Inferring Cognitive Wellness from Motor Patterns

Yiqiang Chen, Chunyu Hu, Bin Hu, Lisha Hu, Han Yu and Chunyan Miao

Abstract—Changes in the motor pattern have been shown to be useful advanced indicators of cognitive disorders, such as Parkinson's disease (PD) and cerebral small vessel disease (SVD). It would be highly advantageous to tap into data containing people's motor patterns from motion sensing devices to analyze subtle changes in cognitive abilities, thereby providing personalized interventions before the actual onset of such conditions. However, this goal is very challenging due to two main technical problems: 1) the size of data labeled by doctors is small, and 2) the available data tends to be highly imbalanced (the vast majority tend to be from normal subjects with only a small fraction from subjects with cognitive disorder). In order to effectively deal with these challenges to infer cognitive wellness from motor patterns with high accuracy, we propose the MOtor-Cognitive Analytics (MOCA) framework. The proposed MOCA first uses the random oversampling iterative random forest based feature selection method to reduce the feature space dimensionality and avoid overfitting, and then adds a bias in the optimization problem of weighted extreme learning machine to achieve good generalization ability in handling imbalanced small-sampling dataset. Experimental results on two real-world datasets including SVD and stroke patients show that MOCA can effectively reduce the rate of misdiagnosis and significantly outperform state-of-the-art methods in inferring people's cognitive capabilities. This work opens up opportunities for population-level pre-screening using motion sensing devices and can inform current discussions on reforming the health-care infrastructure.

Index Terms—Correlation analysis, motor pattern, cognitive wellness, imbalanced small-sampling feature selection, imbalanced classification.

1 Introduction

Neurodegenerative disease is one of the most serious types of disease among the elderly population, which threatens the lives of millions of people and causes many social problems. The traditional view in medicine divides neurodegeneration into two categories: 1) cognitive disorders and 2) dyskinesia. They are usually treated with different approaches. In recent years, it has been shown that cognitive disorders are often accompanied by movement disorders while dyskinesia is associated with cognitive degeneration symptoms. In [1], executive functioning was shown to be an important factor for explaining the motor-cognitive link and the predictive power of fine motor skills for early academic achievement. In [2], cognition was found to be closely associated with gait in complex ways. Irene M.J. van der Fels et al. [3] reviewed current literature and concluded that weak-to-strong relations were found between motor functions and cognitive skills, and complex motor intervention programs can be used to stimulate both motor and higher order cognitive skills in prepubertal children. Inspired by these findings, many researchers analyzed the correlation between motor patterns and cognitive abilities with motion sensing technologies [4], [5], [6], sometimes with the help of crowdsourcing [7], [8]. Their findings laid the foundations for personalized monitoring. In [9], apraxia was suggested to be useful in differentiating Alzheimer's disease (AD) from frontotemporal dementia. Previous studies also showed the importance of measuring an elderly person's gait and balance for detecting cognitive diseases such as cerebral small vessel disease (SVD) [10] and white matter changes [11].

Despite its obvious allure, leveraging technology that people already have in their pockets for detecting medically important events is not as simple as it appears. In general, it is very expensive and difficult to label a lot of data for the training of cognitive disorder detection model. Magnetic Resonance Imaging (MRI), which is essential for the diagnosis of cognitive wellness, is very costly. Besides, it is hard to find a large number of patients with cognitive disorder to participate in the data collection. Thus, data collected for cognitive disorder detection often suffers from the negative effects of small-sampling [12], [13], [14] and data imbalance [15]. With an imbalanced small-sampling dataset, it is difficult to identify the most discriminant motor pattern features for accurate diagnosis [12], [13], [14]. Imbalanced datasets [15] will lead to a natural tendency that the majority class (normal people) is favored in comparison with the minority class (patients with cognitive disorder).

However, many existing works do not address these two problems in motor cognitive analysis. In [1], the AMOS 19 software [16] was used to construct structural equation models to analyze the correlations between motor pattern and cognitive capability. The influence of imbalanced small-sampling dataset was overlooked in this study. In [2], multivariate linear regression and Spearman's correlation techniques were adopted to investigate the associations between cognition and gait. In this research, the imbalanced data distribution was not taken into consideration in their statistical analysis. In [17], step-wise multiple-regression was applied to analyze the relationship between mental rotation and motor processes. Although there were only 65 participants, they didn't pay attention to the small-sampling issue in their correlation analysis. All the above works did not address the small-sampling and imbalanced issues.

- *Yiqiang Chen, Chunyu Hu and Lisha Hu are with the Institute of Computing Technology, Chinese Academy of Sciences, Beijing, China, with University of Chinese Academy of Sciences, Beijing, China, and also with Beijing Key Laboratory of Mobile Computing and Pervasive Device, Beijing, China (Corresponding author: Yiqiang Chen, e-mail: yqchen@ict.ac.cn).*
- *Bin Hu is with the School of Information Science and Engineering, Lanzhou University, China.*
- *Han Yu and Chunyan Miao are with Joint NTU-UBC Research Centre of Excellence in Active Living for the Elderly (LILY), Nanyang Technological University, Singapore.*

In our previous work [18], we proposed a coarse-to-fine feature selection method to discover the most significant features from a high dimensional feature set from small-sampling data. In this paper, we propose a novel MOtor-Cognitive Analytics (MOCA) framework, which aims at inferring cognitive wellness from motor patterns. MOCA extends the IRFFS feature selection method from our previous work [18] to IRFFS-O to handle the oversampling issue as a result of imbalanced small-sampling datasets in this work. In addition, we propose a novel classification method, b-WELM, which adds a bias in the optimization process of weighted extreme learning machine in order to achieve good generalization ability in handling imbalanced small-sampling datasets.

Experimental results based on real-world datasets demonstrate that the proposed MOCA is effective in feature selection and performs better than other state-of-the-art feature selection methods in [19]. Furthermore, experimental results obtained from MOCA also demonstrate its effectiveness in accurate classification of imbalanced data. The proposed framework opens up opportunities for population-level pre-screening using smart mobile devices and can inform current policy discussions on reforming the health-care infrastructure.

The rest of the paper is organized as follows. Section 2 introduces existing works most related to the proposed framework. Section 3 describes the proposed correlation analysis framework. In Section 4, we conduct experiments on two datasets to demonstrate the effectiveness of the proposed framework. Conclusions and future work are given in Section 5.

2 Related Work

2.1 From Motor Patterns to Cognitive Wellness

In recent years, the rapid advancement of computer science makes it possible for machine learning to be used for inferring cognitive wellness based on motor patterns. For example, Termenon *et al.* [15] built a brain MRI morphological pattern extraction tool based on Extreme Learning Machine (ELM) [20] and majority voting classification. Chen proposed an approach for automatic stroke detection with high accuracy through a Trail Making Test [21]. In [22], Support Vector Machine (SVM) was trained and tested on Fractional Anisotropy and Mean Diffusivity data to detect patients with AD. Savio *et al.* [23] proposed a feature extraction method based on artificial neural network and SVM to detect two neurological disorders with cognitive impairment: Myotonic Dystrophy of Type 1 and AD.

Changes in the motor pattern have been shown to be associated with cognitive impairment in PD [24], AD [9] and SVD [18]. Some researches have been conducted to study the correlation between motor pattern and cognitive impairment with different data collection and correlation analysis methods. Delaram *et al.* [25] used the e-AR sensor to collect acceleration data for PD patients. An iterative algorithm was proposed for gait analysis in their work. Ana Lisa *et al.* [26] adopted the quantitative digitography (QDG), which used a MIDI interfaced keyboard to collect the data in repetitive alternating finger-tapping task, to analyze the relationship between QDG and UPDRS III scores. In [27], Kita *et al.* proposed two smart sensing systems to detect the specific motion symptoms of PD. One of their systems consisted of a single sensor integrated in the headphone, and the other one was composed of two sensors on the shins. Accelerometer, gyroscope, and magnetometer were integrated into each sensor unit. In [28], low-cost motion detectors and labeled images were used to monitor the abnormal behavior of AD patients and alarm triggering. To improve the efficiency of data annotation, Yang *et al.* [29] proposed a novel video annotation method, which exploited Web images to help learn robust semantic video indexing classifier. In recent years, game-based data collection systems have also emerged which investigated people's cognitive capabilities in the forms of digital games and unobtrusively collected fine-grained complex inter-temporal decisions-making behaviors [30], [31], [32].

2.2 The Challenges of Imbalanced Small-sampling Dataset

In disease detection, a common challenge is how to select the most significant features from an imbalanced small-sampling dataset for accurate detection. Furthermore, the imbalance characteristic often leads to false negative phenomenon in which the majority class (i.e. normal people) are favored in comparison to the minority class (i.e. patients with cognitive disorder). Feature expression is an important part in correlation analysis. A good feature expression is critical to the performance of the correlation analysis method. Many researchers have paid attention to this issue. In [33], Jun *et al.* proposed a novel stacked deep polynomial network algorithm for the texture feature representation learning on small ultrasound datasets. In [34], discriminating features for the detection of PD were selected using mutual information based approach. In [35], the features selected by genetic programming were significant in the survival prediction of a small size of oral cancer prognosis. In [36], Zhu *et al.* proposed a novel self-taught dimensionality reduction approach, which was the first work that employed external information for dimensionality reduction on the high dimensional and small-sized data. However, additional external information was required to assist the dimensionality reduction on high-dimensional and small-sized target data. In [37], Genuer *et al.* proposed a feature selection strategy based on random forests. However, feature selection using random forests is not suitable for imbalanced small-sampling dataset. This is mainly because even a little perturbation would cause a large difference in out-of-bag error for small-sampling data. On the other hand, information gain or Gini index based algorithms, such as decision tree and random forests, always show poor performance on imbalanced datasets, which can be attributed to the skew sensitive splitting criteria [38].

To solve the problems caused by imbalanced data distribution, random oversampling, random under-sampling and synthetic minority oversampling technique (SMOTE) are the most common used data sampling methods. With random undersampling technique, all minority instances are reserved and some majority instances are randomly removed, which will lead to serious information loss of majority class for small-sampling dataset. With SMOTE, new synthetic data for the minority training set is generated by randomly interpolating pairs of nearest neighbors. However, SMOTE is not very effective for high-dimensional data [39]. For classification of the imbalanced dataset, Rahman *et al.* [40] proposed an improved under-sampling technique to balance cardiovascular data, and Niehaus [41] applied the random under-sampling technique to factor out the class imbalance. Ren *et al.* proposed an adaptive over-sampling algorithm in [42], which has gained good performance in detection of microaneurysm. Another category of methods to tackle imbalanced dataset is the weighted approach. Weighted SVM [43] and weighted ELM [44] have been successfully applied to handle with many imbalanced issues.

In order to address these problems, we propose the MOCA framework, which is effective even when only imbalanced small-sampling data is available. In this framework, we first propose the IRFFS-O approach to reduce the feature space dimensionality and reduce the problem of over-fitting. The random over-sampling technique is adopted to balance the training data, and repeated random sampling and iterative mechanism are used to reduce the effect caused by perturbation on small-sampling datasets. Then, we propose a novel classification model b-WELM which adds a bias in the optimization problem of weighted ELM to improve the generalization ability which enables the framework to distinguish only a few patients with cognitive disorders from a large number of normal control subjects better.

