Journal of Ambient Intelligence and Humanized Computing
https://doi.org/10.1007/s12652-019-01390-3

ORIGINAL RESEARCH

Personalized gait trajectory generation based on anthropometric features using Random Forest

Shixin Ren[1,2] · Weiqun Wang[1,2] · Zeng-Guang Hou[1,3] · Badong Chen[4] · Xu Liang[1,2] · Jiaxing Wang[1,2] · Liang Peng[1,2]

Received: 3 February 2019 / Accepted: 4 July 2019

Abstract
Using lower limb rehabilitation robots (LLRRs) to help stroke patients recover their walking ability is attracting more and more attention presently. Previous studies have shown that gait rehabilitation training with natural gait pattern can improve the therapeutic outputs. However, how to generate the personalized gait trajectory has not been well researched. In this paper, a personalized gait generation method based anthropometric features is proposed. Firstly, gait trajectories are fitted and simplified into Fourier coefficient vectors, which are used to represent gait trajectories. Secondly, fourteen body features are used to generate the personalized gait trajectories and the feature set is further optimized based on the minimal redundancy maximal relevance criterion for easy application on the LLRR. Then, the relationship between the optimized feature set and gait trajectories is modeled by using the RF algorithm. Finally, the performance of the proposed method is demonstrated by several comparison experiments.

Keywords Personalized gait · Gait generation · Random Forest · Anthropometric features · Rehabilitation training

1 Introduction

Stroke and spinal cord injury are leading causes of the motor dysfunction, which has a long-term impact on patients' mobility, muscle control ability and gait patterns (Winnen et al. 2014). It has been shown that patients' motor function can be improved by rehabilitation training (Barbeau et al. 1987; Niu et al. 2014; Wirz et al. 2011). Furthermore, the earlier the rehabilitation training is carried out, the better the rehabilitation results will be (Bernhardt et al. 2008; Hu et al. 2010). In traditional rehabilitation training, each patient needs one or two physiotherapists' assistant. With the increase of patients, the number of physiotherapists is difficult to meet the requirements by the traditional method. At the same time, the intensive labor is a heavy load for physiotherapists. Hence, a large number of patients cannot get effective rehabilitation.

Rehabilitation robots can be used to help patients do rehabilitation training with the advantages of repeatability, reliability and intellectualization (Calabro et al. 2016; Chen et al. 2013). With robotics' assist, patients can perform more daily living activity training (Song 2016), and the same and even better therapeutic outputs can be obtained by rehabilitation robots (Kwakkel et al. 2008; Amirabdollahian et al. 2007; Hogan et al. 2006). Under the assistant of lower limb

✉ Weiqun Wang
weiqun.wang@ia.ac.cn

Shixin Ren
renshixin2015@ia.ac.cn

Zeng-Guang Hou
zengguang.hou@ia.ac.cn

Badong Chen
chenbd@mail.xjtu.edu.cn

Xu Liang
liangxu2013@ia.ac.cn

Jiaxing Wang
wangjiaxing2016@ia.ac.cn

Liang Peng
liangpeng@ia.ac.cn

1 The State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China

2 University of Chinese Academy of Sciences, Beijing 100049, China

3 The CAS Center for Excellence in Brain Science and Intelligence Technology, Beijing 100190, China

4 Institute of Artificial Intelligence and Robotics, Xi'an Jiaotong University, Xi'an 710049, China

Published online: 20 July 2019

rehabilitation robotics (LLRRs), the legs of patients can be trained in several actions, such as walking and cycling, to prevent muscle atrophy and strengthen control ability of muscle.

Walking is a very important ability in daily life, and plenty of LLRRs focus on walking training. For example, the famous treadmill training system, Lokomat (Colombo et al. 2000), is designed to help stroke patients do walking training by two exoskeleton legs. Patients with spinal cord injury can stand up and walk with assistant of eLEGS, which passed the certification by the United States Food and Drug administration in 2012 (Ming et al. 2017).

It has been indicated that training with the reasonable and personalized training trajectory, the patients' participation and the efficiency of rehabilitation can be improved (Vallery et al. 2009). Human gait trajectories are definitely personalized (Murray 1967). Thus, the training trajectory provided by LLRRs must be suitable for patients and adjusted according to patients' features. However, present LLRRs usually cannot provide the personalized gait trajectory and only rough methods to replicate the gait trajectories are applied (Colombo et al. 2000; Wang et al. 2005). Lokmat, as an example, only allows therapists to adjust the gait trajectory according to patients' height (Aoyagi et al. 2007). Hence, how to generate personalized trajectory based on patients' features is an important research topic.

In this paper, a personalized gait trajectory generation method based on Random Forest (RF) algorithm is proposed to model the relationship between anthropometric features and gait trajectory. In order to obtain the low computational load and the convenience of control, the finite Fourier series (FFS) are adopted to fit the discrete joint trajectories. The obtained Fourier coefficient vectors, are taken as the representatives of gait trajectories. Meanwhile, for the purpose of easy application on LLRRs, fourteen body features are used and optimized by using the minimal redundancy maximal relevance criterion, which is based on the mutual information between inter-body features and Pearson correlation coefficients between features and Fourier coefficients. The experiment results show that the performance of gait prediction model based on the optimized feature set is better than that based on feature set adopted by Luu et al. (2011). Besides, the gait generation model based on RF has shown better prediction accuracy than other four regression prediction models.

The remainder of this paper is organized as follows. Section 2 summarizes the works relating to gait generation. A lower limb rehabilitation robot developed at Institute of Automation, Chinese Academy of Sciences and the gait data are introduced in Sect. 3. Section 4 describes the gait trajectory fitting method and the feature subset optimization and gait generation methods. The comparison results of the RF models based on different feature subsets, and the performance comparison between RF model and other four trajectory generation models are given in Sect. 5. Finally, the discussion and conclusion of this paper are presented in Sect. 6.

2 Related work

In the last decade, many studies on the gait generation methods have been carried out. Some studies are applied mainly for hemiplegic patients. For example, Vallery et al. (2009) proposed an approach, named as complementary limb motion estimation (CLME), to ensure stable gait for the LOPES rehabilitation robot. The gait trajectory of patient's affected limb could be generated according to the healthy limb. The experiment results indicated that walking training with CLME was relatively natural. Liu et al. (2016) adopted a long-short term memory model trained by the collected normal gait data, to predict and correct knee joint trajectory based on other joints. The pre-collected gait data was collected via the encoders of lower-limb exoskeletons. Although these methods can obtain good results, they are not suitable for patients with two affected legs.

