

Received June 29, 2021, accepted July 19, 2021, date of publication July 27, 2021, date of current version August 6, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3100580

A Novel Sensor-Based Human Activity Recognition Method Based on Hybrid Feature Selection and Combinational Optimization

YIMING TIAN¹, JIE ZHANG², (Senior Member, IEEE), LIPENG LI¹, AND ZUOJUN LIU³

¹College of Information Engineering, Tianjin University of Commerce, Tianjin 300134, China

²School of Engineering, Merz Court, Newcastle University, Newcastle upon Tyne NE1 7RU, U.K.

³School of Artificial Intelligence and Data Science, Hebei University of Technology, Tianjin 300130, China

Corresponding author: Yiming Tian (yimingtian@tjcu.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 61803143 and Grant 61703135, and in part by the National Key Technology Research and Development Program of the Ministry of Science and Technology of China under Grant 2015BAI06B03.

ABSTRACT In recent years, sensor-based human activity recognition (HAR) has become a hot topic due to the advancement of sensing technologies, wireless communication technologies and nano-technologies. Since the sensor signals are usually non-stationary and quite noisy, both selecting the discriminant feature representations and finding out the optimal parameters for recognition algorithm play an important role for the enhanced performance and robustness of an HAR system. However, most of the previous research focused on one of them ignoring their interactions. Very few studies focused on these two aspects simultaneously. Considering the two factors separately may lead to inferior HAR performance. This paper presents a novel HAR framework which can optimize the feature set and the parameters of recognition algorithm synchronously for robust and optimal system performance. A new hybrid feature selection methodology using game-theory based feature selection (GTFS) and binary firefly algorithm (BFA), called GTFS-BFA, is proposed. GTFS-BFA is a hybrid methodology combining evidence from both filter and wrapper feature selection methods. It consists of two phases, namely pre-selection phase and re-selection phase. Pre-selection phase relies on game-theory-based filter method, while the re-selection phase uses binary firefly algorithm (BFA) as a wrapper method. The popular and efficient algorithm kernel extreme learning machine (KELM) is utilized as a classifier. The experimental results indicate that the proposed method can obtain better comprehensive performance in terms of four performance measures through a comparison to other existing methods on daily activity dataset from five body positions.

INDEX TERMS Human activity recognition, hybrid feature selection, combinational optimization, wearable sensor, binary firefly algorithm.

I. INTRODUCTION

In recent years, as the development of inertial measurement unit (IMU) sensors and wireless transmission technology, human activity recognition (HAR) has become a promising research area in academic and application fields. HAR technology is an effective way to achieve better information interaction between humans and the external environment. It can determine the type of activity and show it by displaying

The associate editor coordinating the review of this manuscript and approving it for publication was Xi Peng.

text or animation, which can help us to make efficient decisions for future human activities. The vision-based HAR systems often suffer from privacy and insufficient illumination [1]. Besides, the vision-based HAR systems can only monitor users in a specific area, which greatly limits their actual use. The wearable sensor-based system is another ideal choice for HAR, which has the advantages of being light and compact. The gap between low-level sensor data and high-level meaningful applications can be reduced by sensor-based HAR system. For example, sensor-based HAR systems have been widely adopted for elderly care fields [2],

such as fall detection and assisted living. Besides, for health-care services, sensor-based HAR system can provide effective information for health conditions of human beings and some diseases, such as Parkinson's and stroke [3], are related to the mobility of the human body. Doctors can detect and assess the rehabilitation of some diseases by using the information from HAR. In addition to these, sensor-based HAR plays an important role in physical training, such as swimming [4] and badminton training [5].

Previous works on sensor-based HAR can be roughly divided into two categories according to the sensor hardware systems. The first one is the body sensor network-based approach [6] and this technique combines sensor information from different positions of the body for HAR. However, excessive sensors on the body will cause inconvenience to the person's daily life, especially in long-term monitoring. Moreover, this approach requires the extra cost of the equipment. The other HAR approach is based on a single sensor [7], [8], which is typically an accelerometer or other inertial sensor mounted at the waist or other positions to recognize activities. Compared with body sensor network-based approach, this approach is low-cost and less intrusive. Therefore, more and more researches have been focused on single sensor-based HAR systems.

Various features have been utilized in sensor-based HAR and these include the traditional time and frequency domain features such as mean, median, standard, Fast Fourier transform, spectral energy, etc. These features are effective for linear signals but sometimes are not helpful because of the non-stationary of the activity data. Therefore, time-frequency domain-based features such as Hilbert-Huang features [9] and EEMD-based features [10] have been proposed as important feature vectors in HAR. However, if all the extracted features are employed for HAR, there will be some redundant and irrelevant features, which could increase the computational cost and reduce the performance of the HAR system.

