



Detecting Parkinson's Disease Using Gait Analysis with Particle Swarm Optimization

Xu Chen¹, Xiaohui Yao^{2(✉)}, Chen Tang¹, Yining Sun³, Xun Wang⁴,
and Xi Wu²

¹ Institute of Industrial and Equipment Technology,
Hefei University of Technology, Hefei, China

² School of Computer and Information,
Hefei University of Technology, Hefei, China

mia_xiaohui@163.com

³ Institute of Intelligent Machines, Chinese Academy of Sciences, Hefei, China

⁴ Hospital Affiliated to Institute of Neurology,
Anhui University of Chinese Medicine, Hefei, China

Abstract. Gait analysis is the study of human movements by analyzing temporal and spatial gait features. Research has shown that Parkinson's disease can degenerate human mobility, thereby causing afflicted individuals to behave differently in terms of gait characteristics. In this work, we propose an optimized method that assists us in better distinguishing people with Parkinson's disease from normal subjects. The spatial-temporal gait features are extracted by using a real U-shaped pressure-sensitive gait-sensing walkway. After pre-processing optimizations, including nondimensionalization and normalization of the raw features, we feed the features to an SVM classifier for training. The Particle Swarm Optimization algorithm is adopted to optimize the classification model. Experimental results show that the optimized method outperforms its predecessor by improving the accuracy from 87.12% to 95.66%, which shows the effectiveness of our proposed method in detecting Parkinson's Disease patients.

Keywords: Gait analysis · PSO · Support vector machine
Parkinson's disease

1 Introduction

Gait is the pattern of movement of the limbs of humans. Gait analysis is the systematic study of human movements by analyzing different spatial-temporal gait parameters, such as step length, stride speed, cadence, stance time and swing time [1]. As with biological characteristics such as fingerprints and irises, different individuals can exhibit different gait patterns with different gait parameters. By analyzing gait components of patients with Parkinson's disease (PD), Knutsson [2] found that some of their component movements of walking are abnormal compared with their normal counterparts (NP). Thus, gait analysis can be used as a promising tool in the assessment of human walking pattern and in distinguishing a particular group of people, such as PD patients, from others [3, 4]. It remains a challenge to effectively extract useful gait

parameters. This is a classical feature selection problem that has a wide range of applications [5, 6] and could also be regarded as a search process and various search strategies could be applied [7].

Early simple methods, such as visual observation and paper walkway, are applied. Although they are inexpensive and easy to implement, those methods usually fail to obtain accurate and reliable gait features [8, 9]. With the evolution of sensor technology, more researchers than ever before have started to take advantage of it. Vision-based and sensor-based approaches are two major categories of methods. The choice of method depends on the data that are collected.

Vision-based approaches collect clips of recorded videos during the subjects' walking processes. Then, gait parameters are extracted by analyzing and calculating the key frame images of the video, or by acquiring and calculating signal information that is generated from special sensors that are attached on key parts of the human body.

Beijer et al. [10] obtained the gait features of early PD patients using cheap handheld cameras. They fixed the camera above the ground at a distance of 0.5 m and all the subjects were walking away from the camera with their accustomed speeds. Compared with the professional GAITRite electronic walkway system, it is found that although the handheld video camera can obtain the average single step time and double support times with high accuracy at low cost, the stability of the acquired single step time is relatively low. Chien-Wen et al. [11] acquired image sequences of human silhouettes from a gait video that was captured by a SONY HDR-HC3 camcorder and extracted the intrinsic features by linear discriminant analysis.