Some results can be applied for both hemiplegic and paraplegia patients. For example, Feng et al. (2008) simplified each human lower limb into two-links, and hence, two lower limbs formed a four-links model with a common connected point on the hip joint. Based on the constraints between swim limb tip and hip joint, they modeled a constraint function to calculate trajectories of limb tips. Finally, the angle trajectories of knee and hip joints can be obtained by limb tips' trajectories, and also adjusted by three constraint parameters. Luu et al. (2010) proposed a multiple linear regression model based on FFS for gait prediction. Stride length and cadence were used as the inputs of regression models. Furthermore, Luu et al. also adopted the multilayer neural network and the general regression neural network algorithm to learn the relationship between Fourier coefficients and some features including stride parameters and features of human legs. Koopman et al. (2014) selected six key events to describe individual's gait patterns in a gait cycle. Then linear models were established to predict timing, angle, angle velocity and acceleration for each key event based on the walking speeds and the subject's height. When the specified walking speed and height were given, dynamic parameters of six key event could be predicted. Then, a new gait trajectory could be reconstructed by these six key events. For the above researches, the body features were taken into account in gait generation models, but the number of body features are still not enough for accurate estimation of the personalized gait trajectories.

Yun et al. (2013) applied a Gaussian process regression algorithm (GPR) to build relationship model between multiple body features and human gait trajectories. Fourteen body features recorded from healthy subjects were selected as the inputs of GPR model and the discrete joint trajectory points were directly used as the outputs of GPR model. Although multiple body feature are taken account into GPR model, it will take more than one or two days to optimize the hyperparameters of GPR model. Furthermore, when the training gait data are changed, the hyperparameters need to be optimized.

At present, deep learning methods have shown tremendous potential and have achieved good results in many fields. However, the difficulty of collecting gait trajectories data leads to small gait trajectory data. The application of deep learning methods is limited by small data in the gait generation.

3 Lower limb rehabilitation robot and gait data

3.1 A lower limb rehabilitation robot

Because the disease severity and onset time of patients are different with each other, the required rehabilitation training strategies also should be different for each patient (Meng et al. 2015). While the rehabilitation training provided by each LLRR is only suitable for patients at specific rehabilitation phases. Therefore, a multi-pose LLRR has been developed at Institute of Automation, Chinese Academy of Sciences recently. Based on this LLRR, multiple training modes can be provided for patients at different rehabilitation phases. For example, patients can carry out cycle training at early rehabilitation phases. At middle and later rehabilitation phases, patients can perform gait training and up and down stair activity training.

As shown in Fig. 1, the LLRR is in gait rehabilitation training mode. Each lower limb exoskeleton has three active joints include hip, knee and ankle joints, which are driven by DC motors. The length of lower limb exoskeletons' thigh and calf can be adjusted according to patients' height. In order to ensure the safety of patients, the maximum scope of each joint is restricted in the mechanical structure. At the same time, two limited switches are used to detect the maximum angle of each joint. Meanwhile, the body weight supported system (BWS) is designed to load partial weight of patient. With the help of BWS, patients can be assured of adequate safety in gait rehabilitation training. Previous studies have shown that with the natural and personalized training trajectory, the efficiency of rehabilitation training can be improved (Vallery et al. 2009). Therefore, how to generate natural and personalized gait trajectories based on this LLRR will be studied in the next contents.

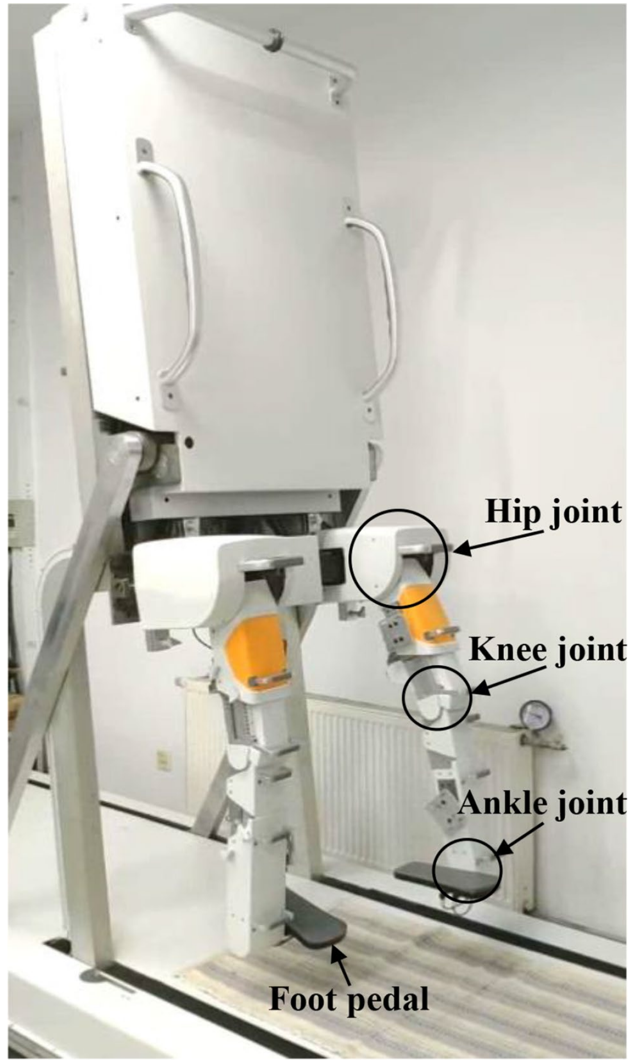

Fig. 1 Lower limb rehabilitation robot developed by Institute of Automation, Chinese Academy of Sciences

3.2 Gait data

In this paper, the body features and gait trajectory data are from Yun et al. (2013). And the feature set consists of twelve anthropometric parameters, age and gender. The ranges of the body features are shown in Table 1, and the definition of some body features can be found in Chandler et al. (1975).

The angle definitions of three joints are shown in Fig. 2. The angle between the centerline of thigh and the line of gravity of the body is hip joint angle θ_h, the angle between thigh's centerline and calf's centerline is knee joint θ_k. l_1 is the line that is perpendicular to calf's centerline, and l_2 is the line connecting toe and center of ankle joint. θ_a' is the angle between l_1 and l_2, and the ankle angle $\theta_a = \theta_a' - c$, where c is the value of θ_a' at standing position. Because the gait trajectory is applied on LLRR, there must be no

Table 1 Ranges of the body features

Features	Ranges
Age (years old)	20–69
Bi-iliac width (cm)	26.1–35.8
Height (cm)	149.0–185.0
ASIS breath (cm)	20.0–30.6
Mass (kg)	43.3–99.0
Knee diameter (cm)	8.2–13.0
Gender	F/M(0/1)
Foot length (cm)	20.5–28.0
Thigh length (cm)	27.5–41.6
Malleolus height (cm)	5.2–9.0
Calf length (cm)	30.5–46.3
Malleolus width (cm)	5.5–8.0
Bi-trochanteric width (cm)	28.8–38.6
Foot breath (cm)	6.4–11.0