3 The MOCA Framework

In Figure 1, we give an illustration of the proposed MOCA framework. The most important components of MOCA are feature selection and classification. In addition, data collection and feature extraction are also presented.

In our MOCA framework, feature selection aims to reduce the feature space dimensionality and reduce the problem of over-fitting. With this measure, we can simplify the activity data collected. Only the motion corresponding to the selected features should be collected in the subsequent data collection phase. Once the optimal features are selected, we hope to build a light classification model, which is able to reduce the rate of missed diagnosis by appropriately improving the sensitivity. With such a classification model, we are able to incorporate it into a light smart devices to assist the diagnosis. Thus, we propose a light-weight classification model – b-WELM, which is very effective in handling imbalanced datasets.

3.1 Data Collection

Activity data collected in this step are usually summarized and simplified according to several existing and widely accepted scales. These include the Scale for Assessment and Rating of Ataxia [45], the Short Physical Performance Battery [46], the Tinetti Mobility Test (TMT) [47] and the Unified Parkinson Rating Scale (UPDRS) [24]. Different data collection schemes are designed for different research purposes. For example, the activities from the UPDRS are summarized to study the correlation between motor pattern and PD. As TMT is an effective indicator for brain damages [48] and a powerful predictor of stroke [49], it is usually used for stroke detection.

Various data collection devices (e.g., Kinect, smartphone and smart band) can be used to collect motion related data from the subjects in this step. Different data collection devices are suitable for different data acquisition requirements. In order to reduce the disturbance of data collection devices on the subjects, we collect the raw motion data using the depth sensor and the RGB camera embedded in Kinect. The data acquired by Kinect is usually recorded in the format of depth image, RGB image and skeleton data. Accelerometers and gyroscopes embedded in wearable devices (e.g., smart band, smartwatch and smartphone) are commonly used to collect motion data in daily life for a long time. The readings from accelerometers and gyroscope record the accelerated velocity and angular velocity respectively. Electromyographic (EMG) armband mounted on the upper arm, which contains some electrodes, is a kind of wearable EMG devices. The data from EMG armband is electric potential generated by muscle cells. Wearable EMG armband, which is made up of electrodes, can be used to collect EMG signal in a wearable way. The data acquired from EMG armband is a set of electric potentials generated by muscle cells.

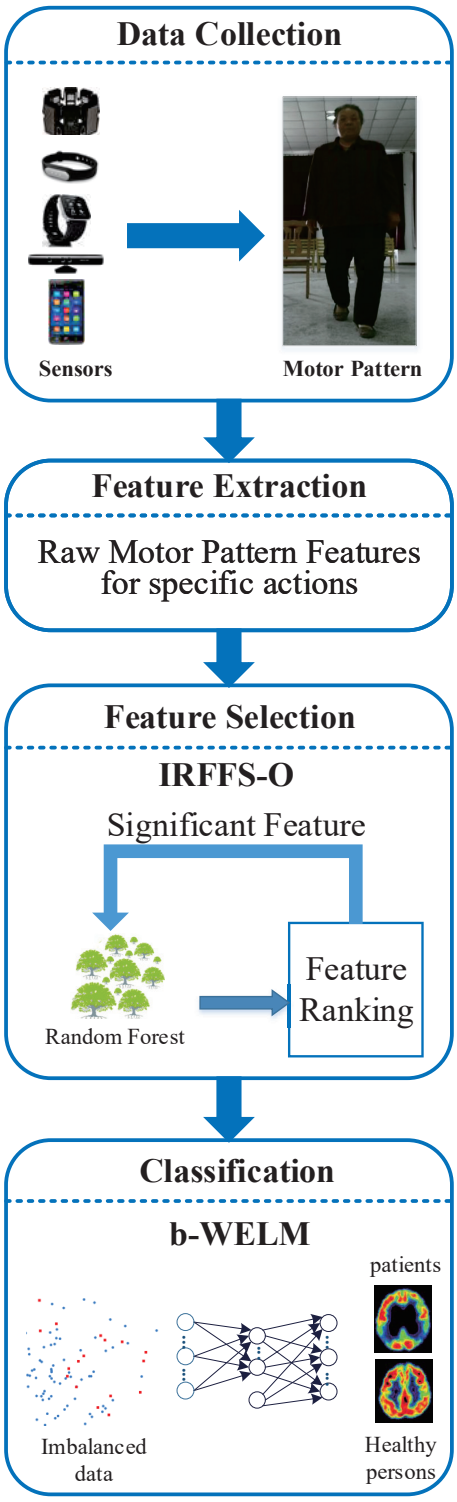


Fig. 1. The proposed MOCA framework.

3.2 Feature Extraction

Feature extraction is commonly used to identify numeric variables that can reflect the intrinsic data properties [50]. In this step, we try to uncover hidden information from raw motion data which can reflect the correlation analysis between motion behavior and cognitive ability. In this work, we focus on three types of features: 1) time domain features, 2) frequency domain features and 3) motor features. Time domain features commonly refer to time-varying features (e.g., mean, standard deviation, zero crossing rate and maximum/minimum). Frequency domain features are often used to find out the periodical information of signals (e.g., direct current, amplitude and power spectral density). Motor features are related to specific actions performed by the person being monitored. These features are used to extract the characteristics of actions. Medical instructions from doctors and characteristics used in clinical cognitive assessment are taken into consideration in the feature extraction step of MOCA, which makes the research more reasonable. Though the selection of motor features is largely dependent on the empirical knowledge of researchers, these features are able to reflect the intrinsic characteristics of motor pattern, which is useful for the correlation analysis.

3.3 Feature Selection

Through feature extraction, we have obtained sufficient features to describe the acquired motor patterns. However, there are two main problems to be solved. On one hand, the number of extracted features could be huge. This high dimensional data will cause the curse of dimensionality. On the other hand, a large number of irrelevant features will increase the difficulty for further correlation analysis. Feature selection is a useful data preprocessing method to deal with these issues. Thus, in Step 3, feature selection is performed to reduce the feature space dimensionality and reduce the problem of over-fitting.

In this subsection, we propose a novel feature selection method, named random Oversampling Iterative Random Forest-based Feature Selection (IRFFS-O) which is an extension to the random forest feature selection technique [51]. This method automatically measures the importance of the extracted features, retain the most significant features and discard unnecessary ones. Its main advantage is the ability to handle imbalanced small-sampling dataset.

The whole process of IRFFS-O is first briefly introduced. Before constructing the random forests, we randomly divide the whole dataset into training sets and testing sets repeatedly and perform the random over-sampling technique on each training set. Then, the IRFFS-O enters an iterative process. Several random forests will be constructed following the learning method proposed by Breiman *et al.* [52]. In each loop, IRFFS-O ranks the features according to their occurrence frequencies and select the most common and frequent features to rebuild new random forests. The process is repeated until no further classification performance improvements can be achieved. Let $\boldsymbol{T} = (\boldsymbol{x}_i, \boldsymbol{t}_i) \in \boldsymbol{R}^n \times \boldsymbol{R}^m, i \in \{1, 2, ..., N\}$ denote the initial training set, where $\boldsymbol{x}_i = [x_{i1}, x_{i2}, ..., x_{in}]^T$ is an input vector with n features, and $\boldsymbol{t}_i = [t_{i1}, t_{i2}, ..., t_{im}]^T$ is the corresponding target vector. $n(t)$ is the number of features in the tth iteration.

Figure 2 shows the flow chart of IRFFS-O method. The main procedure of IRFFS-O is described as the following:

1) Initialization. Set iterative index $t = 0$, initial training set $\boldsymbol{T}(0) = \boldsymbol{T}$, initial average classification accuracy ($\overline{acc} = 0$).
2) Random sampling. Randomly sample N_r samples from the training set $\boldsymbol{T}(t)$ for S times, obtaining S sub-training sets $\boldsymbol{T}^s(t) = \{\boldsymbol{x}_i, \boldsymbol{t}_i\} \in \boldsymbol{R}^{n(t)} \times \boldsymbol{R}^m, i \in \{1, 2, ..., N_r\}$ and S sub-testing sets $\boldsymbol{V}^s(t) = \{\boldsymbol{x}_i, \boldsymbol{t}_i\} \in \boldsymbol{R}^{n(t)} \times \boldsymbol{R}^m, i \in \{N_r + 1, N_r + 2, ..., N\}, s \in \{1, 2, ..., S\}$.
3) Random oversampling on sub-training set. Samples from the minority class are randomly selected with replacement to balance the sub-training sets $\boldsymbol{T}^s(t)$. After random oversampling, the balanced sub-training sets are represented by $\boldsymbol{T}^{'s}(t)$.
4) Random forest building. Build S random forests based on S balanced sub-training sets $\boldsymbol{T}^{'s}(t)$ using the learning method proposed in [51]. Here, the number of trees in each random forest is set to B, and $\lfloor\sqrt{(n(t)}\rfloor$ features are used in each split.
5) Feature validation. Compute the average classification accuracy $\overline{acc}(t)$ of the S built random forests using the corresponding sub-testing sets $\boldsymbol{V}^s(t)$. If $\overline{acc}(t)$ is higher than $\overline{acc}$, set $\overline{acc}$ to $\overline{acc}(t)$ and go to Step 6; else drop the iterative process and select the current $n(t)$ features as the most significant features.
6) Feature sorting. In order to measure the importance of each feature $f_j, j \in \{1, 2, ..., n(t)\}$, we introduce two variables: (1) the number of random forests where f_j occurs, denoted as N_j, and (2) the frequency at which f_j appears in the random forests, denoted as F_j. N_j can be counted directly. F_j is computed as follows:

$$F_j = \sum_{s=1}^{S} \sum_{b=1}^{B} \frac{N_{b_j}^s}{N_b^s} \tag{1}$$

where $N_{b_j}^s$ and N_b^s denote the number of internal nodes testing on feature f_j and the number of total internal nodes in the bth tree of the sth random forest, respectively. Since a more discriminant feature would be selected in more trees of more random forests built using different training sets, a feature with large N_j and F_j values is of greater importance. IRFFS-O then sorts the $n(t)$ features according to N_j and F_j in descending order.
7) Feature selection. In this step, N_j is firstly considered. When the number of built random forests where f_j occurs is the same as another feature f_i, then we further compare F_i and F_j. The feature with a larger value of F is considered more important. Select the top r ranked features as the most discriminant features. Since a large r would cause redundant looping, while a small r may discard significant features, we set r as half of the current number of features, i.e. $r = \lfloor\frac{n(t)}{2}\rfloor$. If r is greater than the minimum number of features $n_{\min}$, assign r to $n(t+1)$ and go to Step 8; else drop the iterative process and select the current $n(t)$ features as the most significant features.
8) Loop. Set $t = t + 1$ and go to Step 4.

