Prediction of knee trajectory based on surface electromyogram with independent component analysis combined with support vector regression

Meng Zhu¹, XiaoRong Guan¹, Zhong Li¹, YunLong Gao¹, KaiFan Zou¹, XinAn Gao¹, Zheng Wang¹, HuiBin Li¹ and KeShu Cai²

Abstract
In recent years, surface electromyogram signals have been increasingly used to operate wearable devices. These devices can aid to help workers or soldiers to lower the load in the task to boost efficiency. However, achieving effective signal prediction has always been a challenge. It is critical to use an appropriate signal preprocessing method and prediction algorithm when developing a controller that can accurately predict and control human movements in real time. For this purpose, this article investigates the effect of various surface electromyogram preprocessing methods and algorithms on prediction results. Walking data (surface electromyogram angle) were collected from 10 adults (5 males and 5 females). To investigate the effect of preprocessing methods on the experimental results, the raw surface electromyogram signals were grouped and subjected to different preprocessing (bandpass/principal component analysis/independent component analysis, respectively). The processed data were then imported into the random forest and support vector regression algorithm for training and prediction. Multiple scenarios were combined to compare the results. The independent component analysis-processed data had the best performance in terms of convergence time and prediction accuracy in the support vector regression algorithm. The prediction accuracy of knee motion with this scheme was 94.54% ± 2.98. Notably, the forecast time was halved in comparison to the other combinations. The independent component analysis algorithm’s “blind source separation” feature effectively separates the original surface electromyogram signal and reduces signal noise, hence increasing prediction efficiency. The main contribution of this work is that the method (independent component analysis + support vector regression) has the potency of best prediction of surface electromyogram signal for knee movement. This work is the first step toward myoelectric control of assisted exoskeleton robots through discrete decoding.

Keywords
sEMG, exoskeleton robot, independent component analysis, support vector regression, knee trajectory, random forest

Date received: 29 June 2021; accepted: 18 July 2022

Introduction
Over the last few decades, interest in research on assisted exoskeleton robots has grown. Assisted exoskeleton robots are wearable robots with human-like drive joints that can assist humans with mobility. They are designed to assist...
workers or soldiers in reducing the weight they bear during duties to maximize efficiency. In practice, excellent human–machine communication has always been critical to ensuring the overall exoskeleton system’s performance and comfort. However, because conventional exoskeleton sensors are incapable of anticipating movement trends, the intent of the wearer’s movement constrains the development of this technology. Existing research has demonstrated that electromyogram (EMG) signals can provide information about neuromuscular activity and can be used to operate exoskeletons. EMG signals have been frequently utilized in human–machine cooperation systems in recent years. Among these, the acquisition mode method based on noninvasive surface electromyogram (sEMG) has demonstrated considerable promise in the study of EMG signals. In comparison to angle and force measurements, an EMG signal can indicate muscle contraction force 30–100 ms earlier. Indeed, the EMG signal has a wide range of applications. It can be used to develop rehabilitation robots, analyze human kinematics, and make clinical medical diagnoses.

Due to the fact that the frequency of human kinematics (about 20 Hz) is low in comparison to the frequency of sEMG. Consequently, EMG signals and features can be recovered from such a “temporal redundancy window” and these features can be used to estimate the next phase of kinematic trends. As a result of this finding, it becomes practical to use EMG data to predict motion trends and send control commands to the exoskeleton. Therefore, it is crucial to investigate the relationship between EMG signals and human motion states, as well as the ability of accurate and fast control of mechanical devices. Currently, research on using EMG data to predict human motion intention is mostly focused on classification and regression models. Mukhopadhyay and Samui classified diverse hand movements using deep neural network (DNN) models and compared them to existing machine learning algorithms. The experimental results indicated that the DNN-based system outperformed other classifiers such as K-nearest neighbor (KNN), random forest (RF), and decision tree. Au and Kirsch investigated the predictive capacity of a time-delay artificial neural network for shoulder and elbow movements. To estimate the angles of the four joints, six distinct EMG signals from the shoulder and elbow muscles were employed as inputs. The results indicated that the prediction’s average root mean square error (RMSE) was less than 20°. Tsuji et al. and Kiatpanichagij and Azfulpurkar classified and predicted human hand postures using a variety of machine learning approaches, including linear discriminant analysis, artificial neural networks, and (support vector machines (SVM).