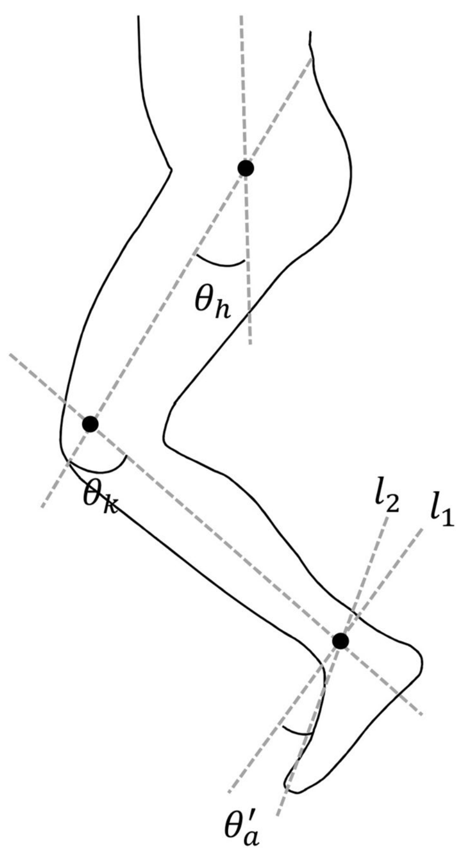

Fig. 2 Angle definitions of hip joint, knee joint and ankle joint

trajectories which does not comply with structure design constraints. Hence, the number of retained samples is eighty.

4 Method

4.1 Fitting of the joint trajectory

The angle variations of the joint are continue and smooth in walking, so the similar angle variations should be provided by the LLRR. But the collected joint trajectory points are discrete, it's not applicable for motor control on the LLRR. At the same time, it can be seen that the joint trajectories have one or two peaks and troughs. Hence, the FFS are adopted to fit joint angle trajectories, as follows:

$$f(t_n) = a_0 + \sum_{i=1}^{n} (a_i \cos(i\omega t) + b_i \sin(i\omega t)), \quad i = 1, \ldots, n, \tag{1}$$

where $\omega = \frac{2\pi}{T}$, T is the period of the gait pattern. n is the order of FFS, a_i and b_i are Fourier coefficients.

The precise of fitting is closely related to the order n. In order to select the optimal order n, the orders from 1 to 8 are used to fit three joint trajectories. The mean absolute deviation (MAD) is calculated for evaluation:

$$MAD = \frac{1}{m} \sum_{j=1}^{m} \left| d_j^* - d_j \right|, \quad j = 1, \ldots, m. \tag{2}$$

where m is the number of joint trajectory sampling points. d_j^* and d_j are sampling points of the actual joint trajectory and the fitting joint trajectory, respectively.

As shown in Fig. 3, with the increase of Fourier order, the MAD decreases gradually and becomes relatively stable when the order is five. It also can be found in the Fig. 4 that

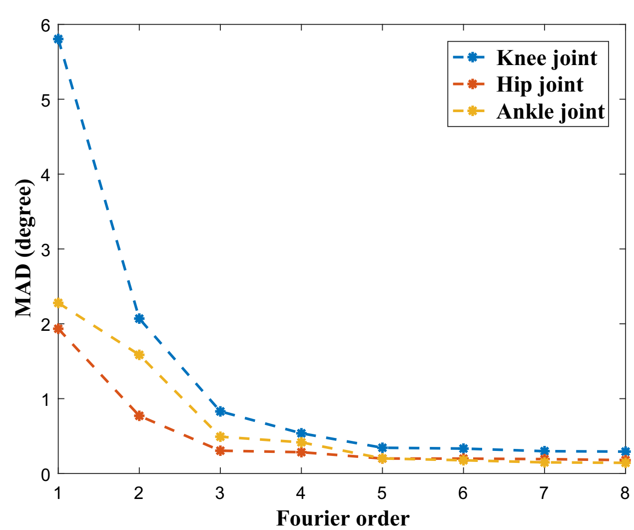

Fig. 3 The MAD between actual trajectories and fitting trajectories using different Fourier order

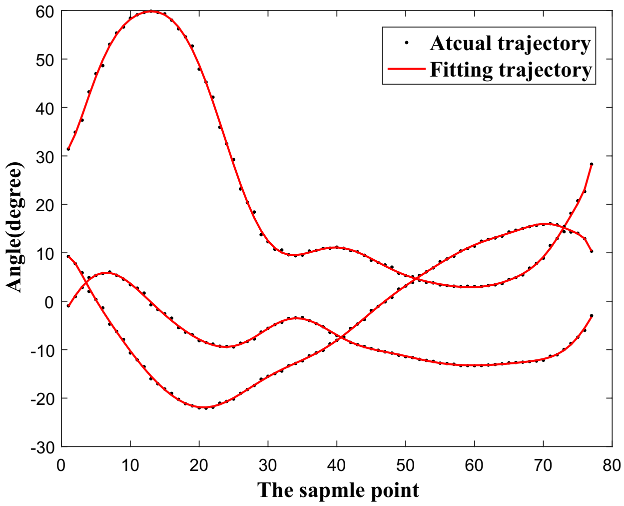

Fig. 4 Actual joint trajectories and fitting joint trajectories

the fitting results of three joint trajectories are quite good when the order is five. Therefore, the five order FFS are adopted to fit joint trajectories. By Eq. (1), 11 Fourier coefficients of each joint trajectory can be calculated, as follows:

$$\mathbf{Y}_{(i,j)} = (a^0_{i,j}, a^1_{i,j}, b^1_{i,j}, a^2_{i,j}, b^2_{i,j}, a^3_{i,j}, b^3_{i,j}, a^4_{i,j}, b^4_{i,j}, a^5_{i,j}, b^5_{i,j}) \tag{3}$$

where $i = 1, 2, \ldots, P$. P is the number of subjects. j is equal to 1, 2, 3, corresponding to the hip, knee and ankle joint, respectively. In order to reduce the computational load, each joint trajectory can be represented by a Fourier coefficient vector which consists of 11 coefficients. The Fourier coefficient vector is to be used as the output of the gait generation model.

4.2 Feature optimization based on the minimal redundancy maximal relevance criterion

Selecting an optimized feature set, in which features are closely related with the gait trajectory, can improve the performance of gait trajectory generation model. Moreover, reducing the number of redundancy feature is also helpful to relieve the load of body feature measurement. For example, when a patient firstly participates in gait rehabilitation training, it will take little time to generate a personalized gait trajectory for him/her.

So, the minimal redundancy maximal relevance criterion is applied to optimize the feature data set. There are two optimization principles: one is that the redundancy information of inter-features is least in the optimized feature subset, another one is that the selected features must have high correlation with Fourier coefficients.