With the continuous development of microelectromechanical system technology, inertial sensors, such as 3D accelerometers and gyroscopes, are widely applied. Researchers fix the sensors either on the key parts of the subject's body or in the insole [5, 12, 13]. Stacy et al. [12] designed a gait shoe system called GaitShoe, which has three orthogonal accelerometers, three orthogonal gyroscopes, four force sensors, two bidirectional bend sensors and electric field height sensors. A comprehensive analysis of the data that are obtained by these sensors can accurately detect heel-strike and toe-off motions and estimate foot orientation and position. Mitsuru et al. [14] built their accelerometry-based gait analysis system with the help of one acceleration sensor that was mounted on the human trunk. They attached the sensor to the trunk of each subject and collected the acceleration signals of 11 healthy people and 12 PD patients. By utilizing the cross-correlation and anisotropy properties of the signal, they could detect the gait peaks due to stride events more accurately. Tiffany et al. [15] devised an inertial sensor system by fixing sensors on the lower extremities. They acquired the acceleration information of subjects' lower limbs and converted them into computer-generated animations. Using the generated animation to evaluate the gait freeze of Parkinson patients, a more robust result was obtained. This approach provides a new possibility for the frozen diagnosis of Parkinson's gait outside the clinic. There is no doubt that these inertial sensors can collect more accurate gait data. However, fixing the sensors on some parts of the human body, especially the legs, will make people feel uncomfortable and even affect their daily lives [14]. Moreover, limited battery life makes it impossible to collect continuous data for a long period of time.

Although many different methods for gait data extraction are available, it is still necessary to find better ways of extracting those data accurately without invasion.

In this paper, we choose a U-shaped electronic gait-sensing walkway that is based on flexible array pressure sensors for collecting our gait spatial-temporal features. The flexible array pressure sensor has the advantages of high sensitivity to low pressure, fast response time and good stability, which guarantee the accuracy of the collected data compared to the previous systems. Since the sensor is flexible, it can be applied to various laboratory scenarios, including irregular surfaces, thereby making it immune to extreme conditions such as illumination. In addition, the subjects don't have to wear any sensor devices; they only have to walk normally, thereby making it possible to reduce the potential influence of stress of the subjects on the collected data. Moreover, the walkway system can obtain gait data at a turning point since it is U-shaped.

Previous work has been done on collecting gait parameters of PD patients and age-matched controls and an SVM classification model has been constructed for distinguishing PD patients with an accuracy of 87.12% [16]. However, in practice, misdiagnosing a PD patient as healthy or a healthy person as having PD will have a great impact on the person. Thus, the classification model still has much room for optimization. In this paper, we propose an optimized Parkinson's disease detection method for improving the classification performance. Spatial-temporal gait features are extracted from the U-shaped electronic gait-sensing walkway system. Compared with the previous model, we eliminate the influence of height on the gait features and apply max-min normalization. Then, the Particle Swarm Optimization algorithm is applied to optimize the parameters of the SVM classifier. Finally, we use a ten-fold cross-validation technique to evaluate the performance of the model. Experimental results show that the performance of the optimized model is greatly improved.

The structure of this paper is as follows: Sect. 2 details the data processing techniques and PSO and SVM algorithms for building our classification model. Experiments on data collection, feature extraction, and classification model construction are described in Sect. 3. In Sect. 4, we conclude this work and discuss our future work.

2 Method

A flow chart of our method is shown in Fig. 1. It has three parts: data acquisition, data preprocessing and classification model construction.

2.1 Nondimensionalization

By experience, compared to taller people, shorter people tend to walk with shorter step length but higher cadence. Therefore, people of shorter stature may have closer gait parameters to people with Parkinson's disease than to taller normal subjects. Thus, it is of great importance to define a "normal step length" in gait analysis. AL Hof [17] proposed a definition of "normal step length". He divides step length by the corresponding human height, as shown in formula 1. By this treatment, the step lengths of normal people of different heights will cluster around a fixed number. Similarly, for other gait features, there are corresponding dimensionless formulas. For all length-relevant features, such as stride length, formula (1) is used to eliminate the influence of height. For speed-relevant features, such as gait velocity, formula (2) is