Selecting representative features which characterize different activities rather than adopting all features is very important in enhancing the HAR system performance. The current feature selection methods can be broadly categorized into filter and wrapper based methods [11]. The filter based methods use the statistical characteristics of the training data to evaluate the importance of features. Then several most important features are selected as a feature subset. This type of methods has the advantage of efficiency, but they are less effective because they ignore the influence of a classifier [12]. The wrapper based methods select the feature subset by utilizing the performance of a classifier. So, the wrapper based methods can achieve high recognition performance for the particular classifier. However, they have high computational cost and are time-consuming [13]. As mentioned above, both filter and wrapper based methods have their advantages and disadvantages. Hence, a hybrid method combining the advantages of filter and wrapper based methods could improve the performance of feature selection.

Recently, various efficient filter based methods have been proposed and have been widely utilized in dealing with high-dimensional feature selection problems. However, most of these methods tend to ignore the relationship between features. These methods only focus on the features that have strong discriminatory power as individuals and neglect the ones that have strong discriminatory power as a group. Recently a novel filter-based method inspired from cooperative game theory has emerged in feature selection and is named game-theory based feature selection [14]. Its powerful feature selection ability has been demonstrated in many real-world applications [10], [15]. Accordingly, we can first apply GTFS to select some features with the highest weights as candidates for the wrapper program. In this way, the raw feature data with high dimensionality can be pre-optimized into low-dimensional feature dataset, which can lower the computational burden of the wrapper process. However, the classifier in the wrapper based methods has an important influence on the selected features and the performance of HAR. Therefore, constructing an effective classifier and re-selecting features in the wrapper process is crucial for the HAR system performance.

Kernel extreme learning machine (KELM) is a popular machine learning method, which has comparable performance to ELM [16]. It has many unique advantages such as extremely fast learning speed and good generalization performance. Hence KELM is very suitable for HAR. However, the establishment of a KELM classifier requires proper selection of the kernel function, the kernel parameters, and the soft margin constant. Appropriate parameter selection is important for improving KELM performance. There have been many studies on the optimization of classifiers using optimization methods such as genetic algorithm and particle swarm optimization [17]–[19]. However, the optimization of features and setting the classifier parameter are usually done separately in the wrapper phase. To obtain the optimal recognition performance of HAR, both feature subset and KELM parameters must be optimized simultaneously. Performing these two aspects separately in the wrapper phase may not lead to optimal recognition performance.

Increasingly evolutionary algorithms and swarm intelligence algorithms have been employed to solve the feature selection problem. Firefly algorithm (FA) was initially proposed by Yang [20] in 2009 and shows the superior optimization ability in various applications. Specifically, some studies have demonstrated the superiority of the FA against the genetic algorithm (GA) and particle swarm optimization (PSO) [21]–[22]. Therefore, a novel hybrid feature selection method named GTFS-BFA is proposed in this paper to optimize the HAR system performance. The proposed method includes combinational optimization of the classifier in the wrapper-based feature selection phase. Firstly, the GTFS is applied to select some important features beneficial to recognition. However, it is less effective without considering the influence of a classifier. Hence, in the next wrapper phase, these features are used as candidates, which will be further

optimized by the BFA. Moreover, in the wrapper phase, the classifier is optimized to fit the selected feature set. The novelty and main contributions of this paper are highlighted as follows:

(1) Activity recognition framework: considering feature subset selection influences the appropriate classifier parameters and vice versa, this paper presents a novel HAR framework that can optimize the feature set and the parameters of the recognition algorithm synchronously. This helps optimize the feature selection and the parameters of the recognition algorithm effectively and achieve robust and optimal system performance. In addition, it is a general framework that can be applied to various classification tasks.

(2) Feature selection method: a novel hybrid feature selection method based on GTFS and BFA is proposed to improve the recognition accuracy and efficiency of recognition. While existing studies have applied the filter or wrapper method for feature selection in HAR, limited work has investigated the hybrid of both methods in the literature of HAR. Moreover, our proposed GTFS-BFA method is different from the previous studies in terms of its filter selection criteria as well as the search implementation in the wrapper process. The hybrid approach GTFS-BFA proposed in this work has a novel contribution to the literature of HAR.

(3) Experimental evaluation: we utilize the acceleration data from five body positions to comprehensively analyze the effectiveness of the proposed scheme and compare it with six well-established counterparts in the literature. The proposed scheme is shown to have a better performance against six well-established counterparts in terms of feature optimization and combinational optimization. This provides a clear indication that the proposed scheme can be a promising alternative for HAR.