For the first principle, the mutual information (MI) are used to evaluate the redundancy (Peng et al. 2005; Ding and Peng 2005). And the mutual information of two feature sets F_m and F_n is calculated, as follows:

$$\begin{aligned} I(F_m, F_n) &= \sum_i^P \sum_j^P p(f_{m,i}, f_{n,j}) \log\left(\frac{p(f_{m,i}, f_{n,j})}{p(f_{m,i})p(f_{n,j})}\right), \\ &0 \le m,\ n \le N, m \ne n,\ f_{m,i} \in F_m,\ f_{n,i} \in F_n. \end{aligned} \tag{4}$$

where N is the number of features. P is the number of subjects. $p(f_{m,i})$ and $p(f_{n,j})$ are the marginal probability functions and $p(f_{m,i}, f_{n,j})$ is joint probability distribution function. For a feature set D_N consisting of N features $(F_1, \ldots, F_N)$, the mutual information of set D_N is defined by:

$$I(D_N) = \frac{1}{N^2} \sum_{m=1}^N \sum_{n=1}^N I(F_m, F_n). \tag{5}$$

For the second principle, the Pearson correlation coefficient is adopted for the criterion of the relevance information. Moreover, the relevance between the features F_m and Fourier coefficients Y^c can be defined by:

$$R(F_n, Y^c) = \frac{\sum_{i=1}^I (f_{n,i} - \bar{F}_n)(y^c_i - \bar{Y}^c)}{\sigma_{F_n}\sigma_{Y^c}}, \quad y^c_i \in Y^c. \tag{6}$$

where $c = 1, 2 \ldots 11$. σ_{F_n} and σ_{Y^c} are standard deviations. The relevance between features set D_N and Fourier coefficients Y^c can be calculated by:

$$R(D_N) = \frac{1}{N} \sum_{n=1}^N R(F_n, Y^c). \tag{7}$$

By minimizing Eq. (5) and maximizing Eq. (7), the above two principles can be guaranteed. And an optimization criterion can be obtained through the combination of Eqs. (5) and (7), as follow:

$$W = I(F_n, Y^c) - \sum_m^{k2} I(F_n, F_m), \quad F_n \in D_{k1}, \quad F_m \in D_{k2}. \tag{8}$$

where D_{k1} is an already selected feature subset with $k1$ features, D_{k2} is an unselected feature subset with $k2$ features $(N = k1 + k2)$. One feature can be selected from D_{k2} into D_{k1} by maximizing Eq. (8) each time. Therefore, the feature importance ranking can be obtained according to the selection sequence. The specific process can be found in Algorithm. 1.

Algorithm 1 Feature selection based on the minimal redundancy maximal relevance criterion for a Fourier coefficient

Input: Feature set D_N, Fourier coefficient set Y^c
Output: The feature importance ranking for Y^c

Initialize D_{k1} to an empty set.
for each $m \in [1, 14]$ **do**
  $R(m) \leftarrow$ the relevance of F_m and Y^c;
end for
$k1 = 1$, $F_{k1} \leftarrow \underset{F_m \in D_N}{max}\left[R(m)\right]$;
Put the F_{k1} into D_{k1};
for each $k1 \in [2, 14]$ **do**
  $K1 \leftarrow$ the number of D_{k1};
  $K2 = N - K1$;
  for each $k2 \in [2, K2]$ **do**
    $R(k2) \leftarrow$ the relevance of F_{k2} and Y^c;
    $I(k2) \leftarrow$ the average MI of F_{k2} and
        every F_{k1}(in D_{k1});
  end for;
  $F_{k1} \leftarrow max\left[R - I\right]$;
  Put the F_{k1} into D_{k1};
end for;

For Fourier coefficient vector of each joint, there are 11 feature importance rankings corresponding to every Fourier coefficient, and the 11 feature importance rankings are merged into a ranking. According to property of FFS, the amplitude of each frequency item is decided by the corresponding coefficient (Tolstov 2012). Therefore, the mean absolute value of each Fourier coefficient is adopted to determine the corresponding weight value of feature importance ranking in the average process. Then, the feature importance ranking for three joints is obtained through mean way. The overall flow chart of feature importance ranking is shown in Fig. 5.

4.3 The gait generation model based on Random Forest algorithm

Random Forest (RF) is a machine learning algorithm that combines the advantages of Bagging and Decision trees in classification or regression (Breiman 2001). It is good interpretable, concise, and robust. Meanwhile, in the RF algorithm, the number of parameters that should be adjusted is relatively small. Due to the good performance of RF algorithm, it has been adopted in many fields (Wager and Athey 2017; Malekipirbazari and Aksakalli 2015; Youssef et al. 2016). RF model is an ensemble of decision trees which can map discrete inputs to continue outputs for regression. The original difficult problem can be split into many smaller and simpler problems by each decision tree. At the same time, the generation and robustness of RF model can be improved by Bagging. Hence, the complex relationship between inputs and outputs can be learned better. So, RF algorithm is applied to describe the relationship of body features and Fourier coefficients.

In this paper, there is a training dataset $S_{N\times P}$, which consists of feature dataset $D_{N\times P}$ and Fourier coefficient dataset $Y_{C\times P}$. P is the number of subjects. N and C are the number of body features and Fourier coefficients, respectively. The process of RF algorithm designed is as follows:

- According to the presetting number of T decision trees, $T(T < P)$ sample subsets are randomly extracted from dataset $S_{N\times P}$ based on the bootstrap method.
- Building T regression trees based on T sample subsets. For each node of every regression tree, $K(K < N)$ features are randomly selected from N features and form a subset Θ_K. Then, according to optimization rules, the optimum feature of Θ_K is selected to split data into two sets. Based on the above way, every regression tree continues to grow until termination conditions are reached.
- Let the response value of a tree for an input sample x be $f_t(x)$, the output value of the RF model can be given, as follows:

$$H(x) = \frac{1}{T}\sum_{t=1}^{T} f_t(x). \tag{9}$$

In the training process of RF model, the final RF model's performance has high correlation with two parameters (Liaw and Wiener 2002): *ntree* (number of the trees), and *mtry* (number of the features in the subset Θ_K). The value of mtry is p / 3 recommended by Breiman (2001). For parameter *ntree*, the N-fold cross validation method (Kohavi 1995) is used to determine its value.

5 Results

5.1 Results of feature optimization

In Sect. 4.2, the feature importance ranking for one joint can be obtained through weighted mean method, and the feature importance ranking for three joints is also obtained through mean method. The mean feature importance ranking is shown in Table 2.