High-quality signals provide more information required for intent prediction, hence increasing prediction accuracy. However, during the collecting of sEMG signals, different interferences and disturbances are unavoidable. During the experiment, patch position changes, sweat on the human skin surface, and transmission problems in the EMG sensors can all have an effect on data gathering. Many signal processing applications, notably in the communications and medical fields, require preprocessing of data gathered by sensors to decrease noise. How to reduce the influence of signal interference becomes a challenge. Indeed, numerous scientists have investigated ways to mitigate signal noise pollution. Tapia et al. filtered the EMG signal using the empirical modal decomposition algorithm. The processed data result in a more accurate sequence of motion activation during walking. The experimental results demonstrate that this approach is capable of overcoming the detrimental effects of volume conduction, such as cross talk and signal loss in the tissue, as well as unidentified noise sources. Li et al. established a knee joint motion estimation model based on the EMG signal using principal component analysis (PCA). PCA-processed EMG signal data have been shown to enhance prediction accuracy. Kilner et al. recorded electromyography (EMG) from five distinct forearm muscles. The (independent component analysis (ICA) algorithm was used to minimize the signal’s noise. The experimental results showed that the ICA algorithm successfully eliminated cross talk in the signal. Li et al. proposed an enhanced ICA model for the removal of EMG artifacts from electroencephalograph (EEG). In this novel method, EMG signals from neck and head muscles are supplied to the ICA algorithm’s input to “push” the majority of the power associated with EMG artifacts into multiple distinct components. The results demonstrate that ERASE successfully reduces EMG artifacts (on average, around 75% of EMG artifacts are reduced when genuine EMG is used as a reference artifact) while keeping predicted EEG motion properties. Kim et al. investigated the effects of various disturbances on the quality of EMG signals (power and motion artifacts as well as neighboring muscle activity). The ICA algorithm and nonnegative matrix decomposition were used to generate a low-dimensional input signal in their tests. They were employed to perform orientation-based motion classification after splitting the noise into superficial and deep muscles. The results demonstrate that their proposed decomposition strategy enhances classification performance when input dimensions are kept to a minimum. Hassan and Karami investigated the effect of the ICA algorithm on removing noise from brain wave signals. After reducing noise, the enhanced independent components are reassembled to obtain the original signal in its natural state.

In practical applications, sEMG signals were gathered in advance and trained offline in these devices. Typically, these exercises consist of a series of fixed movements (mainly involving upper limb movements). The offline system predicted the sEMG signals utilized to drive the robotic arm. The system determines the present movements and issues control commands for various movement patterns based on the input signals. Kiguchi et al. and Sato and Yagi developed an EMG-based upper limb
exercise-assisted exoskeleton for humans. The RMS was immediately extracted as the control command in their work. Chen et al.\(^28\) proposed a control strategy based on EMG signals to control the lower limb exoskeleton. EMG signals are collected during the user’s standing and sitting movements. By collecting and processing the EMG signals, the lower limb exoskeleton can determine whether the user wants to leave the chair for a walk or sit down to rest. Following that, the control system provides instructions to assist the user in moving based on the detection of EMG signals and the prediction of EMG signal trends. Gordleeva et al.\(^29\) designed an exoskeleton based on the control of a multimodal signaling system. The system was tested on healthy subjects who operated the exoskeleton under different conditions. It was demonstrated that the developed system can analyze up to 15 signals simultaneously in real time during the movement. With the assistance of EMG signals reflecting the movement intention, the control system was able to differentiate and initiate the movement of the right and left legs with a high degree of reliability. He and Kiguchi\(^30\) proposed an EMG-based lower limb exoskeleton control method for assisting the movement of disabled people. EMG signals from the skin surface are mainly used as input information to the controller. To generate flexible and smooth movements, fuzzy neural control methods have been applied to the controller. Experimental results have illustrated the effectiveness of the designed EMG-based controller. Kiguchi and Imada\(^31\) developed a control approach based on the muscle model by employing the EMG signal as the control signal. It enables the activation of a power-assisted lower limb robot in response to the user’s motor intention. In their method, they assess the user’s motion intention in real time using a matrix representing the link between muscle activity and generated joint torque.

It can be inferred that the majority of previous research on surface EMG and human motion intention recognition has concentrated on the challenge of fixed pose categorization. This leads to limitations in the movement of the device. Also, the effect of signal interference and noise on prediction accuracy has received less attention in these research. Simple filtering has been applied to the signals used for prediction. While several previous research separated some noise, they did not examine its effect on classification or regression performance. Although the ICA approach has shown promise in a number of instances, no scholar have explored the regression performance of the data after ICA processing. It has been shown that the SVM algorithm is effective in predicting action recognition problems, but little research has been done on their regression performance.

As a result, this article examines the effect of data preparation techniques on signal regression prediction based on the relationship between sEMG signal and lower extremity mobility. Walking data (sEMG + angle) were collected from 10 adult males (5 males and 5 females). To assess the effects of various algorithms and data pretreatment methods on prediction results, different preprocessing methods (bandpass/PCA/ICA) and prediction algorithms (support vector regression (SVR)/RF) were used. By comparing the outcomes of each, the optimal combination of regressor processing in this model was determined.

