiers Neurorobot.*, vol. 7, p. 17, Oct. 2013.

[25] M. H. Jahanandish, N. P. Fey, and K. Hoyt, “Lower limb motion estimation using ultrasound imaging: A framework for assistive device control,” *IEEE J. Biomed. Health Inform.*, vol. 23, no. 6, pp. 2505–2514, Nov. 2019.

[26] J. Yan, X. Yang, X. Sun, Z. Chen, and H. Liu, “A lightweight ultrasound probe for wearable human–machine interfaces,” *IEEE Sensors J.*, vol. 19, no. 14, pp. 5895–5903, Jul. 2019.

[27] X. Yang, Z. Chen, N. Hettiarachchi, J. Yan, and H. Liu, “A wearable ultrasound system for sensing muscular morphological deformations,” *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 51, no. 6, pp. 3370–3379, Jun. 2021.

[28] J.-Y. Guo, Y.-P. Zheng, H.-B. Xie, and T. K. Koo, “Towards the application of one-dimensional sonomyography for powered upper-limb prosthetic control using machine learning models,” *Prosthetics Orthotics Int.*, vol. 37, no. 1, pp. 43–49, Feb. 2013.

[29] X. Sun, Y. Li, and H. Liu, “Muscle fatigue assessment using one-channel single-element ultrasound transducer,” in *Proc. 8th Int. IEEE/EMBS Conf. Neural Eng. (NER)*, Shanghai, China, May 2017, pp. 122–125.

[30] X. Yang, X. Sun, D. Zhou, Y. Li, and H. Liu, “Towards wearable a-mode ultrasound sensing for real-time finger motion recognition,” *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 26, no. 6, pp. 1199–1208, Jun. 2018.

[31] X. Yang, J. Yan, Z. Chen, H. Ding, and H. Liu, “A proportional pattern recognition control scheme for wearable A-mode ultrasound sensing,” *IEEE Trans. Ind. Electron.*, vol. 67, no. 1, pp. 800–808, Jan. 2020.

[32] Y. Geng, P. Zhou, and G. Li, “Toward attenuating the impact of arm positions on electromyography pattern-recognition based motion classification in transradial amputees,” *J. NeuroEng. Rehabil.*, vol. 9, no. 1, pp. 1–11, Oct. 2012.

[33] A. T. Belyea, K. B. Englehart, and E. J. Scheme, “A proportional control scheme for high density force myography,” *J. Neural Eng.*, vol. 15, no. 4, Jun. 2018, Art. no. 046029.

[34] M. R. Cutkosky, “On grasp choice, grasp models, and the design of hands for manufacturing tasks,” *IEEE Trans. Robot. Autom.*, vol. 5, no. 3, pp. 269–279, Jun. 1989.

[35] X. Yang, Y. Zhou, and H. Liu, “Wearable ultrasound-based decoding of simultaneous wrist/hand kinematics,” *IEEE Trans. Ind. Electron.*, vol. 68, no. 9, pp. 8667–8675, Sep. 2021.

[36] X. Zhu, J. Liu, D. Zhang, X. Sheng, and N. Jiang, “Cascaded adaptation framework for fast calibration of myoelectric control,” *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 25, no. 3, pp. 254–264, Mar. 2017.

[37] Z. Qing, Z. Lu, Y. Cai, and J. Wang, “Elements influencing sEMG-based gesture decoding: Muscle fatigue, forearm angle and acquisition time,” *Sensors*, vol. 21, no. 22, p. 7713, Nov. 2021.

[38] N. Zhang, J. Xiong, J. Zhong, and K. Leatham, “Gaussian process regression method for classification for high-dimensional data with limited samples,” in *Proc. 8th Int. Conf. Inf. Sci. Technol. (ICIST)*, Granada, Spain, Jun. 2018, pp. 358–363.

[39] T. K. Ho, “Random decision forests,” in *Proc. 3rd Int. Conf. Document Anal. Recognit.*, Montreal, QC, Canada, 1995, pp. 278–282.

[40] L. Breiman, “Random forests,” *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, Oct. 2001.