In order to select the optimal feature subset, all feature data is divided into fourteen subsets. The rules of division are as follows: The first feature subset D_{s1} consists of the

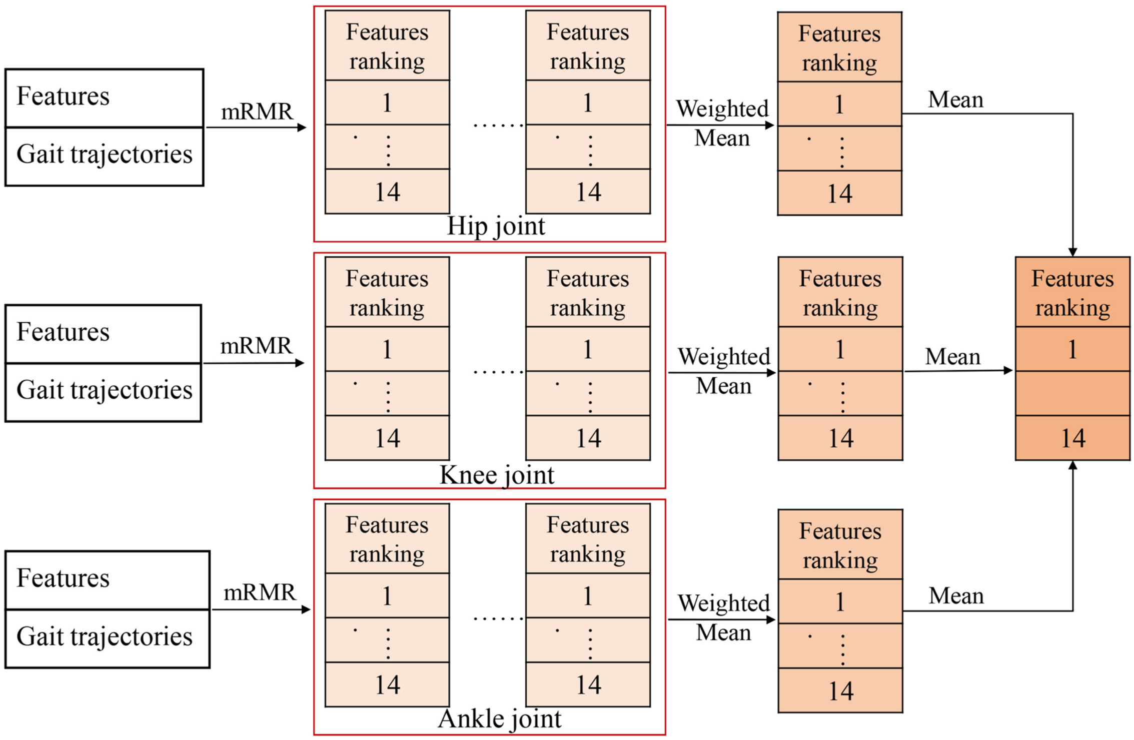

Fig. 5 The overall flow chart of features importance ranking

Table 2 The importance ranking of anthropometric features based on mRMR

Features	Hip joint	Knee joint	Ankle joint	Mean order
mas	1	9	2	1
thigh	3	8	1	1
gender	7	2	4	3
calf	2	3	8	3
bi-t	4	1	9	5
ASIS	5	6	3	5
age	9	4	5	7
bi-i	6	7	6	8
height	10	5	7	9
Knee	8	10	11	10
Footlength	11	11	10	11
Malleolusheight	12	12	13	12
Footbreath	13	14	12	13
Malleoluswidth	14	13	14	14

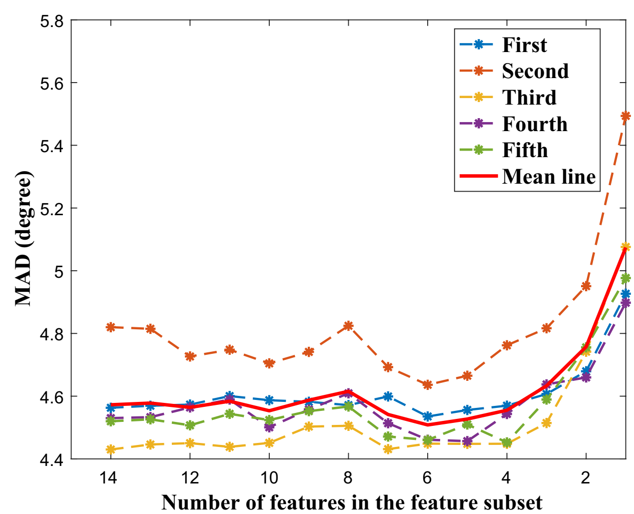

Fig. 6 The MAD of RF models based on different feature subsets for hip joint

first feature in the order of feature importance ranking; The second subset D_{s2} consists of the first and second features in the order of features importance ranking, and so on. By adding one feature into the subset each time, fourteen feature subsets can be finally obtained.

Based on each one of the fourteen feature subsets, there are fourteen RF models trained individually. Fivefold method is adopted to test the performance of each RF model, and MAD is used as the criterion of evaluation. The results of each fold are shown in Fig. 6. It can be found that MAD is the smallest based on the feature subset including the first six features for hip joint. And the similar trends are also found for ankle and knee joints.

5.2 Training and optimization of the RF model

The gait dataset is randomly divided into training set (75%) and testing set (25%). The six features and 11 Fourier coefficients are used as the inputs and outputs of models, respectively. The performance of RF model is influenced by two parameters: *ntree*, and *mtry*. The value of *mtry* is 3, which is recommended by Breiman (2001). For another parameter *ntree*, the numerical searching method is applied to determine the parameter. The search range of *ntree* is from 50 to 600, and the interval is 50. The optimization results of *ntree* are shown in Fig. 7, it can be found that the MAD is no longer reduced with the number of trees increasing from 200 trees. Therefore, the parameter *ntree* of RF model is set to 200.

5.3 Comparison of the RF models based on different feature sets

The optimized feature subset has been obtained in Sect. 5.1. Based on this feature subset, gait trajectory generation models of Fourier coefficients can be obtained through training. And predicted Fourier coefficients are used to reconstruct gait trajectories further.

In order to verified whether the optimized feature subset D_1 is more helpful to the generation of gait trajectory, the feature subset D_2 adopted by Luu et al. is also used to generate the gait trajectory (Luu et al. 2011). In the feature subset D_2, there are six features including four anthropometric features and two gait parameters (stride length and cadence). Four body features can be obtained directly in this paper, but two gait parameters can only be obtained by calculation, as follow:

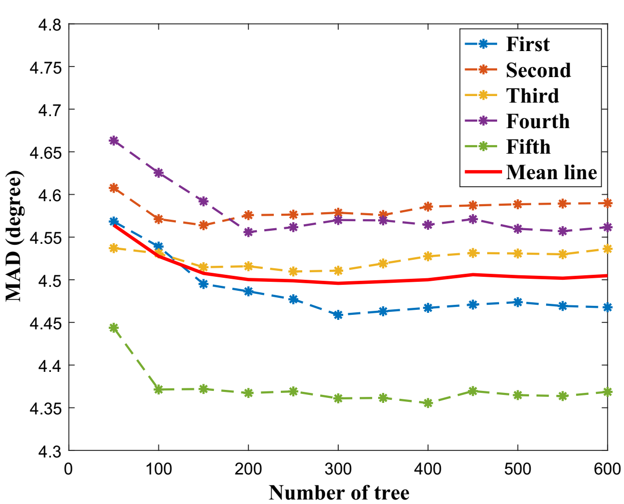

Fig. 7 The MAD of the RF models at each folder of fivefold cross validation for hip joint

Table 3 The performance comparison by using feature dataset D_1 and D_2

Joint	D_1		D_2	
	e	*c*	*e*	*c*
Ankle	4.74	0.76	4.91	0.74
Knee	7.38	0.91	7.41	0.90
Hip	4.60	0.93	4.81	0.92

$$Cadence = 2 \times \frac{1}{T}, \tag{10}$$

$$Length = \frac{V}{T}. \tag{11}$$

where T is the period of a gait. V is the walking speed.