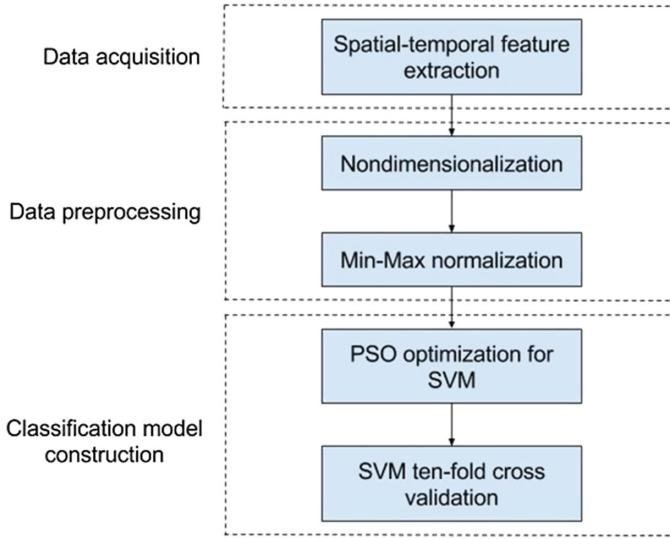


Fig. 1. Flow chart of our method

employed to remove the influence of height. All frequency-related features, including cadence, are processed by formula (3). All time-related features, including stance time, swing time, pre-swing time, gait cycle, double support time, and turning time, are processed by formula (4).

$$\hat{l} = \frac{l}{l_0} \tag{1}$$

$$\hat{v} = \frac{v}{\sqrt{gl_0}} \tag{2}$$

$$\hat{f} = \frac{f}{\sqrt{g/l_0}} \tag{3}$$

$$\hat{t} = \frac{t}{\sqrt{l_0/g}} \tag{4}$$

In the above formulas (1–4), l represents a length-related feature, v represents a speed-related feature, t represents a time-related feature, f represents a frequency-related feature, l_0 is the height of the subject, and g represents the gravitational acceleration constant. In this paper, we choose g to be 9.81.

2.2 Normalization

The data need to be normalized before the classification experiments are carried out. Normalization is the scaling of data in proportion to a small specific interval. The main

benefit of normalization is apparent when the features vary a lot in scale. Because of the different nature of each parameter, some parameters tend to be very large while others tend to be very small. When the sizes of the parameter are substantially different, if we directly use the original parameter value for analysis, we will highlight the roles of larger parameters in the comprehensive analysis and weaken the roles of smaller parameters. Therefore, to ensure the reliability of the results, it is necessary to normalize the data of the original parameters.

As the values of the gait features that are extracted from the experiment are positive, this experiment adopts the min-max normalization method. A linear transformation is applied to the original data to map the results to [0,1]. The specific formula is as follows: For sequence $x_1, x_2, x_3, \dots, x_n$, we obtain new sequence $y_1, y_2, y_3, \dots, y_n \in [0, 1]$ by

$$y_i = \frac{x_i - \min\{x_1, x_2, x_3, \dots, x_n\}}{\max\{x_1, x_2, x_3, \dots, x_n\} - \min\{x_1, x_2, x_3, \dots, x_n\}} \quad (5)$$

2.3 Classification Model

Support Vector Machine. The support vector machine (SVM) algorithm was invented by Vladimir N. Vapnik and Corinna Cortes [18] in 1993 and published in 1995. It is a powerful binary classification model for high-dimensional data with small data sets. For problems with linearly separable training samples, a hyperplane model is used to distinguish the samples. The parameters of the model are derived by maximizing the margin, which is the distance of the support vector to the hyperplane. For non-linear classification problems, a kernel function is adopted to transform the linearly inseparable samples from low-dimensional feature space into higher-dimensional feature space. After the mapping procedure, those samples become linearly separable. The SVM algorithm constructs the optimal hyperplane in feature space based on the concept of structural risk minimization, which globally optimizes the model. The SVM model is expressed in formula (6). This paper chooses a radial basis function (RBF) as the mapping function, which is expressed in formula (7). The code is from LIBSVM, which was developed by Professor Lin Chih-Jen from Taiwan University [19].