With a small-sampling dataset, even a little perturbation would cause a large difference in the out-of-bag error, which will result in a very different feature selection result. Thus, we adopt the repeated random sampling technique and iterative mechanism in

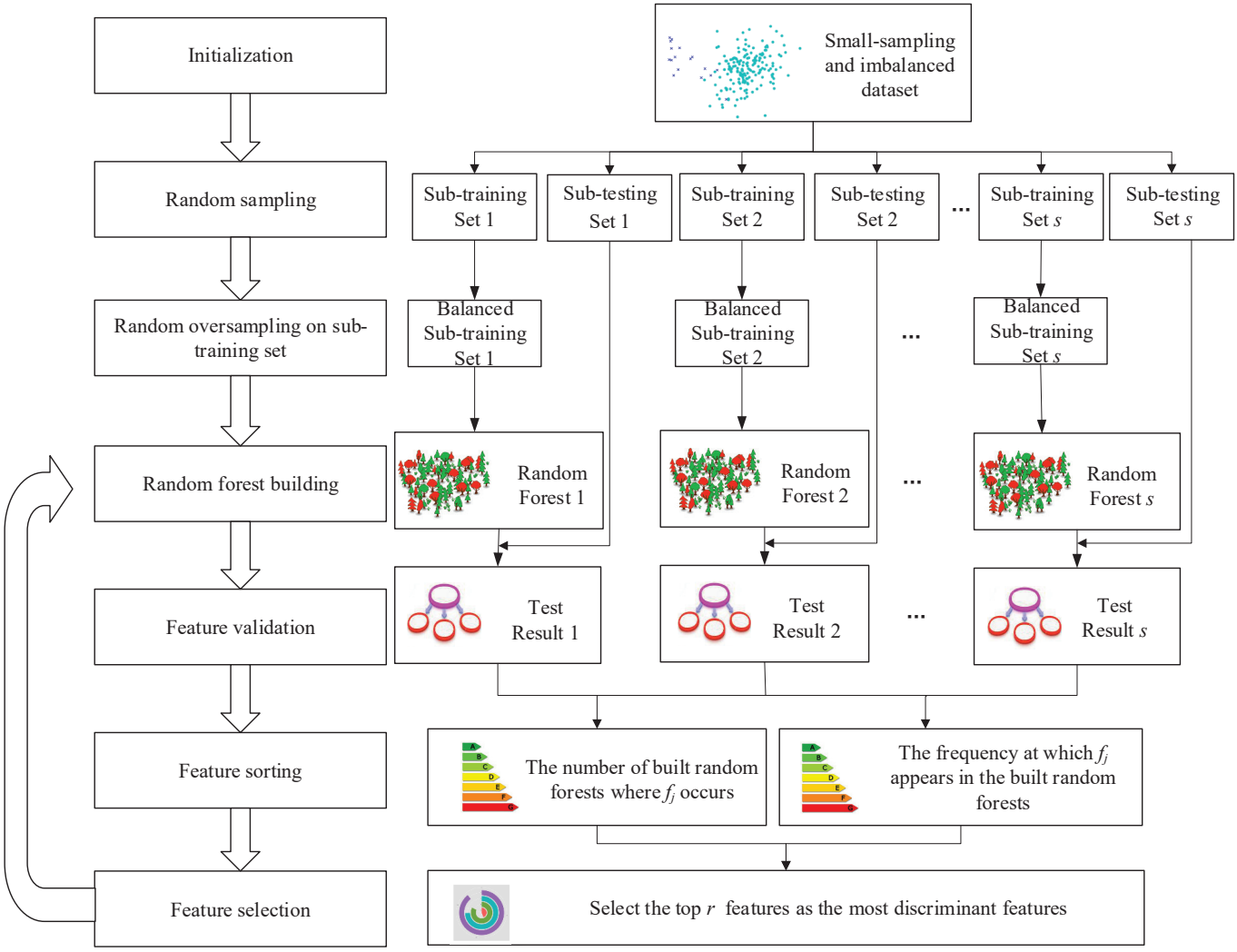


Fig. 2. The flow chart of the IRFFS-O method.

the proposed IRFFS-O method. The repeated random sampling, which corresponds to step 2), is used to decrease the perturbation caused by a single division, and the iterative mechanism is used to reduce the perturbation caused by a singe feature selection process and select the optimal features step by step.

To avoid the problems caused by imbalanced data distribution, the random over-sampling technique (presented in step 3)) is used in IRFFS-O. For each training set, we perform the random over-sampling technique, which will contribute to balanced training sets. Then, the impact of imbalanced data distribution on feature selection is decreased before the building of random forests.

3.4 Classification

In order to build an accurate correlation analysis model with imbalanced data, we propose a novel classification approach – b-WELM. b-WELM is a learning method built on b-COELM [53] and weighted Extreme Learning Machine (weighted ELM) [44], which achieves good generalization in dealing with data with imbalanced class distribution. Since it is difficult to obtain patient data while comparatively easy to collect data from healthy people, the bias in performance caused by imbalanced class distribution is very common in disease detection. b-WELM is designed to distinguish patients with cognitive disorders from normal control subjects.

Extreme Learning Machine (ELM) is a single hidden layer feedforward neural network, which is proposed as a unified framework for binary, multiclass classification and regression problems [54]. In ELM, learning is made without iterative tuning and is very efficient and effective when the training set is small [55]. With comparable predictive accuracy with SVM, ELM performs with a faster learning speed. To improve the generalization ability of ELM, Hu *et al.* [53] proposed the b-COELM by adding a bias item in the optimization problem of COELM. To cope with imbalanced data distribution in the training samples, weighted ELM [44] constructs a model using the weighted trade-off parameters for different training samples. In this paper, we propose a novel learning method, named b-WELM, with better generalization ability in handling imbalanced datasets.

Given N arbitrary distinct samples $(\boldsymbol{x}_i, \boldsymbol{t}_i) \in \boldsymbol{R}^n \times \boldsymbol{R}^m, i \in \{1, ..., N\}$, where $\boldsymbol{x}_i$ is a $n \times 1$ input vector $\boldsymbol{x}_i = [x_{i1}, ..., x_{in}]^T$ and $\boldsymbol{t}_i$ is an $m \times 1$ target vector $\boldsymbol{t}_i = [t_{i1}, ..., t_{im}]^T$. The network architecture of b-WELM is shown in figure 3.

The architecture of b-WELM consists of three layers (from

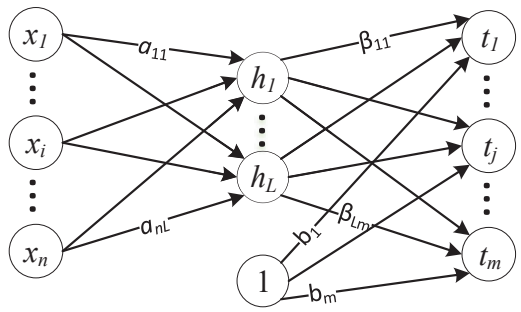


Fig. 3. The network architecture of b-WELM.

left to right): the input layer, the hidden layer and the output layer. Given each arbitrary sample $(\boldsymbol{x}, \boldsymbol{t})$, the n nodes in the input layer correspond to the features of the input vector $\boldsymbol{x}$. In the hidden layer, $\boldsymbol{x}$ is nonlinearly mapped into the vector $\boldsymbol{h}(\boldsymbol{x}) = (h_1(\boldsymbol{x}), \ldots, h_L(\boldsymbol{x}))^T$ in an L-dimensional feature space. h_k represents the kth element of the L-dimensional vector when $\boldsymbol{x}$ is not specified. Notations a_{ik} from the input layer to the hidden layer in Figure 3 are the parameters relevant to the nonlinear mapping. In the output layer, $\boldsymbol{h}(\boldsymbol{x})$ is mapped into the target vector $\boldsymbol{t}$ according to a linear transformation with two variables, an $L \times m$ dimensional weight matrix $\boldsymbol{\beta}$ and a bias vector $\boldsymbol{b} = (b_1, \ldots, b_m)^T$, both can be analytically solved.

The primal optimization problem of b-WELM is as follows:

$$\min_{\boldsymbol{\beta},\boldsymbol{b},\boldsymbol{\xi}} \frac{1}{2}(||\boldsymbol{\beta}||^2 + ||\boldsymbol{b}||^2) + \frac{C}{2}\sum_{i=1}^{N} W_{ii}||\boldsymbol{\xi}_{:,i}||^2 \tag{2}$$

s.t.

$$\boldsymbol{\beta}^T\boldsymbol{h}(\boldsymbol{x}_i) + \boldsymbol{b} = \boldsymbol{t}_i - \boldsymbol{\xi}_{:,i}, i \in \{1, ..., N\} \tag{3}$$

Several parameter values need to be determined beforehand. C is the penalty parameter balancing the maximum generalization ability (the first term of objective function) and minimum training error (the second term of objective function). $\boldsymbol{W}$ is a $N \times N$ dimensional diagonal weight matrix in which W_{ii} is the weight of sample $\boldsymbol{x}_i$.

Based on the parameters above, when the training process starts, the two variables $\boldsymbol{\beta}$ and $\boldsymbol{b}$ are computed. $\boldsymbol{\beta}$ is an $L \times m$ dimensional output matrix, $\boldsymbol{b}$ is an $m \times 1$ bias vector $\boldsymbol{b} = (b_1, \ldots, b_m)^T$. Besides, $\boldsymbol{\xi}$ is an $m \times N$ dimensional slack variable in which the ith column of $\boldsymbol{\xi}$, denoted by $\boldsymbol{\xi}_{:,i}$, is the training error of the instance $\boldsymbol{x}_i$.

In (2), $\boldsymbol{W}$ is the weight matrix, which is associated with individual training instance $\boldsymbol{x}_i$. If $\boldsymbol{x}_i$ belongs to a minority class, W_{ii} will be assigned a larger value. In contrast, W_{ii} will be assigned a smaller value if $\boldsymbol{x}_i$ belongs to a majority class. Thus, with $\boldsymbol{W}$, the weights of the minority and majority classes are rebalanced with this technique during the classification process. With the bias item '$\boldsymbol{b}$', b-WELM has strong convexity and satisfies the non-essential requirement of Mercer's positive definition condition [56]. With both $\boldsymbol{W}$ and $\boldsymbol{b}$ introduced, b-WELM achieves a better generalization ability in handling imbalanced datasets.