Therefore, two RF models based on feature dataset D_1 and D_2 are obtained. Fivefold cross validation is used to evaluate the performance of two models. The comparison results are shown in Table 3. The *e* and *r* are average values of MAD and Pearson correlation coefficients of fivefold cross validation for five times, respectively. The results show that MAD of RF models based on D_1 is lower than that based on D_2 for all three joints, and the correlation coefficients of RF models based on D_1 are more higher. It also indicates the optimized feature subset in this paper is more helpful to improve the performance of gait trajectory generation models based on RF. Moreover, the actual and predicted trajectories of three joints based on D_1 and D_2 are shown in Fig. 8. It can be found that the predicted trajectories are closer to actual trajectories.

5.4 Comparison of the different gait trajectory generation models

In order to validate the performance of gait trajectory generation models based on RF, multiple linear regression (MLR) model method (Luu et al. 2010) is used for comparison. Based on MLR, the relationship between body features and Fourier coefficient vector is described, as follows:

$$\mathbf{Y} = \begin{bmatrix} a_0 \\ \vdots \\ a_5 \\ b_1 \\ \vdots \\ b_5 \end{bmatrix} = \begin{bmatrix} 0 & \cdots & u_0^6 \\ 0 & & \\ \vdots & \cdots & \vdots \\ u_5^0 & \cdots & u_5^6 \\ u_6^0 & \cdots & u_6^6 \\ \vdots & \cdots & \vdots \\ u_{10}^0 & \cdots & u_{10}^6 \end{bmatrix} \begin{bmatrix} 1 \\ fea_1 \\ \vdots \\ fea_6 \end{bmatrix} = \mathbf{UF} \tag{12}$$

where $\mathbf{U}$ is 11×7 coefficients matrix. $\mathbf{F}$ is a vector including 1 and six features of the optimized feature subset. By using the training data set, coefficients matrix $\mathbf{U}$ can be calculated

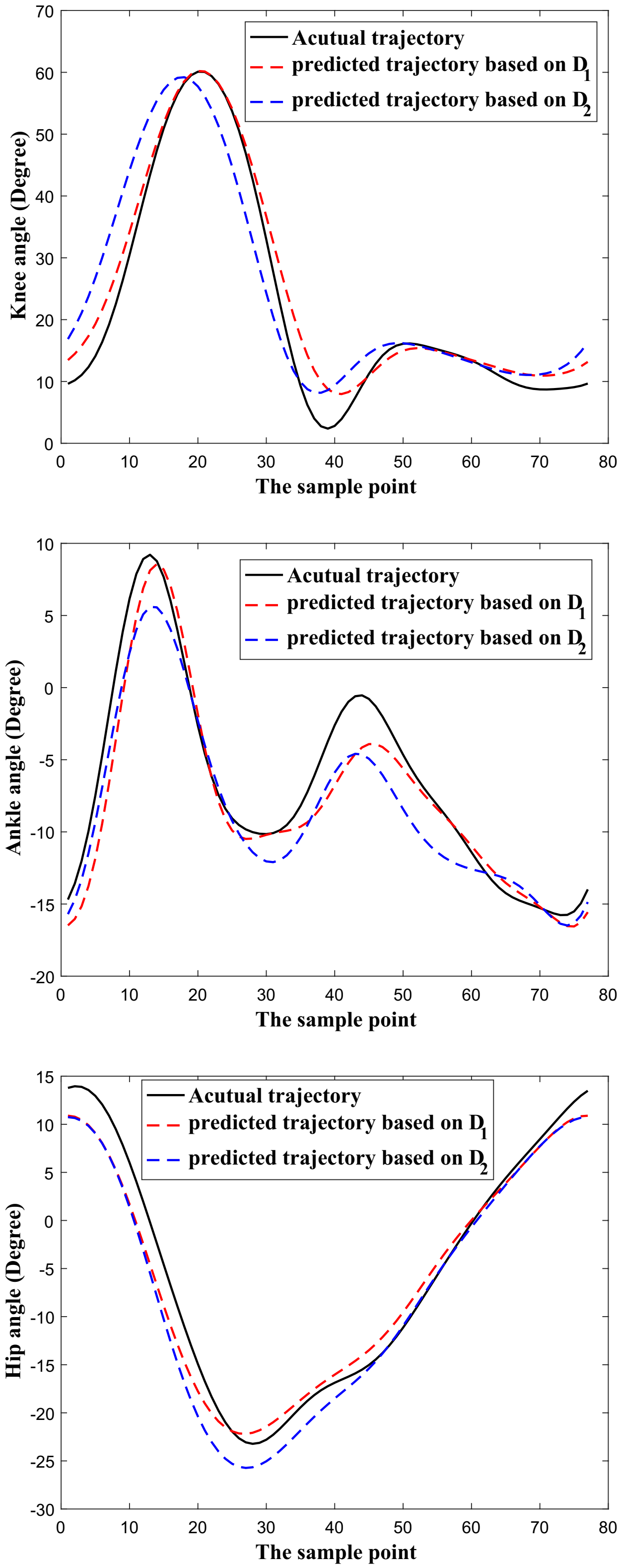

Fig. 8 The comparison results of actual joint trajectories, reconstructed joint trajectories based on different feature subsets

based on the least squares approach. When new features are given, Fourier coefficient vector can be calculated through coefficients matrix **U** to reconstruct the joint trajectory.

Further more, three regression algorithms, which include ε support vector regression (ε-SVR) (Chang and Lin 2011), back propagation neural network (BPNN) and eXtreme gradient boosting (Xgboost) (Chen and Guestrin 2016), are also added to comparison experiments in this study. For the fair comparison, the parameters ε and γ of ε-SVR are optimized by grid search method. For the optimized BPNN, the number of hidden neurons and training function are set to fifty and bayesian regularization, respectively. At the same time, the parameters *max_depth*, *learning_rate* and *n_estimatos* of Xgboost are also determined by grid search method, and are 0.3, 0.1 and 150, respectively.

The performance of RF models and other four models is evaluated by MAD and Pearson correlation coefficients using fivefold cross validation. The comparison results are shown in Table 4. It can be seen that RF models have better performance, i.e., lower MAD and higher correlation than other models.