$$\min \frac{1}{2} \|w\|^2 + C \sum_{i=1}^n \xi_i \quad s.t. \quad y_i [(w^T x_i) + b] \geq 1 - \xi_i \quad i = 1, 2, \dots, n, \xi_i \geq 0 \quad (6)$$

$$K(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2) \quad (7)$$

Particle Swarm Optimization. Particle Swarm Optimization (PSO) is an evolutionary computational technique, which was proposed by Drs. Eberhart and Kennedy in 1995. The algorithm is inspired by the foraging behaviors of birds. The basic strategy of PSO is to find the optimal solution through collaboration and information sharing among

particles in the group. PSO has a high convergence rate and has been widely applied to algorithm parameter optimization [20], neural network training [21], fuzzy system control and other genetic algorithms. The basic principle is shown in formula (8) and formula (9).

$$v[i] = w*v[i] + c_1 * rand() * (pbest[i] - present[i]) + c_2 * rand() * (gbest - present[i]) \tag{8}$$

$$present[i] = present[i] + v[i] \tag{9}$$

Each particle ($present[i]$) represents a candidate solution to the optimization problem. It evolves from its own “memory term” ($w * v[i]$), “self-recognition term” ($c_1 * rand() * (pbest[i] - present[i])$) and “group cognitive term” ($c_2 * rand() * (gbest - present[i])$), as shown in formula (8). A flow chart of the PSO algorithm is shown in Fig. 2.

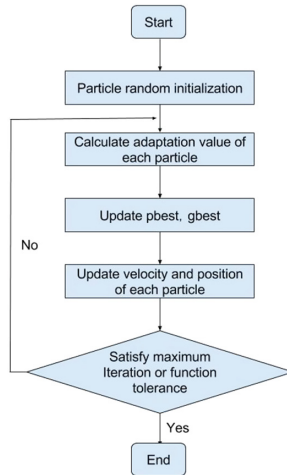


Fig. 2. Flow chart of the PSO algorithm

2.4 Performance Evaluation

To comprehensively evaluate the performance of the classification model, we need to compare the actual label and the predicted value. Table 1 shows the confusion matrix, which indicates the comparison results between the actual labels and predicted values. TP (True Positive) represents the number of real patients with Parkinson’s disease who are predicted to be PD patients; FP (False Positive) represents the number of people without PD who are predicted to be PD patients; FN (False Negative) represents the number of PD patients who are predicted to not have PD; TN (True Negative) represents the number of people without PD who are predicted to not have PD.

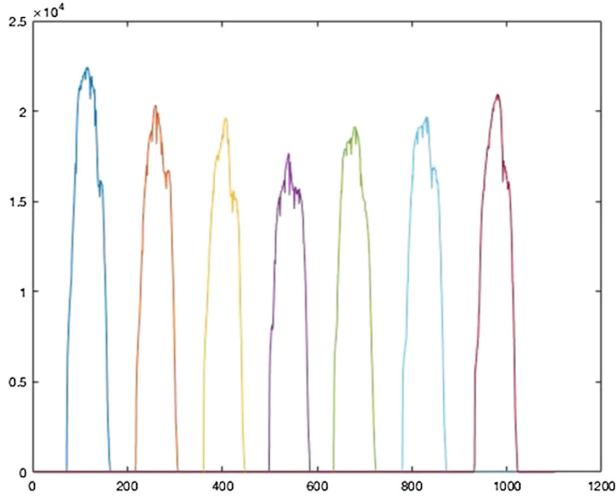


Fig. 3. Pressure signal of one foot

Table 1. Confusion matrix

Prediction	Actual	
	PD	NP
PD	TP	FP
NP	FN	TN

We use Accuracy, Precision, Recall and F-measure to evaluate the performance, which are expressed in formulas (10–13).