The remainder of the essay is organized as follows. The second section discusses the studies in detail, including the experimental design, data collecting, and data preprocessing. The third section examines the prediction errors and Pearson correlation coefficients for data preprocessed using various algorithms. The fourth section discusses the outcomes of the calculations. Finally, the final section has the conclusion.

**Materials and methods**

**Data acquisition**

*Joint angle acquisition.* Codamotion is a supplier of motion capture equipment to the academic research, clinical and other related life science markets (info@codamotion.com). In 1970, David Mitchelson created the company in London, United Kingdom. The experiment used Codamotion to acquire angular data from the motion of the human knee. Two cameras (to record the trajectory of the marker points), a computer (to solve and save data), and several marker points comprise the equipment. The positioning of these marker points follows the Rizzoli scheme guidelines.\(^32\) The experiment’s sampling rate was set to 200 Hz. Each marker is individually coded. As a result, they have an extremely low latency of approximately 0.5 ms. As illustrated in Figure 1, three markers (1,2,3) were attached to the subject’s knee joint. Data were collected during the experiment. Throughout the experiment, the camera recorded the marker’s spatial coordinates. Through the data collecting box,
the marker transmitted the data it acquired to the host computer.

**EMG signal acquisition.** EMG is a research technique used to investigate the genesis, recording, and interpretation of EMG signals. EMG signals are generated when the condition of the muscle fiber membrane changes physiologically. Delsys is a world leader in the design, manufacture, and marketing of high-performance electromyographic (EMG) equipment (sales@delsys.com). Since its founding in 1993 in Natick, Massachusetts, USA, Delsys has focused on solving the engineering challenges associated with wearable EMG sensors. These challenges include low signal artifacts, low cross talk, signal reliability, and signal coherence. sEMG signal was collected in this experiment using the TrignoTM Wireless EMG (a wireless surface EMG collecting device from Delsys). The device’s sensing delay is less than 500 μs. The device is equipped with 16 sensors that operate at a maximum sampling rate of 4000 Hz. A sampling rate of 2000 Hz was chosen in this experiment to avoid signal loss. The muscle groups near the knee play a significant role in human walking. As a result, the quadriceps, medial, and lateral femoris muscles of the thigh muscle group, and the gastrocnemius and soleus muscles of the calf muscle group, were assigned as the EMG signal acquisition locations. Before the experiment’s formal start, the selected sites were wiped with medicinal alcohol to guarantee an efficient capture of EMG signals. It is critical to remember that the sensors must be properly attached to the muscles to avoid the EMG signal collector being deflected during activity. Otherwise, data loss may occur. Figure 2 illustrates the pasted position.

**Experimental procedures**

Ten adults (five males and five females) were selected as volunteers. The subjects’ data are listed in Table 1 below.

Prior to the experiment, the subject’s leg was free of sprains and other impairments of motor function, and there was no pain in the muscle. All subjects were informed and signed a consent form before participating in the experiment. The experiment was approved and conducted in line with the Declaration of Helsinki by the review board of The First Affiliated Hospital of Nanjing Medical University.

Participants were instructed to stroll freely and relaxedly throughout the experiment. It should be mentioned that the 3D motion capture system (Codamotion) activated the EMG acquisition system concurrently with the data acquisition in this experiment. Thus, the experiment’s data collection was synchronized. This guaranteed that when data analysis was performed, the EMG data corresponding to the knee angle data. As advised, participants were invited to complete 3–5 min of low-intensity exercise before the experiment’s start. The individual was advised to remain immobile and relaxed for 3s before and after each round of experiments. To ensure the collection of as much data as feasible, each participant was invited to repeat the experiment 60 times. A 5-min pause was taken between every five sets or whenever the subject became fatigued before the experiment was continued. The experiment utilized the Codamotion 3D motion capture system (which includes the motion acquisition marker and camera) and the TrignoTM Wireless EMG acquisition device. Finally, 600 sets of data were acquired from 10 distinct individuals. The experiment’s setup is depicted in Figure 3.

The experiment was performed as shown in Figure 4, where the subjects were asked to walk according to a constant rhythm. The entire experiment lasted 16s. Each set of experiments contains five steps. During the first 3s and the last 3s of the experiment, the subjects were asked to remain in a standing position without moving.

**PCA/ICA algorithm**

PCA is a basic transformation that maximizes the variance of the transformed data. The variance’s magnitude indicates the quantity of information available about a variable. Unlike when explaining data stability, the bigger the estimated variance of the data, the more effective it is trained

| Table 1. Statistical parameters of volunteers. |
|-----------------------------------------------|
| Mass (kg) | Height (cm) | Age (years) | Gender |
| Sub1      | 85          | 186         | 25      | Male   |
| Sub2      | 70          | 173         | 24      | Male   |
| Sub3      | 72          | 176         | 24      | Male   |
| Sub4      | 65          | 168         | 22      | Male   |
| Sub5      | 75          | 177         | 24      | Male   |
| Sub6      | 53          | 168         | 25      | Female |
| Sub7      | 59          | 171         | 23      | Female |
| Sub8      | 61          | 166         | 22      | Female |
| Sub9      | 51          | 163         | 24      | Female |
| Sub10     | 55          | 165         | 27      | Female |
in machine learning. The following diagram illustrates the PCA algorithm’s calculation mechanism.