Based on the Karush-Kuhn-Tucker (KKT) theorem, the problem in (2) is equivalent to its dual problem. The objective function of its dual problem is as shown in (4).

$$\begin{aligned} L_{D_{\text{b-WELM}}} = &\frac{1}{2}(||\boldsymbol{\beta}||^2 + ||\boldsymbol{b}||^2) + \frac{C}{2}\sum_{i=1}^{N} W_{ii}||\boldsymbol{\xi}_{:,i}||^2 \\ &- \sum_{i=1}^{N}\sum_{j=1}^{m} \alpha_{i,j}(\boldsymbol{\beta}_{:,j}^T\boldsymbol{h}(\boldsymbol{x}_i) + b_j - t_{i,j} + \xi_{j,i}) \end{aligned} \tag{4}$$

$\boldsymbol{\alpha}$ is a $N \times m$ matrix in which the ith column of $\boldsymbol{\alpha}$ is the Lagrange multiplier of sample x_i. We refer to $\boldsymbol{\alpha}$ as the Lagrange matrix. Furthermore, let the partial derivative of $L_{D_{\text{b-WELM}}}$ with respect to all the variables be 0, we have the following equations:

$$\frac{\partial L_{D_{\text{b-WELM}}}}{\partial \boldsymbol{\beta}_{:,j}} = \boldsymbol{\beta}_{:,j} - \sum_{i=1}^{N} \alpha_{i,j}\boldsymbol{h}(\boldsymbol{x}_i) = \boldsymbol{0}, \forall j \Leftrightarrow \boldsymbol{\beta} = \boldsymbol{H}^T\boldsymbol{\alpha} \tag{5a}$$

$$\frac{\partial L_{D_{\text{b-WELM}}}}{\partial b_j} = b_j - \sum_{i=1}^{N} \alpha_{i,j} = 0, \forall j \Leftrightarrow \boldsymbol{b} = \boldsymbol{\alpha}^T\boldsymbol{1}_{vec} \tag{5b}$$

$$\begin{aligned} &\frac{\partial L_{D_{\text{b-WELM}}}}{\partial \xi_{j,i}} = CW_{ii}\xi_{j,i} - \alpha_{i,j} = 0, \forall i, j \\ &\Leftrightarrow C\boldsymbol{\xi}\boldsymbol{W} = \boldsymbol{\alpha}^T \Leftrightarrow \boldsymbol{\xi}^T = \frac{1}{C}\boldsymbol{W}^{-1}\boldsymbol{\alpha} \end{aligned} \tag{5c}$$

$$\begin{aligned} &\frac{\partial L_{D_{\text{b-WELM}}}}{\partial \alpha_{i,j}} = \boldsymbol{\beta}_{:,j}^T\boldsymbol{h}(\boldsymbol{x}_i) + b_j - \boldsymbol{t}_{i,j} + \xi_{j,i} = 0, \forall i, j \\ &\Leftrightarrow \boldsymbol{H}\boldsymbol{\beta} + \boldsymbol{1}_{vec}\boldsymbol{b}^T + \boldsymbol{\xi}^T = \boldsymbol{T} \end{aligned} \tag{5d}$$

$\boldsymbol{H}$ is a $N \times L$ matrix in which each row is $\boldsymbol{h}(\boldsymbol{x}_i)^T$. $\boldsymbol{1}_{vec}$ is a $N \times 1$ vector with all elements being 1. Substitute (5a) - (5c) into (5d), we can get (6).

$$\boldsymbol{\alpha}^* = (\boldsymbol{W}\boldsymbol{H}\boldsymbol{H}^T + \boldsymbol{W}\boldsymbol{1}_{mat} + \frac{1}{C}\boldsymbol{I})^{-1}\boldsymbol{W}\boldsymbol{T} \tag{6}$$

$\boldsymbol{1}_{mat}$ is a $N \times N$ matrix with all elements being 1. Substitute (6) into (5a), we can get:

$$\boldsymbol{\beta}^* = \boldsymbol{H}^T\boldsymbol{\alpha}^* \tag{7}$$

$$\boldsymbol{b}^* = (\boldsymbol{\alpha}^*)^T\boldsymbol{1}_{vec} \tag{8}$$

At last, the output function of b-WELM is:

$$\boldsymbol{f}(\boldsymbol{x}) = (\boldsymbol{\beta}^*)^T\boldsymbol{h}(\boldsymbol{x}) + \boldsymbol{b}^* = (\boldsymbol{\alpha}^*)^T(\boldsymbol{H}\boldsymbol{h}(\boldsymbol{x}) + \boldsymbol{1}_{vec}) \tag{9}$$

For any testing sample $\boldsymbol{x}$, $\boldsymbol{f}(\boldsymbol{x})$ is a $m \times 1$ vector. The prediction of b-WELM on $\boldsymbol{x}$ is in (10). $f_k(\boldsymbol{x})$ is the kth element of $\boldsymbol{f}(\boldsymbol{x})$.

$$Prediction(\boldsymbol{x}) = \underset{k\in\{1,\ldots,m\}}{\arg\max}\ f_k(\boldsymbol{x}) \tag{10}$$

In (6), replace $\boldsymbol{H}\boldsymbol{H}^T$ in (6) by the Gram matrix $\boldsymbol{\Omega}(\boldsymbol{\Omega}_{ij} = k(\boldsymbol{x}_i, \boldsymbol{x}_j), i, j = 1, ..., N)$ of a kernel $k(\boldsymbol{u}, \boldsymbol{v})$ and $\boldsymbol{H}\boldsymbol{h}(\boldsymbol{x})$ in (9) by the kernel form, we have the output function of kernel based b-WELM:

$$\begin{aligned} \boldsymbol{f}(\boldsymbol{x}) = \boldsymbol{T}^T\boldsymbol{W}(\boldsymbol{\Omega}\boldsymbol{W} + \boldsymbol{1}_{mat}\boldsymbol{W} + \frac{1}{C}\boldsymbol{I})^{-1}([k(\boldsymbol{x}_1, \boldsymbol{x}), \\ ..., k(\boldsymbol{x}_N, \boldsymbol{x})]^T + \boldsymbol{1}_{vec}) \end{aligned} \tag{11}$$

4 Experimental Evaluation

To study the effectiveness of the proposed framework, we conduct experiments on two datasets. The first dataset is collected on SVD patients, and the second dataset is gathered from stroke patients.

4.1 Comparison methods

Six feature selection methods embedded in WEKA [19] and IRFFS [18] are used as comparison methods against the proposed IRFFS-O approach. They are:

- CorrelationAttributeEval. Evaluates the usefulness of a subset of attributes by considering the individual predictive power of each feature along with the degree of redundancy among them;
- GainRatioAttributeEval. Evaluates the usefulness of an attribute by measuring the gain ratio with respect to the class;
- InfoGainAttributeEval. Evaluates the usefulness of an attribute by measuring the information gain with respect to the class;
- OneRAttributeEval. Evaluates the usefulness of an attribute by using the OneR classifier;
- ReliefFAttributeEval. Evaluates the usefulness of an attribute by repeatedly sampling an instance and considering the value of the given attribute for the nearest instance of the same and different class;
- SymmetricalUncertAttributeEval. Evaluates the usefulness of an attribute by measuring the symmetrical uncertainty with respect to the class;
- IRFFS. Evaluates the usefulness of attributes with the method proposed in [18].

Four state-of-the-art classification methods are used for comparison with the proposed b-WELM approach. They are:

- b-constrained-optimization-based extreme learning machine (b-COELM) [53];
- Weighted extreme learning machine (weighted ELM) [44];
- Weighted support vector machine (weighted SVM) [57];
- Back propagation (BP) algorithm [58].

4.2 SVD Detection

4.2.1 Dataset Description

TABLE 1
The MRI Characteristics of the SVD Patients

No.	Sex	Age	Fazekas Scale (periventricular white matter	Fazekas Scale (deep white matter)	No. lacunar infarcts
1	F	73	3	3	1
2	M	74	3	3	0
3	M	72	3	2	0
4	M	72	3	3	3
5	M	69	2	2	6
6	F	62	2	2	2
7	F	68	3	2	0
8	F	66	3	3	4

Eight SVD patients and 43 healthy subjects were recruited for our experiments. The SVD patients are aged from 62 to 76 and consist of 4 women and 4 men. The healthy subjects are aged from 22 to 76 and consist of 17 women and 26 men. Each subject corresponds to one sample in the dataset. Thus, there are 51 samples in SVD detection dataset with 8 samples of SVD patients and 43 samples of healthy subjects. Each subject was asked to perform all 17 motor actions in order. The completion time for each subject ranged from 3.05 minutes to 8.42 minutes. The individual and MRI characteristics of the 8 SVD patients are shown in Table 1.

4.2.2 Data Collection

We use Kinect to collect raw depth sensor data and RGB camera data of 17 predefined clinical actions. These 17 actions are classified into three categories: **gait**, **balance** and **agility**.

- Gait Test. Gait test includes natural walking test, 3-meter walking and 180 degree turning test. A subject is asked to walk in front of the Kinect camera. Crutches are allowed if the subject has difficulty walking unaided. Figure 4(a) shows the data collection process of 3-meter walking and 180 degree turning test as an example of the gait test. The subject is asked to walk 3 meters in a natural way, then turned back and walk back to the starting line.
- Balance Test. In these tests, a subject is asked to perform some specific actions to test their balance capability. These actions include standing up from a seated position, repeated standing up from a seated position with arms folded across the chest, pulling test, side by side standing, semi-tandem standing, tandem standing, tandem walking and nudge. If there is any notable left or right skew, trend of falling down or difficulty to complete the test, the test will immediately stop. Figure 4(b) shows the data collection process of tandem walking as an example of balance test. The subject is asked to walk in a straight line with the toe of one foot touching the heel of the other foot.
- Agility Test. In order to judge whether a subject has agility problems, we ask the subject to perform actions as quickly as possible. The tests include: finger tracking, finger-to-nose test, rapid alternating up and down movements of the hand, rapid alternating movements of the hand, finger tapping and tapping heel on the ground in rapid successions while raising the entire leg. Figure 4(c) shows the data collection process of finger tracking and tapping heel on the ground as examples of agility test. In finger tracking, the subject is asked to track the doctor's finger with his own forefinger as quickly as possible. In tapping heel on the ground, the subject is asked to raise the heel up and down as quickly as possible 10 times.

4.2.3 Feature Extraction

Some potential SVD related motor pattern features are extracted from the collected dataset. These features include: 1) gait related features, such as stride length, step width, step height, walking velocity, and sagittal and coronal-angular excursions of the shoulder, elbow, hip, knee and trunk, which are calculated as the method proposed in [59]; 2) balance related features, such as angular excursions of the trunk and stand velocity; and 3) agility related features, including smoothness of movement trajectory and variability of movement velocity. In total, 739 features are extracted.

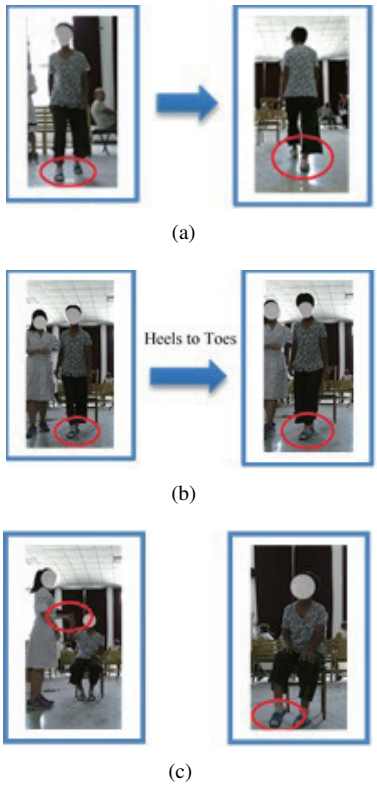


Fig. 4. Data collection for SVD test: (a) Gait test. (b) Balance test. (c) Agility test.

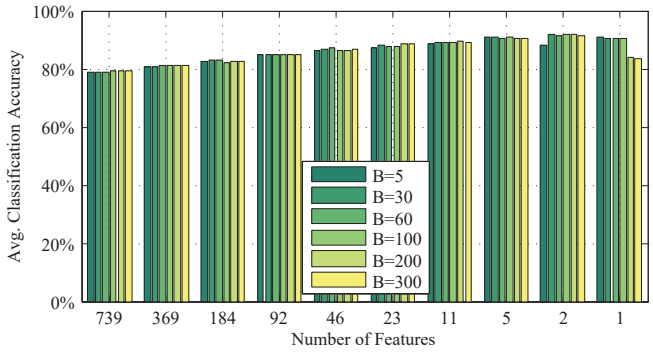


Fig. 5. The average classification accuracy of the random forest model at each iteration using the IRFFS-O method on the SVD dataset.