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} \times 100\% \quad (10)$$

$$Precision = \frac{TP}{TP + FP} \times 100\% \quad (11)$$

$$Recall = \frac{TP}{TP + FN} \times 100\% \quad (12)$$

$$Fmeasure = \frac{2 \times Precision \times Recall}{Precision + Recall} \times 100\% \quad (13)$$

The accuracy rate is the proportion of all correct predictions in the overall population. The higher the accuracy rate is, the better the model performs. Precision is the ratio of true-positive predictions to all positive predictions. The recall rate is the ratio of true-positive predictions to all actual positives. F-measure represents the weighted harmonic mean of recall and precision. It combines the calculation results of recall and

precision, and a high value of F-measure value indicates that the method of classification is effective.

3 Experiments

3.1 Data Acquisition

The U-shaped walkway system [22] was designed by the Hefei Intelligent Machinery Research Institute, Chinese Academy of Sciences. It is composed of a hardware system and a software system. The hardware system has a data acquisition module and a data transmission module. Fourteen flexible pressure-sensitive plates, 5 three-dimensional force-measuring plates, and 1 balance tester are arranged in a U shape. Figures 4 and 5 show a diagram of the distribution of the plates and a real photo with people walking on them respectively. The parameters of the flexible pressure-sensitive plates are as follows: the data acquisition frequency is 100 Hz, the size is 80 cm * 80 cm, and the pressure point density is 4/cm². The parameters of the three-dimensional force-measuring plates are as follows: the data acquisition frequency is 500 Hz and the size is 80 cm * 80 cm. The balance-tester plate has the same size as the other plates. When a person walks on the plates, the system senses the pressure, which is triggered by the change in the resistance values of the resistors that are built into the pressure-sensitive plates.

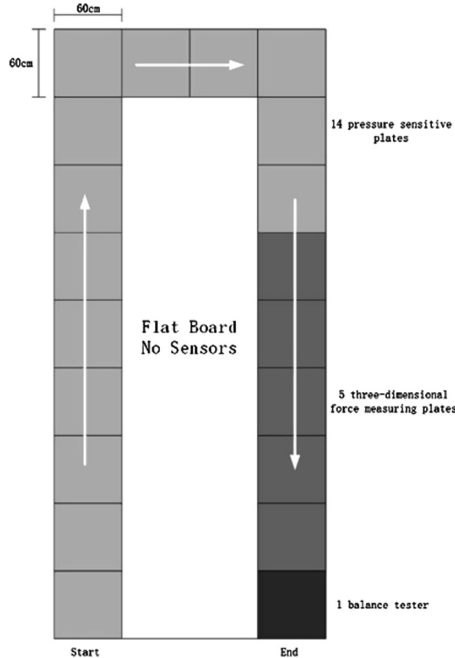


Fig. 4. U-shaped walkway illustration

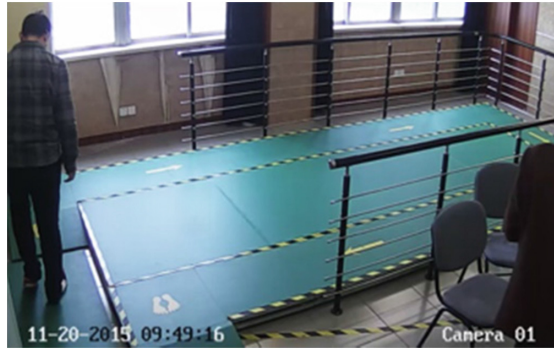


Fig. 5. Real photo of U-shaped walkway with people walking on it

All the flexible pressure-sensitive plates are connected with one another via a power cable and a network cable. The power cable is connected to the regulated power supply and the network cable is connected to the multi-port router. The data that are collected by the sensor are transmitted to the host computer via the router and processed by the software system. The data transmission is based on TCP/IP client server mode. The original data format is shown in Table 2. Together, the plate number, row number, and column number determine the unique coordinates of each pressure point on the walkway. In this way, we can obtain the pressure values of all the pressure points of the walkway and the corresponding time information. Figure 3 shows the pressure signal of one foot during a walking procedure.