First of all, all features were centralized as follows

\[ \tilde{x}_n = \frac{1}{M} \sum_{i=1}^{M} x^i_n \]  

Then, the covariance matrix can be derived

\[ \text{cov}(x_n, x_n) = \frac{\sum_{i=1}^{M} (x^i_n - \tilde{x}_n)^2}{M - 1} \]  

After, all eigenvalues and eigenvectors of the covariance matrix can be found according to the formula

\[ C\mu = \lambda \mu \Rightarrow \{ (\lambda_1, \mu_1), (\lambda_2, \mu_2), \cdots, (\lambda_3, \mu_3) \} \]  

Finally, the original features are projected onto the feature vector of the selection to obtain the new dimensional features after dimensionality reduction

\[ y_n^T = (\mu_1^T x^1_n, x^2_n, \cdots, x^n_n) \]  

ICA is a novel signal processing technique introduced in the 1990s by Jutten and Herault. The goal of this strategy is to breakdown observable data into statistically independent components using a linear decomposition technique. The procedure of solving the blind source separation problem using the ICA algorithm is depicted in Figure 5.

Where \( x(t) \) is the acquired EMG signal in the experiment, which is actually composed of the source signal \( s(t) \) from the muscle and the noise \( v(t) \) from other sensors during the acquisition process. The signal can be approximated
as a linear mixed system, expressed by the following equation

\[ x(t) = As(t) + v(t) \] (5)

The purpose of the ICA algorithm is to isolate the source signal \( s(t) \) from the above equation and, by calculating it, to obtain a signal \( y(t) \) that approximates the original signal.

While both ICA and PCA are multivariate data analysis techniques from a statistical standpoint, the components derived from ICA processing are not only de-correlated but also statistically independent of one another and non-Gaussian distributed. As a result, ICA can more fully show the data’s fundamental structure. Thus, ICA represents a significant advancement over older methods in a number of ways, making it a more promising tool in signal processing. It has been widely applied in a variety of domains, including pattern recognition, signal denoising, and image processing.

**Signal preprocessing**

**Joint angle preprocessing.** The experiment’s coordinates for the marker cannot be utilized directly. It must be turned into an infinitely variable knee angle. The angle of motion of the knee joint must be solved using a link vector model. Two neighboring points \((1 \rightarrow 2, 2 \rightarrow 3)\) are connected as vectors \( \vec{m} \) and \( \vec{n} \) using the spatial coordinates of the three markers (as shown in Figure 6). The angle between the two vectors can then be determined using Codamotion’s built-in solver. The angle is calculated using the following formula

\[ \theta = \arccos \left( \frac{\vec{m} \cdot \vec{n}}{|\vec{m}| \cdot |\vec{n}|} \right) \] (6)

**sEMG signal preprocessing.** To examine the effect of different data preprocessing algorithms on experimental prediction findings, sEMG signals were processed immediately following the signal acquisition. Band-pass filtering (20–450 Hz), PCA/ICA processing, and absolute valuation are all included in the data processing procedure. To begin with, three copies of the data are made for each individual. These three copies of data are processed in the order of Bandpass→Abs, Bandpass→PCA→Abs, and Bandpass→ICA→Abs. Since the joint motion angle acquisition frequency is different from the EMG signal acquisition frequency, it results in different lengths of the two sets of data. To obtain the correspondence of the data, the sEMG signal is interpolated. It enables the synchronization of two sets of data of different sizes. Finally, 1800 sets of data were acquired from 10 distinct individuals. The whole data processing procedure is depicted in Figure 7.
Estimation models for regression

Support vector regression. Support vector regression is a machine learning algorithm that is based on the notion of statistical learning. It enhances learning generalization by attempting to minimize structured risk. As a result, empirical risk and confidence intervals are reduced. In contrast to typical linear regression using a kernel that calculates the error per sample, SVR allows a maximum variation of between the projected output $f(x_n)$ and the true output $y_n$.

As illustrated in Figure 8, this is comparable to establishing an interval band of width centered on the prediction curve and not counting the samples within the band as a loss.