TABLE 2
Two Most Discriminant Features for SVD Detection (Selected by the IRFFS-O Method with $B = 100$)

Feature Type	Motor Action	Feature Extracted
Agility related feature	Rapid alternating hand movements	Mean pronation velocity of the right hand
Balance related feature	tandem walking	Mean of the maximum angular excursions of the trunk in the z direction

4.2.4 *Feature Selection*

We perform feature selection experiment on the data of the 51 subjects with IRFFS-O. In our experiment, the training rate is set as 70%. Thus, the number of training samples is $N_r = \lfloor 43 \times 70\% \rfloor + \lfloor 8 \times 70\% \rfloor = 35$ and the size of testing set is $51 - 35 = 16$. To eliminate the splitting effect, we set the sampling time to 1,000 and the experimental results are averaged. To find out the most discriminant features, IRFFS-O tests different sets of features from all 739 features to a single feature.

The average accuracy at each iteration of IRFFS-O is presented in figure 5. With the number of features decreases from 739 to 2, the average classification accuracy shows a continuous rising trend. The classification accuracy is improved from 78.76%–79.46% to 88.18%–91.86%. This indicates that the proposed IRFFS-O method is able to discard redundant features and preserve significant features continuously. With the number of features decreases from 2 to 1, the average classification accuracy drops from 88.18%–91.86% to 83.67%–90.80%, which means some useful features are discarded in this iteration. Thus, the iterative process is terminated and 2 most discriminant features are selected. With the number of features set as 2, the average classification accuracy obtains the highest value at $B = 100$. Then, we take the feature selection results at this parameter settings as the most significant features. The selected features are shown in Table 2. As we can see, the selected features include an agility related feature and a balance related feature. With this experimental result, we can infer that rapid alternating movements of the hand and tandem walking may be more relevant to the detection of SVD.

To further validate the discriminating ability of the feature selected by IRFFS-O, we conduct a group of comparison experiments against IRFFS [18] and 6 other feature selection methods provided in WEKA [19]. Ten classification tests with AdaBoostM1 and MultilayerPerceptron are performed in WEKA on the features selected by different feature selection methods with default parameter settings. Ten folds cross-validation is adopted in this experiment. The average test accuracies are listed in Table 3. From Table 3, it can be observed that the proposed IRFFS-O and IRFFS both gains the best classification results with AdaBoostM1 algorithm. While with the MultilayerPerceptron algorithm, IRFFS-O slightly outperforms IRFFS. Overall, the test accuracy on SVD dataset of both AdaBoostM1 and MultilayerPerceptron algorithm based on two features selected by the IRFFS-O method is better than that based on two features selected by other feature selection methods, which indicates that the IRFFS-O method is more effective in selecting the most significant features for accurate SVD detection from imbalanced small-sampling data.

TABLE 3
The SVD Detection Performance of AdaBoostM1 and MultilayerPerceptron based on the Two Features Selected by Different Feature Selection Methods

Feature Selection Method	AdaBoostM1	MultilayerPerceptron
CorrelationAttributeEval	90.63%	93.13%
GainRatioAttributeEval	92.20%	89.81%
InfoGainAttributeEval	92.39%	92.48%
OneRAttributeEval	90.97%	95.44%
ReliefFAttributeEval	90.07%	84.79%
SymmetricalUncertAttributeEval	84.79%	89.93%
IRFFS	**94.11%**	95.49%
IRFFS-O	**94.11%**	**95.63%**

4.2.5 Classification

With the features selected by IRFFS-O, we adopt the proposed b-WELM algorithm to construct an accurate model with high classification accuracy and G-mean.

To evaluate the effectiveness of b-WELM, the repeated random sub-sampling (RRSS) technique is used to generate the training and testing datasets separately. Besides, b-COELM, weighted ELM, weighted SVM and back-propagation (BP) are used for comparisons.

Parameter Selection Gaussian kernel $(k(\boldsymbol{x}_i, \boldsymbol{x}_j) = exp(-g\|\boldsymbol{x}_i - \boldsymbol{x}_j\|^2))$ is adopted by all models except for BP. There are two parameters, penalty parameter C and kernel parameter g, used by b-WELM, b-COELM, weighted ELM and weighted SVM; and another two parameters, number of hidden nodes L and learning rate lr, used by BP. In order to fairly compare the performances of the five models, we need to find the optimal parameter pairs for each algorithm. Grid search technique is employed in this paper. Fifty different values $\{2^{-24}, 2^{-23}, \cdots, 2^{24}, 2^{25}\}$ are tested for both C and g, and 15 different values $\{2^{-15}, 2^{-14}, \cdots, 2^{-1}\}$ are tested for lr. For BP, the number of input nodes is commonly set between the number of features (2 in this experiment) and the number of output nodes is commonly set as 1 [60]. Thus, 2 different values $\{2^0, 2^1\}$ are tested for L.

With RRSS technique, we select 70% of the samples for training and 30% of the samples for testing. Therefore, the training set and testing set consist of 35 and 16 instances, respectively. The weights in b-WELM, weighted ELM and weighted SVM are both set as $1/\#(t_i)$, where $\#(t_i)$ is the number of samples belonging to class $t_i (i = 1, \cdots, m)$. Figure 6 presents the testing accuracy of b-WELM, b-COELM, weighted ELM, weighted SVM and BP with different parameter pairs. The optimal parameter values for each model are listed in Table 4.

TABLE 4
The optimal parameter for each model

	b-WELM	b-COELM	weighted ELM	weighted SVM	BP
C (L)	2^{-6}	$2^{11}/2^{21}$	2^{-5}	2^{1}	2^{0}
g (lr)	2^{-8}	$2^{3}/2^{5}$	2^{-9}	2^{6}	2^{-1}

Classification Performance To evaluate the performance of b-WELM on SVD detection, we adopt test accuracy, test sensitivity, test specificity and test G-mean as evaluation metrics. They are calculated as follows:

$$\text{Accuracy} = \frac{t_p + t_n}{t_p + f_p + t_n + f_n} \tag{12}$$

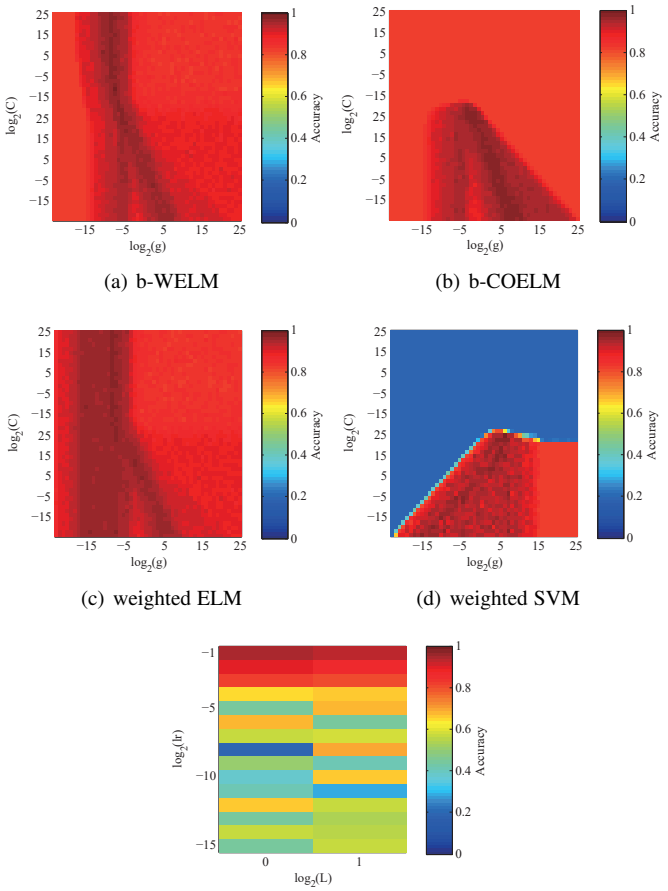
$$\text{Sensitivity} = \frac{t_p}{t_p + f_n} \tag{13}$$

$$\text{Specificity} = \frac{t_n}{t_n + f_p} \tag{14}$$

$$\text{G-mean} = \sqrt{\text{Sensitivity} \times \text{Specificity}} \tag{15}$$

where t_p, f_p, t_n and f_n denote true positive, false positive, true negative and false negative, respectively.

Accuracy is one of the most common metrics used to evaluate the effectiveness of a classifier. For imbalanced dataset, sensitivity and specificity are able to give more insight into the accuracy obtained within each class. G-mean provides an overall view to evaluate a classifier, which is calculated as the geometric mean of sensitivity and specificity. In this section, b-COELM, weighted ELM, weighted SVM and BP are presented as comparisons. Experiments with the optimal parameters are repeated for 1,000 times. Table 5 shows the average SVD detection performances of b-WELM, b-COELM, weighted ELM, weighted SVM and BP based on feature selected by the IRFFS-O method with $B = 100$.



(a) b-WELM (b) b-COELM (c) weighted ELM (d) weighted SVM (e) BP

Fig. 6. Experimental results on the accuracy of b-WELM, b-COELM, weighted ELM, weighted SVM and BP with different parameter pairs.

TABLE 5
The SVD Detection Performance

	Test Accuracy	Test Sensitivity	Test Specificity	Test G-mean
b-WELM	**97.87%**	**100.00%**	**97.38%**	**98.66%**
b-COELM($C = 2^{11}, g = 2^3$)	96.98%	**100.00%**	96.28%	98.09%
b-COELM($C = 2^{21}, g = 2^5$)	97.05%	99.87%	96.40%	98.07%
Weighted ELM	97.08%	98.70%	96.71%	97.60%
Weighted SVM	97.83%	**100.00%**	97.33%	98.64%
BP	95.08%	87.07%	96.92%	88.48%

As shown in Table 5, b-WELM outperforms the other four algorithms in accuracy, sensitivity, specificity, and G-mean. Among these methods, BP achieves the worst performance in accuracy, sensitivity, and G-mean. Except for BP, the advantages of b-WELM over the other three methods are not significant. This is an indirect evidence that the features selected by the proposed IRFFS-O approach are significant in distinguishing SVD patients from healthy persons, which contribute to good classification results of different classifiers. The average accuracy of b-WELM is 97.87%, which is the highest among all comparison approaches.

With imbalanced data distribution, SVD patients are the minority in the dataset. Generally, there is a natural tendency that the majority class (i.e. healthy people) are favored in comparison to the minority class (i.e. SVD patients). In other words, SVD patients are more prone to be misclassified as healthy people. However, with b-WELM, all SVD patients have been classified correctly (i.e. the test sensitivity is as high as 100.00%). This is especially important to disease detection. For specificity, although all five methods obtained good performance, b-WELM performs the best, which means b-WELM can also classify the healthy people better. G-mean is the square root of (positive class accuracy $\times$ negative class accuracy). It is usually used to measure the overall effectiveness of a classifier. As shown in Table 5, b-WELM is able to achieve the highest score in G-mean among all comparison approaches.