The paper is organized as follows. Section II presents some previous HAR works on feature selection and classification algorithm. Section III presents the related algorithms including GTFS, BFA and KELM. Section IV gives the proposed HAR approach, details the feature extraction and hybrid feature selection, as well as illustrates the combinational optimization phase. In Section V, the experimental setup and performance measures are introduced. Section VI presents the experimental results on data from different positions. Some concluding remarks are drawn in Section VII.

II. RELATED WORKS

Some advanced works have been reported on HAR by optimizing the original high dimensional feature set. Feature transformation methods such as LDA [23] and KFDA [24] have been proposed to reduce the feature dimension while also enhancing the distinguishing ability of feature vectors. A game theory-based feature selection method is applied in HAR by Wang *et al.* [10]. The game-based theory is a mathematical method that

describes the phenomenon of conflict and cooperation based on entropy and mutual information theory. Experimental comparisons with the ReliefF and minimum-redundancy maximum-relevance verify the effectiveness of the proposed approach. Ghasemzadeh *et al.* [25] proposed a power-aware feature selection method for mobile-based HAR. To reduce computation complexity, integer programming and greedy approximation approaches are utilized in the method to optimize feature set. Experimental results on data collected from real subjects demonstrate the effectiveness of the proposed method. Considering the influence of the orientation, placement, and subject variations on HAR performance, a method based on coordinate transformation and principal component analysis (CT-PCA) was proposed in [26] to realize location-adaptive activity recognition. Wang *et al.* [27] proposed a hybrid feature selection method to reduce feature dimensions for HAR. This method combined the traditional feature selection methods based on filter and wrapper. However, only the data from the waist was utilized to verify the proposed method. Moreover, un-optimized classifiers may affect the performance of the HAR system.

Recently, a large number of classification algorithms have been utilized for HAR. The classification algorithms such as SVM [28], artificial neural network [29], k-mean clustering [30] and decision tree (DT) [31] are widely applied in HAR. Besides, new branches of machine learning, such as deep learning [32]–[34] and ensemble learning [35]–[37], have also shown their merits in HAR. However, the deep learning-based approaches require a huge dataset for model training, which may not be applicable in actual scenarios. Besides, the high computational load of deep learning makes it unsuitable for real-time human activity detection. The ensemble learning-based approaches can increase the robustness and accuracy of the recognition system. However, there are some weaknesses in establishing an ensemble learning-based recognition system, such as the difficulty of generating fully independent base classifiers and choosing the suitable base classifier.

III. PRELIMINARIES

A. GAME-THEORY BASED FEATURE SELECTION

Mutual information (MI) is used to indicate the degree of correlation between two random variables, which can measure both the linear relationship between variables and the nonlinear relationship between variables. The joint probability distribution of two random variables X and Y is $p(x, y)$ and the mutual information of these two random variables $I(X; Y)$ can be defined as follows:

$$I(X; Y) = \sum_{y \in Y} \sum_{x \in X} p(x, y) \log_2 \frac{p(x, y)}{p(x)p(y)} \quad (1)$$

The conditional mutual information is the MI of the two variables $\{X, Y\}$ given a discrete random variable Z , which

can be expressed as follows:

$$CMI(X; Y | Z) = \sum_{y \in Y} \sum_{x \in X} \sum_{z \in Z} p(x, y, z) \log_2 \frac{p(x, y | z)}{p(x | z)p(y | z)} \quad (2)$$

In the game theory, Shapley value is utilized to measure the powers of game players. In our method, we regard the feature as the player and the feature subset as the coalition. By introducing Shapley value to evaluate the weight of each feature, we can select each coalition as a candidate subset for the final best feature subset. The Shapley value can provide a fair and efficient way to estimate the features' importance corresponding to the contribution of the features, while considering their possible intrinsic and intricate correlative interactions. The relevance, redundancy and interdependence of features can be considered by Shapley value, which is an effective method to measure features. It is formulated as following:

$$\Psi_i(v) = \sum_{L \subset M \setminus \{i\}} \Delta_i(L) \frac{|L|!(m - |L| - 1)!}{m!} \quad (3)$$

and

$$\Delta_i(L) = \psi(L \cup i) - \psi(L)$$

where m represents the number of players and the sum extends over all subsets L of M not including player i .