To evaluate the consistency of the algorithm, the fivefold cross validation are used for five times. The results of every time are shown in Fig. 9. It can be seen that the variance of RF models is smaller than other models in terms of MAD and Pearson correlation coefficients. This indicates that good consistency can be obtained by using RF algorithm. In addition, for the ankle joint, the Pearson correlation coefficients between actual and reconstructed trajectories based on all models are small. This may due to that the angle variations of the ankle joint are relatively complex, and the relationship between anthropometric features and the hip joint is hard to be described.

6 Discussion and conclusion

In this study, a personalized gait generation model based on anthropometric features is proposed by using RF algorithm for the LLRR. Based on this model, Fourier coefficient vectors can be obtained and further to reconstruct the joint trajectories. At the same time, anthropometric features set are optimized through mRMR method and the number of trees in RF model is determined by the numerical searching method. From the experiment results, it can be found that the performance of proposed gait trajectory generation model based on the optimized feature subset by using RF algorithm is satisfactory.

It is worth noting that there are some limitations in this paper. One is that a lower limb rehabilitation robot lacks the degrees of freedom on the hip adduction. However, the gait trajectories dataset is collected from the normal subjects walked on the treadmill. The influences of lack of DoFs on

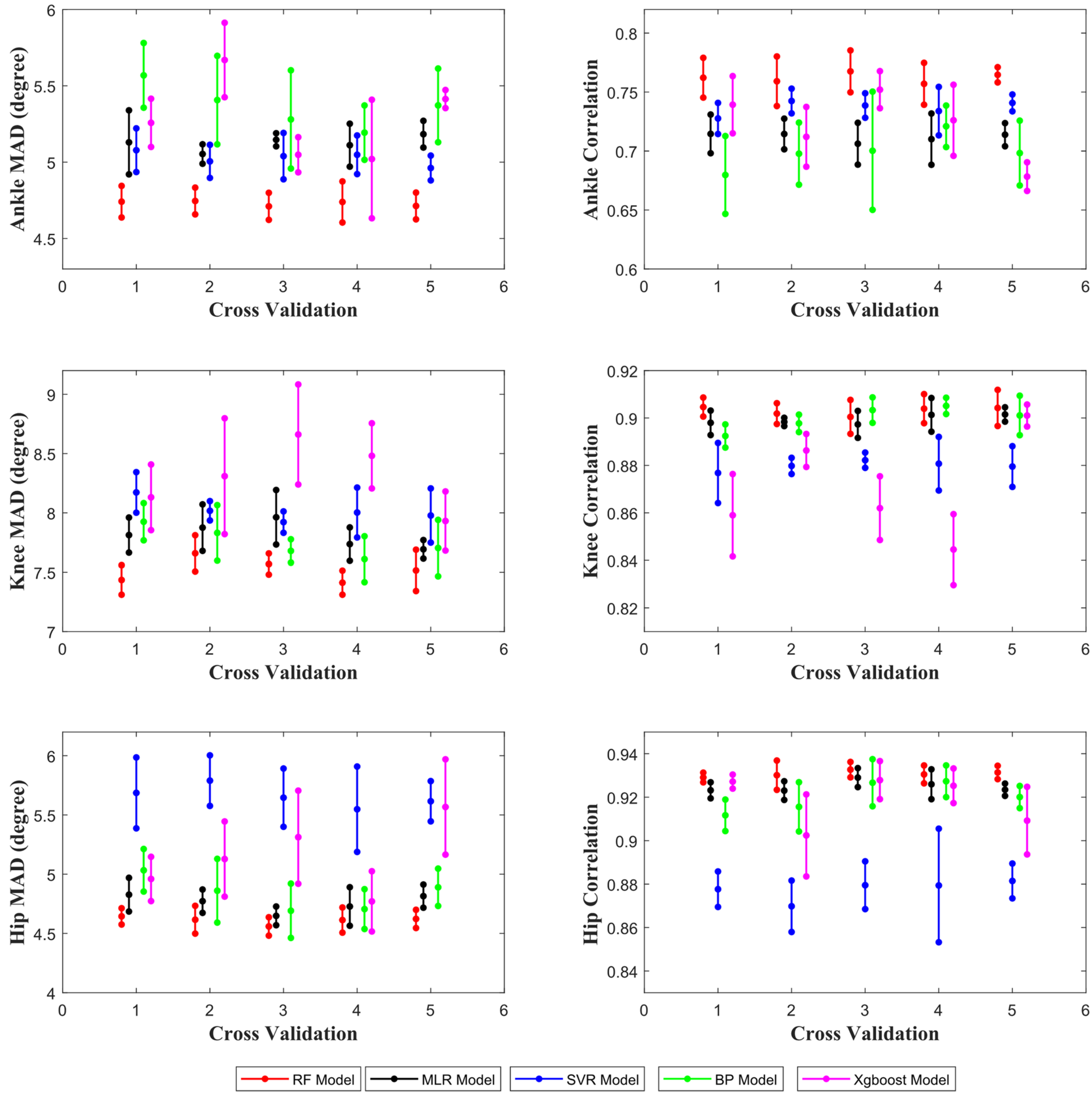

Fig. 9 MAD and Pearson correlation coefficients between actual trajectories and generated trajectories based on RF models and other four models for three joints. Each bar is obtained from a fivefold cross validation. The middle point in bar is the average, and two end points of the bar indicate the standard deviation

Table 4 The comparison experiment results

Joint	RF		MLR		SVR		BPNN		XGboost	
	e	*c*	*e*	*c*	*e*	*c*	*e*	*c*	*e*	*c*
Ankle	4.74	0.76	5.12	0.71	5.04	0.73	5.41	0.69	5.33	0.72
Knee	7.38	0.91	7.81	0.89	8.00	.88	7.73	0.90	8.79	0.87
Hip	4.60	0.93	4.76	0.92	5.63	0.88	4.86	0.92	5.14	0.91

the gait pattern need to be analysed. In the future work, more features which have the high relationship with the gait patterns will be considered into the prediction of the gait patterns. The personalized gait patterns will be added more gait characters, like the step length and step speed. In order to ensure that the predicted gait patterns can be easily applied to the lower limb rehabilitation robot, the gait pattern dataset collection will be done on the robot. Another important work is that the stroke patients will participate in the rehabilitation experiment to prove the efficiency of LLRR.

Acknowledgements This research was supported by the National Natural Science Foundation of China (Grants 91848110, 61720106012, 91648208) and the Strategic Priority Research Program of Chinese Academy of Science (Grant No. XDB32000000).