The above experimental results suggest that there is a strong correlation between the motor pattern changes and SVD, which is consistent with clinical observations. Besides, our MOCA framework is able to reduce the rate of missed diagnosis by appropriately improving the sensitivity of the algorithm. Therefore, the proposed framework is shown to be useful in the diagnosis of SVD and thus can assist doctors in the correlation analysis between motor patterns and SVD.

4.3 Stroke Detection

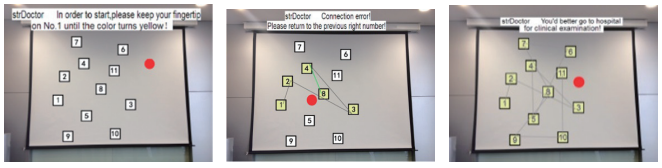
4.3.1 Dataset Description

We collected motion data from 14 patients (8 males and 6 females) and 55 healthy subjects (30 males and 25 females) aged from 30 to 68. Each subject corresponds to one sample. Thus, there are 69 sample instances in stroke detection dataset. All patients suffered from stroke for the first time. They have recovered well and were discharged from the hospital. Each subject was asked to play the specifically designed body sensing game BSG-TMT [61], which is developed based on the traditional TMT to indicate possible cognitive impairment.

4.3.2 Data Collection

In this experiment, Kinect is used for single fingertip tracking of any palm pose. We adopt the palm-pose-adaptive single fingertip tracking method [61] to detect fingertips in natural palm poses. In TMT, the trajectory coordinates and time are recorded in a text file.

The interfaces of BSG-TMT are shown in Figure 7. Subjects should connect a set of 11 dots in order as quickly as possible while maintaining accuracy. If the subject does not connect the dots in order, the game will remind the subject to correct the error before moving on to the next dot.



(a) Pre-start (b) Connecting error (c) Final state

Fig. 7. The user interface of BSG-TMT [62].

4.3.3 Feature Extraction

After data collection, we analyzed the BSG-TMT data and extracted eight features. The drawing length between two numbers is defined as l, the straight-line distance between two consecutively numbered dots is defined as L, and the time needed to connect two consecutively numbered dots is defined as T. The eight extracted features [21] are the total time required to complete the test (ATime), the trail accuracy (TAccuracy), the time the subject took to correct an error (CTime), the mean (M-R-Length), and variance (V-Length), of the $N-1$ ratios of l to L, the mean time duration spent by fingertips at the dots (M-Fingertip), the mean (M-R-Time), and variance (V-Time), of the ratios of T to l.

4.3.4 Feature Selection

In this experiment, we try to find the most distinguishing features for detection of stroke and discard the redundant features. The training rate is set as 70% and the number of sampling is set as 1,000, which are the same as the experimental settings in Section 4.2.4. Thus, the size of training set is $N_r = \lfloor 14 \times 70\% \rfloor + \lfloor 55 \times 70\% \rfloor = 47$. To find out the most discriminant features, we decrease the number of features from 8 to 1.

As shown in Figure 8, with the number of features declining from 8 to 1, the average classification accuracy first rises from 73.99% -75.04% to 74.88% -77.11% and then falls down to 64.44% -66.95%, which means some redundant features are abandoned at the first iteration and some useful information are dropped at the next two iterations. Thus, four significant features are selected by IRFFS-O. With B set as 30, the average classification accuracy reaches the maximum value 77.11%. The details of the four selected features are listed in Table 6. With this experimental result, we can build a more effective stroke detection model.

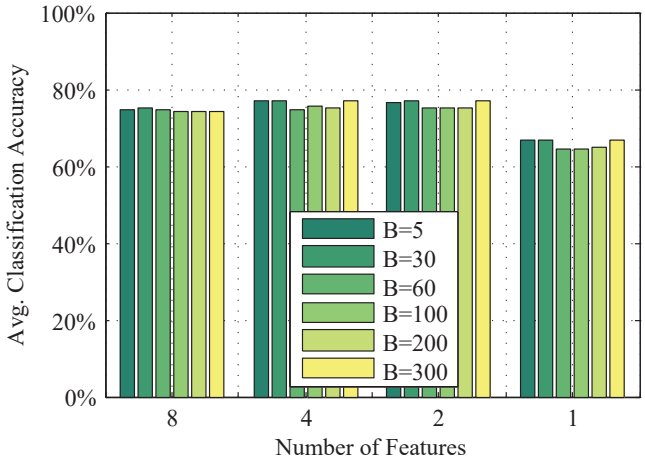


Fig. 8. The average classification accuracy of the random forest model at each iteration using the IRFFS-O method on the stroke dataset.

To further evaluate the effectiveness of the IRFFS-O method, we compare it with IRFFS and the other 6 feature selection methods presented in WEKA [19]. AdaBoostM1 and MultilayerPerceptron are both used as classifiers. The experimental settings are the same as in SVD dataset. As show in Table 7, the stroke detection accuracy of both AdaBoostM1 and MultilayerPerceptron based on the four features selected by the IRFFS-O method

TABLE 6
The Four Most Discriminant Features for Stroke Detection (Selected by the IRFFS-O method with $B = 30$)

Feature Name	Feature Calculation
ATime	$T_a = \sum_{i=1}^{N} T_i$, N is the number of lines drawn in TMT.
CTime	$T_C = \frac{1}{N_C} \sum_{i=1}^{N_C} T_i$, N_C is the number of connecting error.
M-R-Time	$T_{MR} = \frac{1}{N} \sum_{i=1}^{N} \frac{T_i}{l_i}$, t_i is the time spend on drawing the ith line, l_i is the drawing corresponding to the ith line.
M-Fingertip	$F_M = \sum_{i=1}^{N_p} I(\|pos - p_i\| < \mu)$, I is an indicator function, N_p is the number of points, pos is the position of tester finger, p_i is the position of points i and μ is a threshold value.

are better than the other seven feature selection methods, which further demonstrates that the IRFFS-O method is very effective in selecting the most significant features for accurate stroke detection and also further validates the effectiveness of IRFFS-O in handling imbalanced small-sampling data.

TABLE 7
The Stroke Detection Performance of AdaBoostM1 and MultilayerPerceptron based on the Four Features Selected by Different Feature Selection Methods

Feature Selection Method	AdaBoostM1	MultilayerPerceptron
CorrelationAttributeEval	82.43%	83.56%
GainRatioAttributeEval	78.24%	76.35%
InfoGainAttributeEval	78.24%	76.35%
OneRAttributeEval	80.40%	78.84%
ReliefFAttributeEval	82.43%	83.56%
SymmetricalUncertAttributeEval	78.24%	76.35%
IRFFS	78.17%	83.01%
IRFFS-O	**82.88%**	**84.06%**

4.3.5 Classification

With the four selected features, we build a stroke detection model with b-WELM. The RRSS technique is used to generate training and testing datasets. b-COELM, weighted ELM, weighted SVM and BP are used for comparisons.

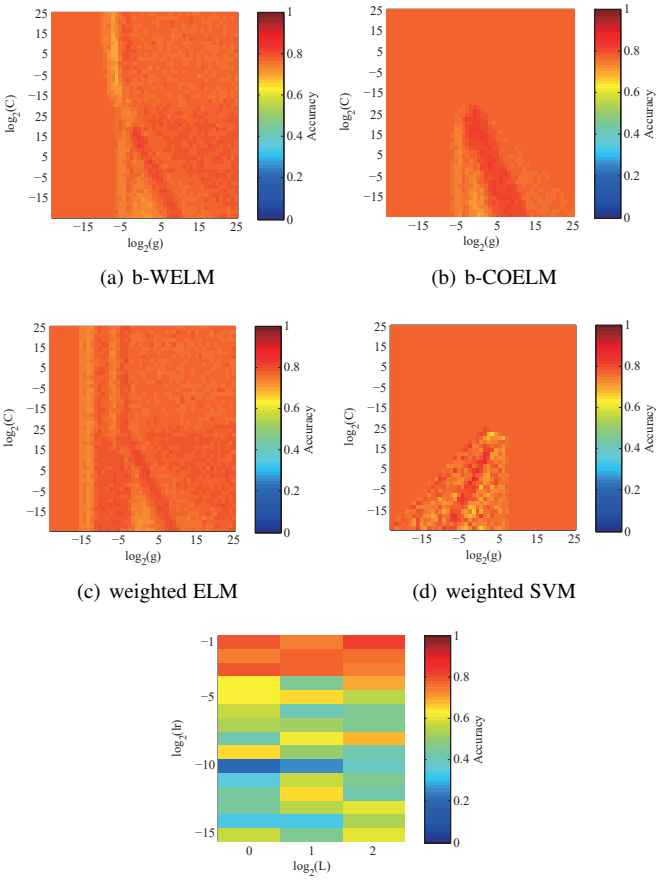
Parameter Selection Similar to Section 4.2.5, we adopt the Gaussian kernel $(k(\boldsymbol{x}_i, \boldsymbol{x}_j) = exp(-g\|\boldsymbol{x}_i - \boldsymbol{x}_j\|^2))$ for all comparison approaches except BP. The penalty parameter C and the kernel parameter g are tuned for b-WELM, b-COELM, weighted ELM and weighted SVM, while the number of hidden nodes L and the learning rate lr are tuned for BP. We adopt the grid search technique to select the optimal parameters for b-WELM, b-COELM, weighted ELM, weighted SVM and BP. Fifty different values $\{2^{-24}, 2^{-23}, \cdots, 2^{24}, 2^{25}\}$ are tested for both C and g, and 15 different values $\{2^{-15}, 2^{-14}, \cdots, 2^{-1}\}$ are tested for lr. We use 70% of the samples for training and 30% of the samples for testing. The weights in b-WELM, weighted ELM and weighted SVM are set to $1/\#(t_i)$, where $\#(t_i)$ is the number of samples belonging to class $t_i (i = 1, \cdots, m)$.

Figure 9 presents the accuracy achieved by b-WELM, b-COELM, weighted ELM, weighted SVM and BP with different parameter settings. The optimal parameter values for each model are shown in Table 8.

TABLE 8
The optimal parameter for each model

	b-WELM	b-COELM	weighted ELM	weighted SVM	BP
C (L)	2^{24}	2^{-1}	2^{8}	2^{9}	2^{2}
g (lr)	2^{3}	2^{-3}	2^{-2}	2^{-5}	2^{-1}

Classification Performance To evaluate the performance of b-WELM on stroke detection, we also adopt accuracy, sensitivity, specificity, and G-mean as evaluation metrics. Experiments using the optimal parameters obtained above in Table 8 are repeated for 1,000 times. Table 9 shows the average stroke detection performance of b-WELM, b-COELM, weighted ELM, weighted SVM and BP based on the features selected by the IRFFS-O method with $B = 30$.



(a) b-WELM (b) b-COELM (c) weighted ELM (d) weighted SVM (e) BP

Fig. 9. Experimental results on the accuracy of b-WELM, b-COELM, weighted ELM, weighted SVM and BP with different parameter pairs.