In this paper, the weight of the feature f_i is expressed by the Shapley value and the function $\Delta_i(L)$ can be redefined associated with feature information:

$$\Delta_i(L) = \begin{cases} 1, & \text{if } CMI(L; C | f_i) \geq 0 \\ & \text{and } \sum_{f_j \in L} \Phi(i, j) \geq \frac{|L|}{2} \\ 0, & \text{otherwise} \end{cases} \quad (4)$$

This equation means that the Shapley value of the feature f_i is positive only when the feature is related to the coalition and interdependent with at least half of the L features. In the formula (4), $\Phi(i, j)$ indicates the interdependence between features, the interdependence between features means that each feature in the relationship cannot function when it is separated from one another, that is, the influence of each feature on the recognition performance cannot be ignored and replaced. Suppose both features f_i and f_j are in the feature set and interdependent, the correlation between the feature f_j and the target class C can be increased conditioned by f_i , that is, the two features f_i and f_j are interdependent if the following formula is satisfied.

$$CMI(f_j; C | f_i) \geq I(f_j; C) \quad (5)$$

where C means the instance class. Therefore, the $\Phi(i, j)$ can be defined as follows:

$$\Phi(i, j) = \begin{cases} 1, & \text{if } CMI(f_j; C | f_i) \geq I(f_j; C) \\ 0, & \text{otherwise} \end{cases} \quad (6)$$

Formula (6) can guarantee the selected features are relevant to the target class C and interdependent on each other, which ensures that redundant features will not be selected.

B. BINARY FIREFLY ALGORITHM

As a population-based stochastic global search algorithm, FA simulates the behavior of fireflies of approaching brightness. For any two fireflies, the less bright firefly is attracted and move to the brighter one. The FA was originally proposed to solve the continuous optimization problem and recent studies have shown the competitiveness of FA in various applications. Specifically, the superiority of the FA compared with other optimization methods has been demonstrated by various studies [21], [22], [38], [39], which motivated us to utilize it in the wrapper-based feature selection phase.

The FA search process is closely related to two important aspects: the variation of brightness and formulation of the attractiveness. The attractiveness of fireflies is proportional to their brightness and the attractiveness decreases as the distance between any two fireflies increases. Another important factor is the absorption coefficient which affects the attractiveness. The brightness of a firefly can be expressed as:

$$B(r) = B_0 e^{-\gamma r} \quad (7)$$

where B_0 is the original brightness, γ is a light absorption coefficient which is always a fixed value. r is the distance between any two fireflies. The distance r between two fireflies at the positions x_i and x_j is calculated by using Euclidean distance. This can be represented by:

$$r_{ij} = \|x_i - x_j\| = \sqrt{\sum_{k=1}^d (x_{i,k} - x_{j,k})^2} \quad (8)$$

where $x_{i,k}$ is the k th component of the spatial coordinate x_i of the i th firefly. As a firefly's attractiveness is proportional to the brightness seen by another firefly, the attractiveness A of a firefly can be expressed as:

$$A(r) = A_0 e^{-\gamma r} \quad (9)$$

where A_0 is the attractiveness when the $r = 0$. The firefly i with a lower brightness is attracted by the firefly j with a higher brightness. The movement of a firefly is formulated by:

$$x_i(d + 1) = x_{id} + A_0 e^{-\gamma r_{ij}^2} (x_{jd} - x_{id}) + \alpha (\text{rand} - \frac{1}{2}) \quad (10)$$

where α is a randomization parameter and rand is a random number uniformly distributed in $[0, 1]$, r_{ij} represents the Euclidean distance between the i th firefly and the j th firefly, and d is the iteration index. As recommended by previous works [20], [22], in this paper, we set $\gamma = 1$, $A_0 = 1$ and $\alpha \in [0, 1]$.

When utilizing formula (10) to calculate the movement of i th firefly to j th firefly, the position of firefly changes from a binary vector to a real-valued vector. In order to obtain

the binary positions of fireflies, a probabilistic rule based on a hyperbolic tangent sigmoid transfer function is applied, which is shown in (11):

$$x_{id} = \begin{cases} 1, & \text{if } rand < S(x_{id}) \\ 0, & \text{otherwise} \end{cases}$$

$$S(x_{id}) = \tanh(|x_{id}|) = \frac{\exp(2 * |x_{id}|) - 1}{\exp(2 * |x_{id}|) + 1} \quad (11)$$

where *rand* is a random number uniformly distributed in [0, 1] and *S*(·) is the hyperbolic tangent sigmoid transfer function.