Table 2. Original data format

TimeID	PlateId	RowNo	ColNo	Pressure
--------	---------	-------	-------	----------

3.2 Feature Extraction

The raw pressure signals that are collected by the hardware system are transmitted to the PC host computer. These raw pressure signals need to be processed by the software system to generate the final useful gait features. We recruit 42 patients with Parkinson’s disease and 93 age-matched normal controls. All of them come from Hospital Affiliated to Institute of Neurology, Anhui University of Chinese Medicine. All subjects took no medications and gave informed consent within 24 h. All of them are asked to walk at their accustomed pace under the supervision of two doctors (Table 3).

The first step is to obtain footprints (shown in Fig. 6) of each subject by means of signal analysis and an image processing method. Then, we use those footprints to calculate the spatial and temporal features. The spatial and temporal features of all gait parameters come from the UPDRS-III scale, which was proposed by the Rancho Los Amigos (RLA) Medical Center in California, USA. These parameters are as follows: step length, stride length, gait velocity, cadence, stance time, swing time, pre-swing time, gait cycle, and double support time. To consider the gait abnormalities that Parkinson’s disease patients may have when they are turning, we added two additional

Table 3. Twenty-two gait features

Physical features	Spatial-temporal gait features	Turning time features
Age	Step length (cm)	Turning time 1 (ms)
Gender	Stride length (cm)	Turning time 2 (ms)
Weight (kg)	Gait velocity (m/s)	
Height (cm)	Cadence (steps/s)	
	Stance time (ms)	
	Swing time (ms)	
	Pre-swing time (ms)	
	Gait cycle (ms)	
	Double support time (ms)	

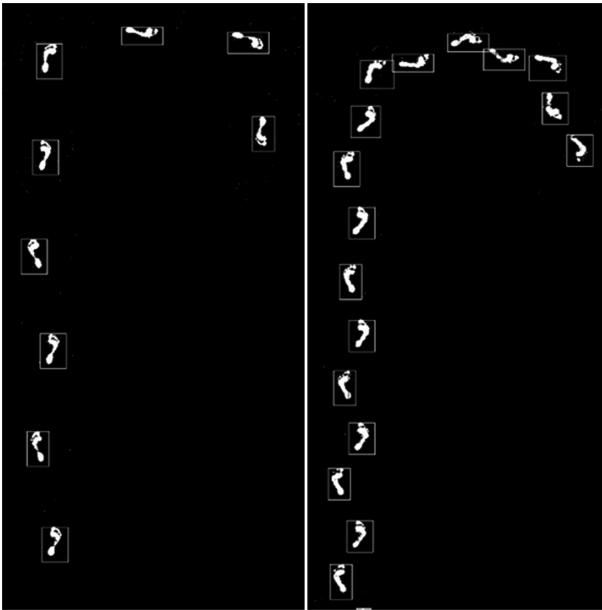


Fig. 6. Footprints of one normal subject (left) and one PD patient (right)

time features of turning at the two tuning points of the U-shaped walkway. Finally, 22 features were extracted. To increase the number of samples, some of the subjects were required to walk multiple times. The final number of samples is 242 (159 PD samples and 93 normal control samples).