TABLE 9
The Stroke Detection Performance

	Test Accuracy	Test Sensitivity	Test Specificity	Test G-mean
b-WELM	**81.77%**	73.08%	84.33%	**77.52%**
b-COELM	81.73%	51.90%	90.50%	65.75%
weighted ELM	79.98%	**73.24%**	81.96%	76.41%
Weighted SVM	81.09%	64.74%	85.89%	73.15%
BP	79.37%	37.92%	**91.56%**	53.36%

As shown in Table 9, the accuracy of each classifier is above 79%, which demonstrates that the features selected by IRFFS-O are significant in distinguishing stroke patients from healthy persons. In this experiment, five methods show comparable performance in test accuracy, while presenting significant differences in test sensitivity, test specificity, and test G-mean. For sensitivity, b-WELM and weighted ELM perform similarly, which are both about 73%. By contrast, the other three methods, especially BP, show poor performance in sensitivity. However, these three methods show relatively better performances in specificity, which means they perform better in classifying the healthy subjects. For G-mean, b-WELM shows the best performance, which indicates

that b-WELM achieves the best overall performance in the detection of stroke. For imbalanced datasets, b-WELM can classify the patients more accurately, which is very useful for diagnosing stroke.

Based on the experimental results shown above, it can be seen that TMT is useful for the detection of stroke. The connect-the-dots behavior can be analyzed for the indication of stroke. MOCA is an effective tool for the doctors to make analysis of the correlation between motor patterns and stroke.

5 Conclusions

In this paper, we propose the MOCA framework to inferring cognitive wellness from motor patterns. We propose a novel feature selection method – IRFFS-O – and a novel classification method – b-WELM – under the MOCA framework which can be used to deal with imbalanced small-sampling data prevalent among disease datasets. Various sensors can be used for data collection in conjunction with our framework. Motor pattern features are extracted according to the collected data. Experimental results on two real-world datasets, including SVD and stroke patients' data, demonstrate that MOCA is suitable for the analysis of correlation between motor patterns and cognitive wellness, where the datasets are usually imbalanced small-sampling. It is able to reduce the rate of missed diagnosis by appropriately improving the sensitivity of the algorithm. Therefore, MOCA can be employed as an effective tool to assist doctors in the diagnosis of cognitive disorders.

In the future, we will conduct experiments on more cognitive disorder related datasets to study the effectiveness of this framework. Besides, we will investigate how to jointly consider the feature selection and extraction steps, possibly with the help of deep learning neural networks.

Acknowledgments

This work is supported by the National Key Research and Development Program of China (No. 2017YFB1002801); Natural Science Foundation of China under Grant No. 61572471 and No. 61502456; Science and Technology Planning Project of Guangdong Province under Grant No. 2015B010105001; the National Research Foundation, Prime Minister's Office, Singapore under its IDM Futures Funding Initiative; the Lee Kuan Yew Post-Doctoral Fellowship Grant; and the Singapore Ministry of Health under its National Innovation Challenge on Active and Confident Ageing (NIC Project No. MOH/NIC/COG04/2017). The corresponding author is Yiqiang Chen.

References

[1] C. M. Roebers, M. Röthlisberger, R. Neuenschwander, P. Cimeli, E. Michel, and K. Jäger, "The relation between cognitive and motor performance and their relevance for children's transition to school: a latent variable approach," *Human Movement Science*, vol. 33, pp. 284–297, 2014.

[2] V. J. Verlinden, J. N. van der Geest, A. Hofman, and M. A. Ikram, "Cognition and gait show a distinct pattern of association in the general population," *Alzheimer's & Dementia*, vol. 10, no. 3, pp. 328–335, 2014.

[3] I. M. van der Fels, S. C. te Wierike, E. Hartman, M. T. Elferink-Gemser, J. Smith, and C. Visscher, "The relationship between motor skills and cognitive skills in 4–16 year old typically developing children: A systematic review," *Journal of Science and Medicine in Sport*, vol. 18, no. 6, pp. 697–703, 2015.

[4] D. Son, J. Lee, S. Qiao, R. Ghaffari, J. Kim, J. E. Lee, C. Song, S. J. Kim, D. J. Lee, S. W. Jun, S. Yang, M. Park, J. Shin, K. Do, M. Lee, K. Kang, C. S. Hwang, N. Lu, T. Hyeon, and D.-H. Kim, "Multifunctional wearable devices for diagnosis and therapy of movement disorders," *Nature Nanotechnology*, vol. 9, pp. 397–404, 2014.

[5] N. Savage, "Mobile data: Made to measure," *Nature*, vol. 527, pp. S12–S13, 2015.

[6] L. Gravitz, "Technology: Monitoring gets personal," *Nature*, vol. 538, pp. S8–S10, 2017.

[7] H. Yu, C. Miao, C. Leung, Y. Chen, S. Fauvel, V. R. Lesser, and Q. Yang, "Mitigating herding in hierarchical crowdsourcing networks," *Scientific Reports*, vol. 6, no. 4, pp. doi:10.1038/s41 598–016–0011–6, 2016.

[8] H. Yu, C. Miao, Y. Chen, S. Fauvel, X. Li, and V. R. Lesser, "Algorithmic management for improving collective productivity in crowdsourcing," *Scientific Reports*, vol. 7, no. 12541, pp. doi:10.1038/s41 598–017–12 757–x, 2017.

[9] S. Ahmed, I. Baker, S. Thompson, M. Husain, and C. R. Butler, "Utility of testing for apraxia and associated features in dementia," *Journal of Neurology, Neurosurgery & Psychiatry*, pp. doi:10.1136/jnnp–2015–312 945, 2016.

[10] K. F. de Laat, A. G. van Norden, R. A. Gons, L. J. van Oudheusden, I. W. van Uden, D. G. Norris, M. P. Zwiers, and F.-E. de Leeuw, "Diffusion tensor imaging and gait in elderly persons with cerebral small vessel disease," *Stroke*, vol. 42, pp. 373–379, 2011.

[11] A. Soumaré, A. Elbaz, Y. Zhu, P. Maillard, F. Crivello, B. Tavernier, C. Dufouil, B. Mazoyer, and C. Tzourio, "White matter lesions volume and motor performances in the elderly," *Annals of Neurology*, vol. 65, no. 6, pp. 706–715, 2009.

[12] G. Prasad, P. Herman, D. Coyle, S. McDonough, and J. Crosbie, "Applying a brain-computer interface to support motor imagery practice in people with stroke for upper limb recovery: a feasibility study," *Journal of NeuroEngineering and Rehabilitation*, vol. 7, no. 60, pp. doi:10.1186/1743–0003–7–60, 2010.

[13] P. Boyle, R. Cohen, R. Paul, D. Moser, and N. Gordon, "Cognitive and motor impairments predict functional declines in patients with vascular dementia," *International Journal of Geriatric Psychiatry*, vol. 17, no. 2, pp. 164–169, 2002.

[14] P. Foster, V. Drago, B. Ferguson, P. Harrison, and D. Harrison, "Quantitative electroencephalographic and neuropsychological investigation of an alternative measure of frontal lobe executive functions: the figure trail making test," *Brain Informatics*, vol. 2, no. 4, pp. 239–251, 2015.

[15] M. Termenon, M. Graña, A. Savio, A. Akusok, Y. Miche, K.-M. Björk, and A. Lendasse, "Brain mri morphological patterns extraction tool based on extreme learning machine and majority vote classification," *Neurocomputing*, vol. 174, pp. 344–351, 2016.

[16] J. L. Arbuckle, Ed., *IBM SPSS AMOS 19 user's guide*. Crawfordville, FL: AMOS Development Corporation, 2010.

[17] J. Lehmann, C. Quaiser-Pohl, and P. Jansen, "Correlation of motor skill, mental rotation, and working memory in 3-to 6-year-old children," *European Journal of Developmental Psychology*, vol. 11, no. 5, pp. 560–573, 2014.

[18] Y. Chen, M. Huang, C. Hu, Y. Zhu, F. Han, and C. Miao, "A coarse-to-fine feature selection method for accurate detection of cerebral small vessel disease," in *Proceedings of 2016 International Joint Conference on Neural Networks (IJCNN'16)*, 2016, pp. 2609–2616.

[19] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, "The weka data mining software: An update," *ACM SIGKDD Explorations Newsletter*, vol. 11, no. 1, pp. 10–18, 2009.

[20] G. Huang, Q. Zhu, and C. Siew, "Extreme learning machine: a new learning scheme of feedforward neural networks," in *Proceedings of 2004 International Joint Conference on Neural Networks (IJCNN'04)*, 2004, p. doi:10.1109/IJCNN.2004.1380068.

[21] Y. Chen, H. Yu, C. Miao, B. Chen, X. Yang, , and C. Leung, "Using motor patterns for stroke detection," *Science*, vol. Advances in Computational Psychophysiology, pp. 12–15, 2015.

[22] M. Graña, M. Termenon, A. Savio, A. Gonzalez-Pinto, J. Echeveste, J. Pérez, and A. Besga, "Computer aided diagnosis system for Alzheimer disease using brain diffusion tensor imaging features selected by Pearson's correlation," *Neuroscience Newsletter*, vol. 502, no. 3, pp. 225–229, 2011.

[23] A. Savio, M. García-Sebastián, D. Chyzyk, C. Hernandez, M. Graña, A. Sistiaga, A. López de Munain, and J. Villanúa, "Neurocognitive disorder detection based on feature vectors extracted from VBM analysis of structural MRI," *Computers in Biology and Medicine*, vol. 41, no. 8, pp. 600–610, 2011.

[24] C. Goetz, B. Tilley, S. Shaftman, G. Stebbins, S. Fahn, P. Martinez-Martin, W. Poewe, C. Sampaio, M. Stern, R. Dodel, B. Dubois, R. Hol-

loway, J. Jankovic, J. Kulisevsky, A. Lang, A. Lees, S. Leurgans, P. LeWitt, D. Nyenhuis, C. Olanow, O. Rascol, A. Schrag, J. Teresi, J. van Hilten, and N. LaPelle, "Movement Disorder Society-sponsored revision of the Unified Parkinson's Disease Rating Scale (MDS-UPDRS): scale presentation and clinimetric testing results," *Movement Disorders*, vol. 23, no. 15, pp. 2129–2170, 2008.

[25] D. Jarchi, A. Peters, B. Lo, E. Kalliolia, I. D. Giulio, P. Limousin, B. L. Day, and G.-Z. Yang, "Assessment of the e-AR sensor for gait analysis of Parkinson's Disease patients," in *Proceedings of the 12th IEEE International Conference on Wearable and Implantable Body Sensor Networks (BSN'15)*, 2015, pp. 1–6.

[26] A. Taylor Tavares, G. Jefferis, M. Koop, B. Hill, T. Hastie, G. Heit, and H. Bronte-Stewart, "Quantitative measurements of alternating finger tapping in Parkinson's disease correlate with UPDRS motor disability and reveal the improvement in fine motor control from medication and deep brain stimulation," *Computers in Biology and Medicine*, vol. 20, no. 10, pp. 1286–1298, 2005.

[27] A. Kita, P. Lorenzi, G. Romano, R. Rao, R. Parisi, A. Suppa, M. Bologna, A. Berardelli, and F. Irrera, "Smart sensing systems for the detection of human motion disorders," *Procedia Engineering*, vol. 120, pp. 324–327, 2015.

[28] A. Larab, R. Bastide, and B. Rigaud, "Using sensors and labeled graphs to detect the space confusion's problem of patients suffering from Alzheimer's disease," in *Proceedings of the 2009 ICSE Workshop on Software Engineering in Health Care (SEHC'09)*, 2009, pp. 29–33.