Therefore, the optimization objective of SVR can be expressed as follows

$$
\min_{w, b} \frac{1}{2} w^T w + C \sum_{i=1}^{m} \epsilon_i (f(x_n) - y_n) 
$$

(7)

where $C$ is the regularization constant and $\epsilon_i$ is referred to as the $\epsilon$-insensitive loss function, expressed in the following form

$$
\epsilon_i = \begin{cases} 
0, & \text{if } |Z| \leq \varepsilon \\
|Z| - \varepsilon, & \text{otherwise}
\end{cases} = \max(0, |Z| - \varepsilon) 
$$

(8)

For the data acquired in this experiment, a compromise point between the prediction result and the input data may be established, and the SVR algorithm obtains the best prediction result. Notably, the SVR algorithm is not affected by the sample input’s dimension.

Random forests. RF is a powerful and adaptable machine learning algorithm. RF regression is composed of many regression trees, each of which is uncorrelated with the others. The RF algorithm employs the principle of minimum mean squared difference during the training process. That is, for any division of feature $A$, corresponding to any
division point S on both sides of the data set $D_1$ and $D_2$, find the feature and eigenvalue division point that minimizes the mean squared difference of the respective sets of $D_1$ and $D_2$, while the sum of the mean squared differences of $D_1$ and $D_2$ is minimized. It is resolved as follows

$$\min_{A,s} \left[ \min_{c_1} \sum_{x_i \in D_1(A,s)} (y_i - c_1)^2 + \min_{c_2} \sum_{x_i \in D_2(A,s)} (u_i - c_2)^2 \right]$$

(9)

where $C_1$ is the sample output mean of $D_1$ data set and $C_2$ is the sample output mean of $D_2$ data set. The regression process of the RF is shown in Figure 9.

To summarize, the RF algorithm predicts using a vote selection method rather than a computational one. This means that as long as the input training set has a sufficiently big sample size, the RF algorithm can produce a satisfactory prediction result.

**Evaluation index**

According to Tang et al.,\textsuperscript{34} the RMSE and Pearson’s correlation coefficient were used to examine the prediction outcomes. The mean square error (MSE) of a parameter is defined as the expected value of the squared difference between its estimated and true values. The RMSE is equal to the arithmetic square root of the MSE. The RMSE value is a measure of how accurate a prediction model is. The following is the RMSE formula

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (Y - X)^2}$$

(10)

Pearson’s correlation, also known as cumulative correlation, is a method for determining linear correlations invented by the British statistician Pearson in the 20th century. The Pearson’s correlation coefficient is used to analyze data that conform to a linear relationship or normal distribution. It can be calculated as

$$\rho_{X,Y} = \frac{\sum (X - \bar{X})(Y - \bar{Y})}{\sqrt{\sum (X - \bar{X})^2}(Y - \bar{Y})^2}$$

(11)

In the above equation, $N$ is the total sample size, $Y$ is the predicted knee angle, $X$ is the true knee angle, $\bar{X}$ is the mean of all predicted knee angles, and $\bar{Y}$ is the mean of all true knee angles. Between 0 and 1, the bigger the RMSE and Pearson correlation coefficient, the more accurately the data predicted.

**The calculation process**

The data were eventually partitioned into a 9:1 training set and a validation set. These data were subsequently used to train the SVR/RF algorithm. Following the completion of the training phase, the validation set is imported for comparison. Python is used to accomplish the calculations in this article. Figure 10 illustrates the computation.
Experimental results

In this chapter, a new data set including 10 different EMG signals and the corresponding knee angles has been composed. Eighteen hundred data sets (360 sets each) were collected from 10 different individuals. Immediately after, these data were imported into the algorithm (RF/SVR) for data prediction, respectively. We investigated the effects of various data processing methods on forecast accuracy and time spent. Additionally, we examined the prediction accuracy and Pearson’s correlation coefficient for different scenarios. The study’s results are presented as the overall mean with standard deviation to illustrate the consistent performance of the algorithm.

To begin, we compared the prediction errors of three distinct processing methods using the RF and SVR algorithms. It can be seen in Figure 11 that the prediction error of the ICA-processed data is less than that of the PCA-processed data in both the RF and SVR algorithms, with an error of roughly 1°. PCA-processed data had a lower prediction error than bandpass-processed data, with an error of roughly 1°. The prediction error in the RF algorithm is around 7° for bandpass-processed data and approximately 6° for PCA-processed data. The prediction error for ICA-processed data is approximately 5°. Similarly, the SVR algorithm has a prediction error of roughly 6° for bandpass-processed data, about 5° for PCA-processed data, and about 4° for ICA-processed data. Moreover, the SVR approach outperforms the RF algorithm in terms of prediction accuracy for bandpass-processed data, PCA-processed...
data, and ICA-processed data. SVR’s prediction error is 1 to 2 times smaller than that of RF for the identical input data.

Additionally, the relative calculation times for data processed by the PCA and ICA algorithms were compared to the calculation durations for data processed by bandpass in the RF/SVR method. Figure 12 compares the calculation times of the RF and SVR algorithms for various data inputs. As can be observed, both the RF and SVR algorithms always reduce the relative calculating time of the data processed by the PCA and ICA algorithms, with the ICA algorithm always having the shortest calculating time. In comparison to bandpass-processed data, ICA-processed data requires only half the calculation time in the SVR algorithm.