3.3 Experimental Results

The experiment is divided into two parts. In the first part, we use the PSO algorithm to find the optimal combination of SVM parameters. The default parameters of PSO are set as follows: $c_1 = 1.5$, $c_2 = 1.7$, number of particles $N = 20$, and the number of

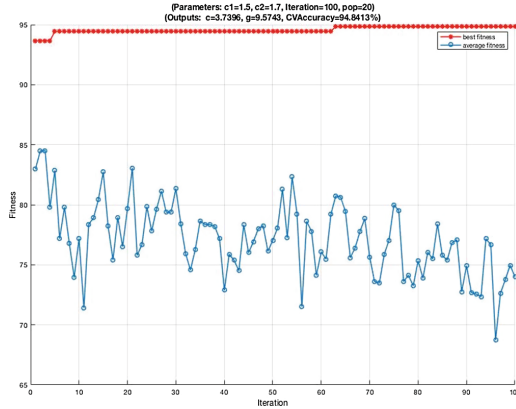


Fig. 7. Fitness line

iterations is set to 100. We use the three-fold cross-validation accuracy of the SVM classifier as the evaluation criterion. Figure 7 shows the fitness line that was obtained. According to the figure, after 100 iterations, the best fitness of the PSO algorithm is 94.8413%. The corresponding parameter values, namely, $c = 3.7396$ and $g = 9.5743$, are chosen for the SVM model.

In the second experiment, we use the parameters that were acquired by PSO to train the SVM model using ten-fold cross-validation. The examples are randomly divided into 10 subsets of equal size and we train the SVM classification model 10 times. Each time, we choose one subset of the examples for testing and use the other nine subsets for training, and obtain one set of results, which consists of a performance evaluation matrix, accuracy, precision, recall and F-measure. Finally, we average the ten sets of results and treat the average results as the final performance evaluation results, which are shown in Table 4. Comparing to its predecessor model without PSO optimization, the new model’s performance is significantly improved: accuracy, precision, and F-measure are increased by 8.54%, 7.39 and 2.04%, respectively, while recall remains approximately the same at 63.53%.

Table 4. Results

Model	Accuracy (%)	Precision (%)	Recall (%)	F-measure (%)
SVM	87.12	89.63	64.29	74.74
PSO-SVM	95.66	97.02	63.53	76.78

4 Conclusions

In this paper, we extract effective gait features from a U-shaped gait-sensing platform that is based on flexible pressure-sensitive sensors. Then, we nondimensionalize the raw features to eliminate the influence of height on the gait parameters and carry out

data preprocessing using max-min normalization. Finally, we construct an SVM classification model, whose parameters are optimized by the PSO algorithm. The accuracy of the model is as high as 95.66%, which indicates that the optimization algorithm can improve the classification performance effectively. Our future work will include determining the prevalence of Parkinson's disease using a multi-class classification algorithm and collecting more features to improve the performance of the classifier.

Acknowledgements. This work was supported in part by the National Natural Science Foundation of China under Grant No. 71661167004, the Anhui Key Project of Research and Development Plan under Grant No. 1704e1002221, the Foshan Science and Technology Innovation Project under Grant No. 2015IT100095, and the Program of Introducing Talents of Discipline to Universities ("111 Program") under Grant No. B14025.