C. KERNEL EXTREME LEARNING MACHINE

Extreme learning machine (ELM) is a single hidden layer feed-forward neural network, which has many advantages such as fast training speed and excellent generalization ability. All these advantages make it successfully applied to HAR research. KELM extends ELM from explicit activation to implicit mapping functions. Some studies demonstrated it has better generalization performance than the traditional ELM algorithm. KELM is described as follows:

For any *N* different samples $(\mathbf{x}_j, \mathbf{t}_j), j = 1, 2, \dots, N$, where $\mathbf{x}_j = [x_{j1}, x_{j2}, \dots, x_{jn}]^T$ is the *j*th *n*-dimensional feature sample, and $\mathbf{t}_j = [t_{j1}, t_{j2}, \dots, t_{jm}]^T$ is the corresponding class label. All samples belong to *m* different classes, and the ELM mathematical model with *L* hidden neurons can be expressed as:

$$\sum_{i=1}^L \beta_i g(\mathbf{w}_i \cdot \mathbf{x}_j + b_i) = \mathbf{t}_j, \quad j = 1, 2, \dots, N \quad (12)$$

where *g*(*x*) is the excitation function, \mathbf{w}_i, b_i , and β_i are the vector of input weights, hidden layer bias and vector of output weights of the *i*th hidden neuron node respectively. Equation (12) can be written in matrix form:

$$\mathbf{H}\beta = \mathbf{T} \quad (13)$$

where β represents the output weight, \mathbf{T} is the corresponding coding class label, and \mathbf{H} is the hidden layer output matrix:

$$\mathbf{H} = \begin{bmatrix} g(\mathbf{w}_1 \cdot \mathbf{x}_1 + b_1) & \cdots & g(\mathbf{w}_L \cdot \mathbf{x}_1 + b_L) \\ \vdots & \cdots & \vdots \\ g(\mathbf{w}_1 \cdot \mathbf{x}_N + b_1) & \cdots & g(\mathbf{w}_L \cdot \mathbf{x}_N + b_L) \end{bmatrix}_{N \times L} \quad (14)$$

Since the activation function in the output layer of ELM is linear, the vector of output layer weights, β , is obtained by the following equation:

$$\beta = \mathbf{H}^\dagger \mathbf{T} \quad (15)$$

where \mathbf{H}^\dagger is the generalized inverse matrix of \mathbf{H} . In order to further improve the generalization ability of ELM, Huang *et al.* [40] introduced a kernel function to avoid the problem of ELM with randomly generating input weight and bias value. The calculation formula of KELM output weight is as follows:

$$\beta = \mathbf{H}^T \left(\frac{1}{C} + \mathbf{H}\mathbf{H}^T \right)^{-1} \mathbf{T} \quad (16)$$

where *C* is regularization coefficient. The output function for the SLFN is:

$$f(\mathbf{x}_j) = h(\mathbf{x}_j)\beta = h(\mathbf{x}_j)\mathbf{H}^T \left(\frac{1}{C} + \mathbf{H}\mathbf{H}^T \right)^{-1} \mathbf{T} \quad (17)$$

where *h*(\mathbf{x}_j) is the output of the hidden nodes and actually maps the data from input space to the hidden layer feature space \mathbf{H} . When the hidden layer function *h*(\mathbf{x}_j) is unknown, the kernel function matrix is calculated as follows:

$$\Omega_{\text{ELM}} = \mathbf{H}\mathbf{H}^T : \Omega_{\text{ELM},i,j} = h(\mathbf{x}_i) \cdot h(\mathbf{x}_j) = K(\mathbf{x}_i, \mathbf{x}_j) \quad (18)$$

where *K*($\mathbf{x}_i, \mathbf{x}_j$) represents the kernel function. In this paper, the most commonly utilized Gaussian kernel function was applied. The form of RBF kernels is as following forms:

$$K(\mathbf{x}, \mathbf{x}_1) = \exp\left(-\frac{\|\mathbf{x} - \mathbf{x}_1\|}{\gamma^2}\right) \quad (19)$$

When using RBF as the kernel function, two major parameters applied in KELM are *C* and γ . In order to achieve a higher classification accuracy, it is necessary to search for the optimal *C* and γ . Then the output function of KELM can be written as:

$$f(\mathbf{x}) = \begin{bmatrix} K(\mathbf{x}, \mathbf{x}_1) \\ \vdots \\ K(\mathbf{x}, \mathbf{x}_N) \end{bmatrix}^T \left(\frac{1}{C} + \Omega_{\text{ELM}} \right)^{-1} \mathbf{T} \quad (20)$$

IV. THE PROPOSED GTFS-BFA BASED HAR FRAMEWORK

A. OVERVIEW OF THE PROPOSED FRAMEWORK

The flowchart of the proposed HAR framework is shown in Figure 1. Using the original high-dimensional feature set to train the classifier will take a lot of computational resources. Therefore, in the proposed approach, the importance of each feature with GTFS is first evaluated to effectively reduce the large feature space. However, the obtained feature ranking does not consider the interaction between features. Then the optimal combination of feature set and parameters of KELM are obtained by using the BFA to maximize the recognition accuracy. The feature re-selection and classifier optimization are conducted synchronously in the wrapper-based phase. Lastly, the optimal KELM and feature set are utilized in HAR. The proposed HAR approach includes original feature extraction, GTFS-based feature pre-selection, BFA-based feature re-selection and KELM parameter optimization and activity recognition. The details of the proposed HAR approach is described as below.