The Pearson’s correlation coefficients for the SVR and RF algorithms are shown in Figure 13 for various data inputs. The ICA-processed data always outperforms the PCA-processed data in both RF and SVR algorithms, followed by the bandpass-processed data. Pearson correlation coefficients for the RF and SVR algorithms are nearly identical when the same data set is used. However, the SVR algorithm produces somewhat better results than the RF algorithm.

Finally, Figure 14 visually shows the accuracy of the RF and SVR algorithms in predicting the knee angle. The input data are the ICA-processed data, as it is evident that the ICA-processed data produce the more accurate prediction outcomes in this experiment. According to human biomechanics, the knee joint changes its angle somewhat prior to accomplishing a full range of motion. As can be seen from Figure 14, both algorithms exhibit superior prediction performance. However, the RF algorithm exhibited a lag in joint angle prediction for several stages, whereas SVR demonstrated excellent predictive stability throughout.

Discussion

By comparing the prediction accuracy, computational time consumption, and correlation coefficients under each combination of different algorithms, a new method for predicting knee joint angle using EMG signals is proposed in this article. By and large, both models (RF/SVR) were capable of estimating the knee joint angle from the input EMG signal. However, for all people examined, the approach (ICA + SVR) produced the best results. The prediction accuracy under different conditions was analyzed and compared. From the data processing process (as shown in Figure 7), it can be found that the amplitude range of the signal is different after different processing methods. The amplitude range of data processed by bandpass is the biggest, followed by the amplitude range of data processed by ICA and PCA. It may be stated that the ICA approach effectively extracts the signal’s effective information and eliminates the interference components present in the original signal.

Figure 11 demonstrates that the prediction error of ICA-processed data is always lower than that of PCA- and filtered-processed data, regardless of whether SVR or RF algorithms are used. There could be several possible explanations for this. According to Kiguchi et al., by extracting features, ICA approaches can decrease data sets and separate signals. The observational signal in our study can be thought of as a linear combination of numerous statistically independent components, and the ICA approach is used to separate the source signal from a mixed-signal. To a certain extent, it is more concerned with the independence of EMG signals. Additionally, the PCA method was employed for comparison. Indeed, the process of data processing using the PCA approach is an information extraction process. It decreases the dimension of the raw data based on the data’s contribution to the forecast accuracy. In actuality, if simply
the energy or variance of the data is desired, rather than the noise, the core signal can be recovered using the PCA approach. Due to the Gaussian distribution of the EMG signal, the ICA approach is more effective in capturing the stochastic statistical features of the variables and suppressing Gaussian noise than the PCA method. As a result, the ICA approach outperforms the PCA method in terms of speed and accuracy.

By comparing the performance of the two different algorithms with the same data input, as shown in Figure 11, for all participants in the test, the prediction error of the SVR was about 1° to 3° lower than that of the RF. The possible explanation for this phenomenon is that the SVR algorithm does not consider losses when training and forecasting on data that is within the interval band. The approach begins calculating losses only when the absolute difference between the expected and true outputs is greater than the predetermined value. In the practical application of sEMG in the field of assisted exoskeleton, an error higher than 5° is unacceptable to achieve soft control. Excessive prediction errors might result in ineffective control, so interfering with the task process. From this perspective, ICA-processed data exhibit the most consistent performance (both in the RF/SVR algorithm). When the SVR algorithm is compared to the RF algorithm in terms of Pearson coefficient, as illustrated in Figure 13, the SVR approach is clearly more stable. In short, SVR optimizes the model by increasing the interval band width and decreasing the total loss. On the other hand, the RF algorithm is a selective algorithm. It works by generating several regression trees, randomly selecting a subset of samples from the training set as the root node of each tree, then splitting the trees according to the most appropriate features. In other words, its final output is the result of the joint selection of multiple trees. However, RF may not make the most correct selection when the knee angle variation is small. As shown in Figure 14, the knee angles predicted by the RF algorithm can appear shaky, which is bad in exoskeleton control. To summarize, these selections may be atypical for certain time periods. As a result, we believe this may be the reason why SVR algorithm performs better than the RF algorithm.

While comparing accuracy, the calculation time required to perform the RF/SVR algorithm was also compared. Figure 12 compares the RF/relative SVR’s calculation time for various data inputs. Take note that the ICA algorithm is always the quickest to calculate, both in the RF and SVR algorithms. The calculating time for data processed by the PCA/ICA method (in SVR) is about two times shorter than that for bandpass-processed data since the ICA algorithm can reproduce the original signal (with less noise and interference) to the maximum extent possible. To summarize, ICA’s “Blind Source Separation Technique” reduces convergence time and increases prediction accuracy.