References

Amirabdollahian F, Loureiro R, Gradwell E, Collin C, Harwin W, Johnson G (2007) Multivariate analysis of the fugl-meyer outcome measures assessing the effectiveness of gentle/s robot-mediated stroke therapy. J Neuroeng Rehabil 4(1):4

Aoyagi D, Ichinose WE, Harkema SJ, Reinkensmeyer DJ, Bobrow JE (2007) A robot and control algorithm that can synchronously assist in naturalistic motion during body-weight-supported gait training following neurologic injury. IEEE Trans Neural Syst Rehabil Eng 15(3):387–400

Barbeau H, Wainberg M, Finch L (1987) Description and application of a system for locomotor rehabilitation. Med Biol Eng Comput 25(3):341–344

Bernhardt J, Dewey H, Thrift A, Collier J, Donnan G (2008) A very early rehabilitation trial for stroke (avert): phase ii safety and feasibility. Stroke 39(2):390–396

Breiman L (2001) Random forests. Mach Learn 45(1):5–32

Calabro RS, Cacciola A, Berte F, Manuli A, Leo A, Bramanti A, Naro A, Milardi D, Bramanti P (2016) Robotic gait rehabilitation and substitution devices in neurological disorders: where are we now? Neurol Sci 37(4):503–514

Chandler R, Clauser CE, McConville JT, Reynolds H, Young JW (1975) Investigation of inertial properties of the human body. Technical report, AIR FORCE AEROSPACE MEDICAL RESEARCH LAB WRIGHT-PATTERSON AFB OH

Chang CC, Lin CJ (2011) LIBSVM: a library for support vector machines. ACM Trans Intell Syst Technol 2:1–27

Chen G, Chan CK, Guo Z, Yu H (2013) A review of lower extremity assistive robotic exoskeletons in rehabilitation therapy. Crit Rev Biomed Eng 41(4–5):343

Chen T, Guestrin C (2016) XGBoost: A scalable tree boosting system. In: Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining, ACM, New York, NY, USA, KDD ’16, pp 785–794

Colombo G, Joerg M, Schreier R, Dietz V (2000) Treadmill training of paraplegic patients using a robotic orthosis. J Rehabil Res Dev 37(6):693–700

Ding C, Peng H (2005) Minimum redundancy feature selection from microarray gene expression data. J Bioinform Comput Biol 3(02):185–205

Feng Z, Qian J, Zhang Y, Shen L, Zhang Z, Wang Q (2008) Dynamic walking planning for gait rehabilitation robot. In: 2008 2nd international conference on bioinformatics and biomedical engineering. IEEE, pp 1280–1283

Hogan N, Krebs HI, Rohrer B, Palazzolo JJ, Dipietro L, Fasoli SE, Stein J, Hughes R, Frontera WR, Lynch D et al (2006) Motions or muscles? some behavioral factors underlying robotic assistance of motor recovery. J Rehabil Res Dev 43(5):605–618

Hu MH, Hsu SS, Yip PK, Jeng JS, Wang YH (2010) Early and intensive rehabilitation predicts good functional outcomes in patients admitted to the stroke intensive care unit. Disabil Rehabil 32(15):1251–1259

Kohavi R et al (1995) A study of cross-validation and bootstrap for accuracy estimation and model selection. Ijcai, Montreal, Canada, vol 14, pp 1137–1145

Koopman B, van Asseldonk EH, van der Kooij H (2014) Speed-dependent reference joint trajectory generation for robotic gait support. J Biomech 47(6):1447–1458

Kwakkel G, Kollen BJ, Krebs HI (2008) Effects of robot-assisted therapy on upper limb recovery after stroke: a systematic review. Neurorehabil Neural Repair 22(2):111–121

Liaw A, Wiener M et al (2002) Classification and regression by randomforest. R News 2(3):18–22

Liu DX, Du W, Wu X, Wang C, Qiao Y (2016) Deep rehabilitation gait learning for modeling knee joints of lower-limb exoskeleton. In: 2016 IEEE international conference on robotics and biomimetics (ROBIO). IEEE, pp 1058–1063

Luu TP, Lim HB, Qu X, Low K (2010) Subject tailored gait pattern planning for robotic gait rehabilitation. In: 2010 IEEE international conference on robotics and biomimetics (ROBIO). IEEE, pp 259–264

Luu TP, Lim H, Qu X, Hoon K, Low K (2011) Subject-specific lower limb waveforms planning via artificial neural network. In: 2011 IEEE international conference on rehabilitation robotics (ICORR). IEEE, pp 1–6

Malekipirbazari M, Aksakalli V (2015) Risk assessment in social lending via random forests. Expert Syst Appl 42(10):4621–4631

Meng W, Liu Q, Zhou Z, Ai Q, Sheng B, Xie SS (2015) Recent development of mechanisms and control strategies for robot-assisted lower limb rehabilitation. Mechatronics 31:132–145

Ming D, Jiang SL, Wang ZP, Qi HZ, Wan BK (2017) Review of walk assistant exoskeleton technology: human–machine interaction. Zidonghua Xuebao/Acta Automatica Sinica 43(7):1089–1100

Murray MP (1967) Gait as a total pattern of movement: Including a bibliography on gait. Am J Phys Med Rehabil 46(1):290–333

Niu X, Varoqui D, Kindig M, Mirbagheri MM (2014) Prediction of gait recovery in spinal cord injured individuals trained with robotic gait orthosis. J Neuroeng Rehabil 11(1):42

Peng H, Long F, Ding C (2005) Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. IEEE Trans Pattern Anal Mach Intell 27(8):1226–1238

Song WK (2016) Trends in rehabilitation robots and their translational research in National Rehabilitation Center, Korea. Biomed Eng Lett 6(1):1–9

Tolstov GP (2012) Fourier series. Courier Corporation, Chelmsford

Vallery H, Van Asseldonk EH, Buss M, Van Der Kooij H (2009) Reference trajectory generation for rehabilitation robots: complementary limb motion estimation. IEEE Trans Neural Syst Rehabil Eng 17(1):23–30

Wager S, Athey S (2017) Estimation and inference of heterogeneous treatment effects using random forests. J Am Stat Assoc **(just-accepted)**

Wang CYE, Bobrow JE, Reinkensmeyer DJ (2005) Dynamic motion planning for the design of robotic gait rehabilitation. J Biomech Eng 127(4):672–679

Winnen E, Beckwée D, Meeusen R, Baeyens JP, Kerckhofs E (2014) Does robot-assisted gait rehabilitation improve balance in stroke patients? A systematic review. Topics Stroke Rehabil 21(2):87–100

Wirz M, Bastiaenen C, Bie RD, Dietz V (2011) Effectiveness of automated locomotor training in patients with acute incomplete spinal cord injury: a randomized controlled multicenter trial. BMC Neurol 11(1):1–5

Youssef AM, Pourghasemi HR, Pourtaghi ZS, Al-Katheeri MM (2016) Landslide susceptibility mapping using random forest, boosted regression tree, classification and regression tree, and general linear models and comparison of their performance at wadi tayyah basin, asir region, saudi arabia. Landslides 13(5):839–856

Yun Y, Kim HC, Shin SY, Lee J, Deshpande AD, Kim C (2013) Statistical method for prediction of gait kinematics with Gaussian process regression. J Biomech 47(1):186