- (1) Original feature extraction: since traditional classification algorithms are not suitable for the time-series sensor data, this type of data needs to be divided into segments before extracting features. Sliding window techniques, including sliding window with overlapping and sliding window without overlapping between two consecutive windows, are widely adopted and has been proved effective. Then features are extracted from these sliding windows to construct the training and

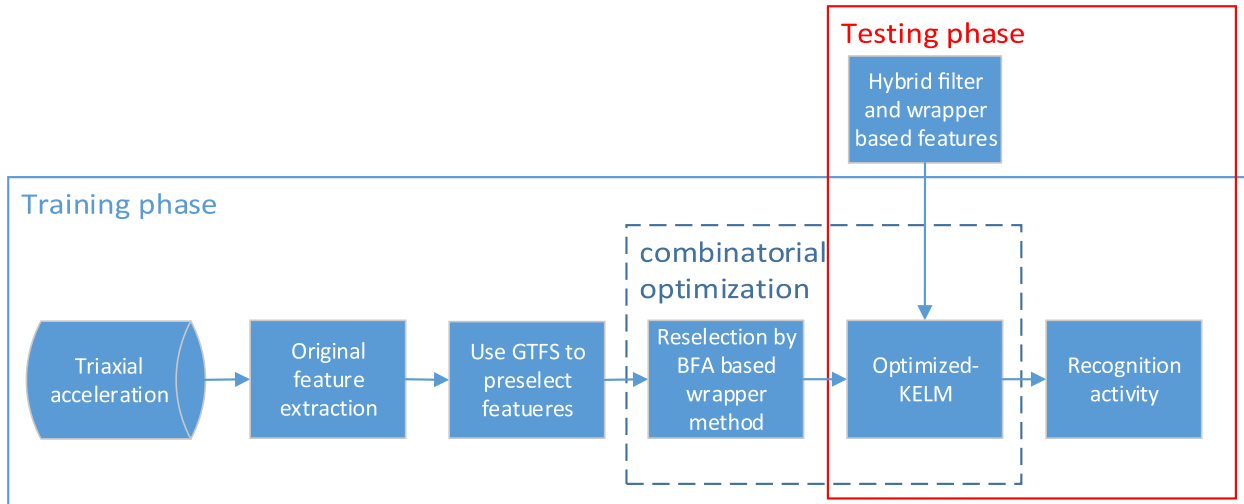


FIGURE 1. The architecture of the proposed HAR framework.

testing sample dataset. The raw feature vector is with high dimensionality and it not only includes relevant features for HAR but also irrelevant and redundant ones. Feature selection once generally cannot obtain representative features that characterize the activity type, which will lead to poor classification accuracy. Therefore, the feature pre-selection phase and re-selection phase are considered in this paper.

- (2) Feature pre-selection phase: using the original high-dimensional feature set as the inputs of the classifier would not only consume a lot of computation time but also decrease the recognition performance. As a filter-based method, GTFS has the merit that it takes less computational resources, so in this phase, the GTFS is utilized to calculate the feature weights and output an optimized feature set, which will reduce the dimensionality of the feature vector and preselect some top-ranked features advantageous to classification. While it is less effective without considering the influence of a classifier. Therefore, the preselected features will be provided to the next wrapper phase.
- (3) Feature re-selection phase and KELM parameter optimization: compared with the filter-based method, the wrapper-based method is more effective because it selects the optimal feature subset with the evaluation of a classifier. Therefore, in this phase, BFA is utilized to optimize the feature set and classifier parameters synchronously. Then the optimal feature set and optimal KELM model with the highest training accuracy will be obtained.
- (4) Activity recognition: in the testing phase, the corresponding optimal feature set from the testing high-dimensional feature vector is selected as the input of the optimized KELM. Eventually, the testing dataset is used to verify the proposed HAR approach.

B. ORIGINAL FEATURE EXTRACTION FROM ACCELERATION SIGNAL

Since the raw acceleration data are noisy and not representative of different activities, features from the raw data are more discriminative representations compared with raw acceleration data. Many features including time domain and frequency domain have been proven to be effective for HAR. For example, the signal magnitude area of acceleration can be utilized to recognize walking and fall. Frequency domain features show the distribution of signal energy, which help recognize dynamic activities from static ones. In this work, 24 features including time and frequency domain features from three-axis acceleration data are extracted as listed in Table 1.