In conclusion, this study’s objective is to determine the effect of signal preprocessing on prediction results. The link between the sEMG signal and the human joint angle is further investigated to develop a high-precision exoskeleton EMG controller in future study. It enables the use of sEMG signals to control the exoskeleton while offering intelligent support to be straightforward and practical. Therefore, recognizing different movement intentions under various conditions and issuing commands to assist the wearer in completing the task is a point worth investigating. In this article, however, only the walking state of the movement is analyzed, which is not enough.

Conclusions

The goal of this work is to investigate the effect of preprocessing the sEMG signal on prediction findings, laying the groundwork for the construction of an accurate and responsive exoskeleton robot controller. The purpose of this article is to offer a novel method (ICA + SVR) for predicting the knee joint angle during human motion. Ten individuals’ experimental data were collected to demonstrate the method’s superiority. It has been demonstrated that the ICA algorithm...
is capable of both noise removal and signal isolation from EMG signals. The method reduces undesirable effects caused by volume conduction, such as cross talk and signal loss in the tissue, as well as unknown sources of noise. By comparing the prediction outcomes under various scenarios, it can be determined that the combination of ICA and SVR produces the best results while remaining computationally efficient. Additionally, the data analysis demonstrates that the ICA algorithm can significantly reduce the time required for the SVR algorithm to produce predictions.

A possible direction for future work in this area is that the types of movements studied should be increased including different poses and states. Correlating force signals, movement signals, and sEMG signals is another area that should be researched in the future. Finally, due to the complexity of the human musculoskeletal model, choosing the best effective place for EMG signal capture is a research and attention-seeking topic.

Acknowledgments
The authors thank the experiment participants and Nanjing University of Science and Technology for their support in this study.

Author contributions
Conceptualization and software: Meng Zhu and XiaoRong Guan; Methodology: Meng Zhu, XiaoRong Guan, and Zhong Li; Validation: Zhong Li; Formal analysis, investigation, data curation, writing—original draft, and visualization: Meng Zhu; Resources: YunLong Gao, KaiFan Zou, XinAn Gao, Zheng Wang, and Hui-Bin Li; Supervision: KeShu Cai.

Declaration of conflicting interests
The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding
The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research was funded by National Defense Basic Scientific Research Program of China, grant number JCKY2019209B003.