References

1. Whittle, M.W.: Gait analysis: an introduction - 3rd edition. *Physiotherapy* **77**(11), 786 (2003)
2. Knutsson, E.: An analysis of parkinsonian gait. *Brain J. Neurol.* **95**(3), 475–486 (1972)
3. Sejdic, E., Lowry, K.A., Bellanca, J., Redfern, M.S., Brach, J.S.: A comprehensive assessment of gait accelerometry signals in time, frequency and time-frequency domains. *IEEE Trans. Neural Syst. Rehabil. Eng.* **22**(3), 603–612 (2014)
4. Bae, J., Kong, K., Byl, N., Tomizuka, M.: A mobile gait monitoring system for gait analysis. In: 11th International Conference on Rehabilitation Robotics, pp. 73–79. IEEE, Kyoto (2009)
5. Wang, A., Chen, G., Yang, J., Zhao, S., Chang, C.Y.: A comparative study on human activity recognition using inertial sensors in a smartphone. *IEEE Sens. J.* **16**(11), 4566–4578 (2016)
6. An, N., Que, X., Ding, H., Yuan, J., Tang, X., Farrer, L.A., Au, R.: Application of deep learning methods in identifying proteomic risk markers for Alzheimer's disease. *IEEE/ACM Trans. Comput. Biol. Bioinform. (TCCB)* (2018) (To appear)
7. Yang, J., Xu, Y., Sun, G., Shang, Y.: A new progressive algorithm for a multiple longest common subsequences problem and its efficient parallelization. *IEEE Trans. Parallel Distrib. Syst.* **24**(5), 862–870 (2013)
8. Eastlack, M.E., Arvidson, J., Snyder-Mackler, L., Danoff, J.V.: Interrater reliability of videotaped observational gait-analysis assessments. *Phys. Ther.* **71**(6), 465–472 (1991)
9. Clarkson, B.H.: Absorbent paper method for recording foot placement during gait. Suggestion from the field. *Phys. Ther.* **63**(3), 345 (1983)
10. Beijer, T.R., Lord, S.R., Brodie, M.A.: Comparison of handheld video camera and GAITRite® measurement of gait impairment in people with early stage Parkinson's disease: a pilot study. *J. Parkinsons Dis.* **3**(2), 199–203 (2013)
11. Cho, C.W., Chao, W.H., Lin, S.H., Chen, Y.Y.: A vision-based analysis system for gait recognition in patients with Parkinson's disease. *Expert Syst. Appl.* **36**(3), 7033–7039 (2009)
12. Bamberg, S., Benbasat, A.Y., Scarborough, D.M., Krebs, D.E.: Gait analysis using a shoe-integrated wireless sensor system. *IEEE Trans. Inf. Technol. Biomed.* **12**(4), 413–423 (2008)

13. Tien, I., Glaser, S.D., Aminoff, M.J.: Characterization of gait abnormalities in Parkinson's disease using a wireless inertial sensor system. In: 32rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society, pp. 3353–3356. IEEE, Buenos Aires (2010)
14. Yoneyama, M., Kurihara, Y., Watanabe, K., Mitoma, H.: Accelerometry-based gait analysis and its application to Parkinson's disease assessment— part 1: detection of stride event. *IEEE Trans. Neural Syst. Rehabil. Eng.* **22**(3), 613 (2014)
15. Morris, T.R., Cho, C., Dilda, V., Naismith, S.L.: Clinical assessment of freezing of gait in Parkinson's disease from computer-generated animation. *Gait & Posture* **38**(2), 326 (2013)
16. Zhang, Y., Wu, X., Xu, S., Yang, X., Wang, X.: U-shape electronic walkway system: dyskinesia assessment in Parkinson's disease. *Comput. Eng. Appl.* **54**(1), 166–171 (2018)
17. Hof, A.L.: Scaling gait data to body size. *Gait & Posture* **4**(3), 222–223 (1996)
18. Chang, C.C., Lin, C.J.: LIBSVM: a library for support vector machines. *ACM Trans. Intell. Syst. Technol.* **2**(3), 1–27 (2011)
19. Cherkassky, V.: The nature of statistical learning theory. *IEEE Trans. Neural Netw.* **38**(4), 409 (2002)
20. Ranaee, V., Ebrahimzadeh, A., Ghaderi, R.: Application of the PSO-SVM model for recognition of control chart patterns. *ISA Trans.* **49**(4), 577–586 (2010)
21. Ren, C., An, N., Wang, J., Li, L., Hu, B., Shang, D.: Optimal parameters selection for BP neural network based on particle swarm optimization: a case study of wind speed forecasting. *Knowl.-Based Syst.* **56**(3), 226–239 (2014)
22. Wu, X., Chen, X., Duan, Y., Xu, S., Cheng, N., An, N.: A study on gait-based Parkinson's disease detection using a force sensitive platform. In: 8th IEEE International Conference on Bioinformatics and Biomedicine (BIBM), pp. 2330–2332. IEEE, Kansas City (2017)