TABLE 1. Extracted features.

| | | |
|-------------------------|--------------------------|------------------------|
| Mean | Root mean square | Kurtosis |
| Signal magnitude vector | Maximum | Minimum |
| Skewness | Signal magnitude area | Standard deviation |
| Impulse Factor | FFT coefficient | Absolute mean value |
| Energy | Correlation between axes | Variance |
| Margin factor | Weighted average | Frequency center |
| Tilt angle | Shape factor | Cross correlation |
| Interquartile range | Crest factor | Root variant frequency |

C. FEATURE PRE-SELECTION PHASE USING GTFS

There are uncorrelated or redundant features in the original training dataset that do not contribute to the recognition accuracy. Furthermore, if the original high dimensional features are utilized as inputs of classifier directly, then high computational cost will be incurred. Therefore, it is necessary to optimize the optimal feature subset to achieve good recognition performance and reduce computational cost. GTFS can evaluate the weight of each feature and select features which have strong discriminatory power as a group. Higher discriminatory features will be preselected by GTFS to form a feature subset. Thus, the pre-selected feature set can be re-selected in the next wrapper phase to obtain the optimal feature set. Before the pre-selection phase, feature values are normalized to the range [0, 1] to eliminate the influence caused by different dimensionalities and orders of magnitude. The normalization equation is as follows:

$$w'_i = \frac{w_i - \min(w_i)}{\max(w_i) - \min(w_i)} \quad (21)$$

where w'_i is the normalized value. w_i is the value of the original feature, $\min(w_i)$ and $\max(w_i)$ are minimum value and maximum value of feature w_i respectively.

D. BFA ENCODING FOR COMBINATIONAL OPTIMIZATION

Filter-based GTFS method is used for the sake of reducing the search dimension and improving computing efficiency as there may still be redundant features in the feature set. In order to obtain the optimal feature subset that can be utilized to improve the performance of HAR, the wrapper method is introduced as a re-selection phase. The features pre-selected by GTFS are utilized as candidates which are further optimized by the BFA. Besides, considering the robustness of the HAR system, BFA is also utilized to simultaneously search for the optimal KELM parameters. In this phase, the two parameters of RBF kernel function C and γ and the best feature subset will be optimized.

By applying the BFA for optimizing feature set and recognition algorithm simultaneously, a population of fireflies are utilized to search in the solution space. The BFA performs through updating the individuals from one iteration to another. To initialize the BFA optimization algorithm, each individual is coded to represent the feature selection state and the values of parameters C and γ . In our method of initializing BFA, the coding state of each firefly is divided into three parts, which include the states of feature selection and the value of parameter C and parameter γ . The firefly encoding in our method is illustrated in Figure 2.

The selection state of features selected by GTFS is represented by bits of the coded sequence in each firefly. Consequently, N_p bits are utilized to represent the feature selection state in a candidate. The bit value “1” indicates that the feature is selected while the value “0” indicates that the feature is abandoned. For parameter C and γ , N_c and N_g bits are utilized to represent their values respectively. The

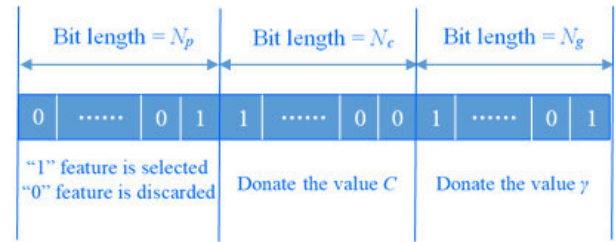


FIGURE 2. Illustration of the firefly encoding in BFA.

following formula is utilized to calculate the decimal value of the parameter from the value of each bit:

$$x_d = \frac{\sum_{i=1}^N (bit(i) \times 2^{i-1})}{2^N - 1} (x_{d\max} - x_{d\min}) + x_{d\min} \quad (22)$$

where N is the number of bits, x_d is the decimal value of C or γ , i is the bit index, $bit(i)$ is the value of the i th bit, $x_{d\min}$ and $x_{d\max}$ are the lower and upper bounds of the searching interval of C or γ respectively.

E. FEATURE RE-SELECTION AND COMBINATIONAL OPTIMIZATION FOR KELM PARAMETERS

After initializing the BFA optimization algorithm, it is used for searching the optimal feature subset and the KELM parameters. In the BFA iteration, the bits of each individual represents a reselected feature subset and the value of the KELM parameter. The fitness value of each individual is calculated by formula (23) to obtain the optimal result in the iteration process.

$$fit = A_{tr} - \alpha N_s \quad (23)$$

where fit is the fitness value, A_{tr} is the training accuracy, N_s is the number of selected features and α is the weighting factor whose value is very small, such as 0.01. As the objective is to obtain higher recognition accuracy with fewer features, the fitness function of the firefly considers the training accuracy and the number of features. This selection criterion finds the combination of features and parameter values for the goal of achieving the highest classification accuracy with a number of features as less as possible. Figure 3 illustrates the proposed synchronous of feature re-selection and KELM parameters optimization with BFA.