ORCID iD
XiaoRong Guan https://orcid.org/0000-0001-5493-3699

References
1. Woo H, Kong K and Rha DW. Lower-limb-assisting robotic exoskeleton reduces energy consumption in healthy young persons during stair climbing. Appl Bionics Biomech 2021. https://doi.org/10.1155/2021/8833461
2. Yamanaka Y, Kashima M, Arakawa H, et al. Verification of the “AB-Wear” semi-exoskeleton-type power-assist suit in providing assistance to the lower back. In: 2021 22nd IEEE international conference on industrial technology (ICIT), Valencia, Spain, 10–12 March 2021, 2021.
3. Zhou X, Liu G, Han B, et al. Design of a human lower limbs exoskeleton for biomechanical energy harvesting and assist walking. Energy Technol 2021; 9(1): 2000726.
4. Tu Y, Zhu AB, Song JY, et al. Design and experimental evaluation of a lower-limb exoskeleton for assisting workers with motorized tuning of squat heights. In: IEEE transactions on neural systems and rehabilitation engineering (Vol. 30). Piscataway, NJ: IEEE, 2022, pp. 184–193.
5. Mudie K, Billing D, Garofolini A, et al. The need for a paradigm shift in the development of military exoskeletons. Eur J Sport Sci 2022; 22(1): 35–42.
6. Proud JK, Lai D, Mudic KL, et al. Exoskeleton application to military manual handling tasks. Hum Factors 2020; 64(3): 527–554.
7. Liu H, Tao J, Lyu P, et al. Human-robot cooperative control based on sEMG for the upper limb exoskeleton robot. Robot Auto Syst 2020; 125: 103350.
8. Amamcherla N, Turlapaty A and Gokaraju B. A machine learning system for classification of EMG signals to assist exoskeleton performance. In: 2018 IEEE applied imagery pattern recognition workshop (AIPR), Washington, DC, USA, 09–11 October 2018, 2018.
9. Asif AR, Waris A, Gilani SO, et al. Performance evaluation of convolutional neural network for hand gesture recognition using EMG. Sensors 2020; 20(6): 1642.
10. Fan Y and Yin Y. Active and progressive exoskeleton rehabilitation using multisource information fusion from EMG and force-position EPP. IEEE Trans Biomed Eng 2013; 60(12): 3314–3321.
11. Chaparro-Rico B, Cafolla D, Tortola P, et al. Assessing stiffness, joint torque and ROM for parietic and non-parietic lower limbs during the subacute phase of stroke using Lokomat tools. Appl Sci 2020; 10(18): 6168.
12. Chaparro-Rico B and Cafolla DJAS. Test-retest, inter-rater and intra-rater reliability for spatiotemporal gait parameters using SANE (an eaSy gAit aNalysis systEm) as measuring instrument. Appl Sci 2020; 10(17): 5781.
13. Turker H. Electrodiagnosis in new frontiers of clinical research. Norderstedt: BoD – Books on Demand, 2013.
14. Bi L, Feleke AG and Guan C. A review on EMG-based motor intention prediction of continuous human upper limb motion for human-robot collaboration. Biomed Signal Process Contr 2019; 51: 113–127.
15. Mukhopadhyay AK and Samui S. An experimental study on upper limb position invariant EMG signal classification based on deep neural network. Biomed Signal Process Contr 2020; 55: 101669.
16. Au AT and Kirsch RF. EMG-based prediction of shoulder and elbow kinematics in able-bodied and spinal cord injured individuals. IEEE Trans Rehabil Eng 2000; 8(4): 471–480.
17. Tsuji T, Shibano T, Shima K, et al. A novel five-fingered myoelectric prosthetic hand and its muscle synergy-based control. Bullet Japanese Soc Prosth Ortho 2010; 26(2): 91–96.
18. Kiatpanichajj K and Afzulpurkar N. Use of supervised discretization with PCA in wavelet packet transformation-based surface electromyogram classification. Biomed Signal Process Contr 2009; 4(2): 127–138.
19. Vijayvargiya A, Kumar R and Dey N. Voting-based 1D CNN model for human lower limb activity recognition using sEMG signal. *Phys Eng Sci Med* 2021; 44(4): 1297–1309.
20. Tapia C, Daud O and Ruiz-del-Solar J. EMG signal filtering based on independent component analysis and empirical mode decomposition for estimation of motor activation patterns. *J Med Biolog Eng* 2017; 37(1): 140–155.
21. Li Z, Guan X, Zou K, et al. Estimation of knee movement from surface EMG using random forest with principal component analysis. *Electronics* 2019; 9(1): 43.
22. Kilner JM, Baker SN and Lemon RN. A novel algorithm to remove electrical cross-talk between surface EMG recordings and its application to the measurement of short-term synchronization in humans. *J Physiol* 2010; 538(Pt 3): 919–930.
23. Li YC, Wang PT, Vaidya MP, et al. Electromyogram (EMG) removal by adding sources of EMG (ERASE)-A Novel ICA-based algorithm for removing myoelectric artifacts from EEG. *Front Neurosci* 2021; 14: 597941.
24. Kim Y, Stapornchaisit S, Miyakoshi M, et al. The effect of ICA and non-negative matrix factorization analysis for EMG signals recorded from multi-channel EMG sensors. *Front Neurosci* 2020; 14:1254.
25. Hassani M and Karami MR. Noise estimation in electroencephalogram signal by using Volterra series coefficients. *J Med Signals Sensors* 2015; 5(3): 192.
26. Kiguchi K, Rahman MH, Sasaki M, et al. Development of a 3DOF mobile exoskeleton robot for human upper-limb motion assist. *Robot Auto Syst* 2008; 56(8): 678–691.
27. Sato M and Yagi E. A study on power assist suit using pneumatic actuators based on calculated retaining torques for lift-up motion. In: *SICE Annual Conference 2011*, Tokyo, Japan, 13–18 September 2011, 2011.
28. Chen DS, Ning M, Zhang BG, et al. Control strategy of the lower-limb exoskeleton based on the EMG signals. In: *2014 IEEE International Conference on Robotics and Biomimetics (ROBIO 2014)*, Bali, Indonesia, 05–10 December 2014, 2014.
29. Gordleeva SY, Lobov SA, Grigorev NA, et al. Real-time EEG-EMG human-machine interface-based control system for a lower-limb exoskeleton. *IEEE Access* 2020; 8: 84070–84081.
30. He H and Kiguchi K. A study on EMG-based control of exoskeleton robots for human lower-limb motion assist. In: *2007 6th International Special Topic Conference on Information Technology Applications in Biomedicine*, Tokyo, Japan, 08–11 November 2007, 2007.
31. Kiguchi K and Imada Y. EMG-based control for lower-limb power-assist exoskeletons. In: *2009 IEEE workshop on robotic intelligence in informationally structured space*, Nashville, TN, USA, 30 March 2009–02 April 2009, 2009.
32. Leardini A, Sawacha Z, Paolini G, et al. A new anatomically based protocol for gait analysis in children. *Gait Posture* 2007; 26(4): 560–571.
33. Cappello A. ABC of EMG. 2010.
34. Tang Z, Zhang K, Sun S, et al. An upper-limb power-assist exoskeleton using proportional myoelectric control. *Sensors (Basel)* 2014; 14(4): 6677–6694.