[29] Y. Yang, Z. Zha, Y. Gao, X. Zhu, and T. Chua, "Exploiting web images for semantic video indexing via robust sample-specific loss," *IEEE Transactions on Multimedia*, vol. 16, no. 6, pp. 1677–1689, 2014.

[30] H. Yu, Z. Shen, C. Miao, C. Leung, Y. Chen, S. Fauvel, J. Lin, L. Cui, Z. Pan, and Q. Yang, "A dataset of human decision-making in teamwork management," *Scientific Data*, vol. 4, no. 160127, p. doi:10.1038/sdata.2016.127, 2017.

[31] H. Liu, B. Liu, H. Zhang, X. Q. Liang Li, and G. Zhang, "Crowd evacuation simulation approach based on navigation knowledge and two-layer control mechanism," *Information Sciences*, vol. 436-437, pp. 247 – 267, 2018.

[32] C. Hu, H. Liu, and P. Zhang, "Cooperative co-evolutionary artificial bee colony algorithm based on hierarchical communication model," *Chinese Journal of Electronics*, vol. 25, no. 3, pp. 570–576, 2016.

[33] J. Shi, S. Zhou, X. Liu, Q. Zhang, M. Lu, and T. Wang, "Stacked deep polynomial network based representation learning for tumor classification with small ultrasound image dataset," *Neurocomputing*, vol. 194, pp. 87–94, 2016.

[34] R. Bharti, J. Akanksha, S. Mohit, G. Sunita, K. S. Senthil, A. RK, and B. Madhuri, "Regions-of-interest based automated diagnosis of parkinson's disease using t1-weighted mri," *Expert Systems with Applications*, vol. 42, no. 9, pp. 4506–4516, 2015.

[35] M. S. Tan, J. W. Tan, S.-W. Chang, H. J. Yap, S. A. Kareem, and R. B. Zain, "A genetic programming approach to oral cancer prognosis," *PeerJ*, vol. 4, p. e2482, 2016.

[36] X. Zhu, Z. Huang, Y. Yang, H. Tao Shen, C. Xu, and J. Luo, "Self-taught dimensionality reduction on the high-dimensional small-sized data," *Pattern Recognition*, vol. 46, no. 1, pp. 215–229, 2013.

[37] P. J. M. Genuer Robin and T.-M. Christine, "Variable selection using random forests," *Pattern Recognition Letters*, vol. 31, no. 14, pp. 2225–2236, 2010.

[38] C. David and C. Nitesh, "Learning decision trees for unbalanced data," *Machine Learning and Knowledge Discovery in Databases*, pp. 241–256, 2008.

[39] B. Rok and L. Lara, "Evaluation of smote for high-dimensional class-imbalanced microarray data," in *Proceedings of the 11th International Conference on Machine Learning and Applications (ICMLA'12)*, vol. 2. IEEE, 2012, pp. 89–94.

[40] M. M. Rahman and D. Davis, "Addressing the class imbalance problem in medical datasets," *International Journal of Machine Learning and Computing*, vol. 3, no. 2, p. 224, 2013.

[41] R. Niehaus, D. S. Raicu, J. Furst, and S. Armato, "Toward understanding the size dependence of shape features for predicting spiculation in lung nodules for computer-aided diagnosis," *Journal of Digital Imaging*, vol. 28, no. 6, pp. 704–717, 2015.

[42] F. Ren, P. Cao, W. Li, D. Zhao, and O. Zaiane, "Ensemble based adaptive over-sampling method for imbalanced data learning in computer aided detection of microaneurysm," *Computerized Medical Imaging and Graphics*, vol. 55, pp. 54–67, 2017.

[43] Y. Huang and S. Du, "Weighted support vector machine for classification with uneven training class sizes," in *Proceedings of 2005 International Conference on Machine Learning and Cybernetics (ICMLC'05)*, vol. 7. IEEE, 2005, pp. 4365–4369.

[44] W. Zong, G. Huang, and Y. Chen, "Weighted extreme learning machine for imbalance learning," *Neurocomputing*, vol. 101, pp. 229–242, 2013.

[45] T. Schmitz-Hübsch, S. du Montcel, L. Baliko, J. Berciano, S. Boesch, C. Depondt, P. Giunti, C. Globas, J. Infante, J. Kang, B. Kremer, C. Mariotti, B. Melegh, M. Pandolfo, M. Rakowicz, P. Ribai, R. Rola, L. Schöls, S. Szymanski, B. van de Warrenburg, A. Dürr, T. Klockgether, and R. Fancellu, "Scale for the assessment and rating of ataxia," *Neurology*, vol. 66, no. 11, pp. 1717–1720, 2006.

[46] J. Guralnik, E. Simonsick, L. Ferrucci, R. Glynn, L. Berkman, D. Blazer, P. Scherr, and R. Wallace, "A short physical performance battery assessing lower extremity function: association with self-reported disability and prediction of mortality and nursing home admission," *Journal of Gerontology*, vol. 49, no. 2, pp. 85–94, 1994.

[47] R. M. Reitan, "The relation of the trail making test to organic brain damage," *Journal of Consulting Psychology*, vol. 19, no. 5, p. 393, 1955.

[48] ——, "Validity of the trail making test as an indicator of organic brain damage," *Perceptual and Motor Skills*, vol. 8, no. 3, pp. 271–276, 1958.

[49] B. Wiberg, "Risk factors for stroke in adult men," Ph.D. dissertation, Faculty of Medicine, Uppsala University, Sweden, 2010.

[50] D. Ping Tian *et al.*, "A review on image feature extraction and representation techniques," *International Journal of Multimedia and Ubiquitous Engineering*, vol. 8, no. 4, pp. 385–396, 2013.

[51] L. Breiman, "Random forests," *Machine Learning*, vol. 45, pp. 5–32, 2001.

[52] L. Breiman, J. Friedman, C. J. Stone, and R. Olshen, *Classification and Regression Trees*, 1st ed. 2000 N.W. Blvd, Boca Raton, Florida 33431: CRC Press LLC, 1998.

[53] L. Hu, Y. Chen, S. Wang, and Z. Chen, "b-COELM: A fast, lightweight and accurate activity recognition model for mini-wearable devices," *Pervasive and Mobile Computing*, vol. 15, pp. 200–214, 2014.

[54] G. Huang, H. Zhou, X. Ding, and R. Zhang, "Extreme learning machine for regression and multiclass classification," *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, vol. 42, no. 2, pp. 513–529, 2012.

[55] K. Han, D. Yu, and I. Tashev, "Speech emotion recognition using deep neural network and extreme learning machine," in *Proceedings of the 15th Annual Conference of the International Speech Communication Association (ISCA'14)*, 2014, pp. 14–18.

[56] O. L. Mangasarian and E. W. Wild, "Proximal support vector machine classifiers," in *Proceedings of the 7th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'01)*, 2001, pp. 77–86.

[57] X. Yang, Q. Song, and Y. Wang, "A weighted support vector machine for data classification," *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 21, no. 05, pp. 961–976, 2007.

[58] D. E. Rumelhart, J. L. McClelland, and C. PDP Research Group, Eds., *Parallel Distributed Processing: Explorations in the Microstructure of Cognition, Vol. 1: Foundations*. Cambridge, MA, USA: MIT Press, 1986.

[59] J. Zijlmans, P. Poels, J. Duysens, J. van der Straaten, T. Thien, M. van't Hof, H. Thijssen, and M. Horstink, "Quantitative gait analysis in patients with vascular Parkinsonism," *Movement Disorders*, vol. 11, no. 5, pp. 501–508, 1996.

[60] J. Heaton, *Introduction to Neural Networks for Java, 2Nd Edition*, 2nd ed. Heaton Research, Inc., 2008.

[61] H. Yu, X. Yang, Y. Chen, and J. Liu, "strDoctor: Indicate stroke for elderly through body sensing game," in *Proceedings of the 2015 IEEE 12th International Conference on Ubiquitous Intelligence and Computing and 2015 IEEE 12th International Conference on Autonomic and Trusted Computing and 2015 IEEE 15th International Conference on Scalable Computing and Communications and Its Associated Workshops (UIC-ATC-ScalCom'15)*, 2015, pp. 360–363.

[62] L. Rokach and O. Maimon, "Top-down induction of decision trees classifiers - a survey," *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 35, no. 4, pp. 476–487, 2005.

Yiqiang Chen received the B.S. and M.S. degrees in computer science from Xiangtan University, Xiangtan, China, in 1996 and 1999, respectively, and the Ph.D. degree in computer science from the Institute of Computing Technology, Chinese Academy of Sciences, Beijing, China, in 2003.

In 2004, he was a Visiting Scholar Researcher with the Department of Computer Science, Hong Kong University of Science and Technology (HKUST), Hong Kong. He is currently a professor and the director of the Pervasive Computing Research Center at the Institute of Computing Technology (ICT), Chinese Academy of Sciences (CAS). His research interests include artificial intelligence, pervasive computing, and human computer interaction.

Chunyu Hu received the B.S. and M.S. degrees in computer science and technology from Shandong Normal University, Jinan, China, in 2012 and 2015, respectively. Now she is a Ph.D. candidate in the Institute of Computing Technology, Chinese Academy of Sciences, Beijing, China. Her current research interests include machine learning, pervasive computing, and activity recognition.

Bin Hu received Ph.D. degree in computer science from Institute of Computing Technology, Chinese Academy of Science, China in 1998. Since 2008, he has been a professor and the Dean of School of Information Science and Engineering, Lanzhou University, China. He had also been a guest professor in ETH Zurich, Switzerland till 2011. His research interests include Pervasive Computing, Computational Psychophysiology, and Data Modeling.

Lisha Hu received the B.S. and M.S. degrees in applied mathematics from Hebei University, Baoding, China, in 2009 and 2012, respectively, and the Ph.D. degree in computer science from the Institute of Computing Technology, Chinese Academy of Sciences, Beijing, China, in 2017. She is currently a lecturer in the Institute of Information Technology, Hebei University of Economics and Business, Shijiazhuang, China. Her research interests include machine learning, wearable computing, and activity recognition.

Han Yu received his B.Eng. (Hons) degree and Ph.D. degree from the School of Computer Engineering, Nanyang Technological University (NTU), Singapore in 2007 and 2014, respectively. He is currently a Lee Kuan Yew Post-Doctoral Fellow (LKY PDF) at the Joint NTU-UBC Research Centre of Excellence in Active Living for the Elderly (LILY). His research focus is on artificial intelligence (AI)-powered crowd work. Specifically, he designs stochastic optimization techniques to enhance crowd workers' experience through data-driven algorithmic management.

Chunyan Miao received the B.S. degree from Shandong University, Jinan, China, in 1988, and the M.S. and Ph.D. degree in Nanyang Technological University, Singapore, in 1998 and 2003, respectively. She is currently a Professor in the School of Computer Science and Engineering at Nanyang Technological University (NTU), and the Director of the Joint NTU-UBC Research Centre of Excellence in Active Living for the Elderly (LILY). Her research focus on infusing intelligent agents into interactive new media (virtual, mixed, mobile and pervasive media) to create novel experiences and dimensions in game design, interactive narrative and other real world agent systems.