As the iteration continues, the position value of all fireflies will be updated until the required number of iterations is reached. Finally, the re-optimized feature set and the optimal KELM parameters with the highest training recognition accuracy are obtained. In the testing phase, the corresponding features are selected from the testing data and the activity type can be obtained by using these features and the optimized KELM.

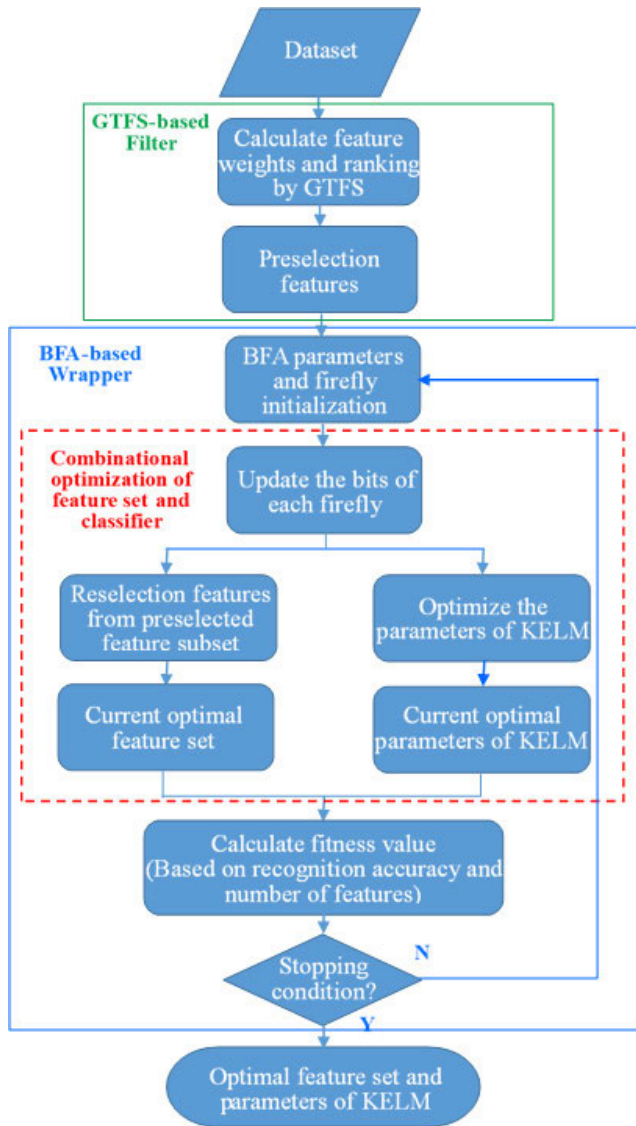


FIGURE 3. The flow chart of combinational optimization for feature set and KELM using BFA.

V. EXPERIMENTAL SETUP AND ACCELERATION DATA ACQUISITION

A. DATASET

In order to evaluate the performance of the proposed approach for HAR, we acquired the dataset by utilizing the TRIGNOTM wireless system from Delsys Company which contains a base station and collection nodes. The collection node integrates a triaxial accelerometer, which has an acceleration range of ±6G with resolution = 0.016 (G is the gravitational constant). The TRIGNOTM wireless system has wireless transmission function and the acceleration signal can be transmitted to the base station from the collection nodes. Once received by the base station, the data can be transmitted and stored in the computer. The datasets were collected when each of the ten volunteers aged between twenty and forty-five performed activities with five collection nodes respectively

attached to the chest, waist, left wrist, left ankle and right arm. Accordingly, the dataset from five different positions of the body can be obtained. Figure 4 presents the fixed positions of the collection nodes and the data collection process. Before the start of data collection, we utilized straps to fix the sensors on the body and checked the sensors were in the same position as the previous subject. The triaxial accelerometer worked at a sampling frequency of 150 Hz. As we mainly focus on the recognition of daily activities, the task on this dataset is to distinguish six basic daily activities which include five dynamic activity walking (W), going upstairs (GU), going downstairs (GD), running (R), jumping (J) and one static activity standing (S). Figure 5 shows the triaxial accelerometer data of different activities from the left ankle. After data acquisition, sliding windows are used to divide the acceleration signal into segments. The window length is 300 data points and adjacent windows contain 50% overlap.

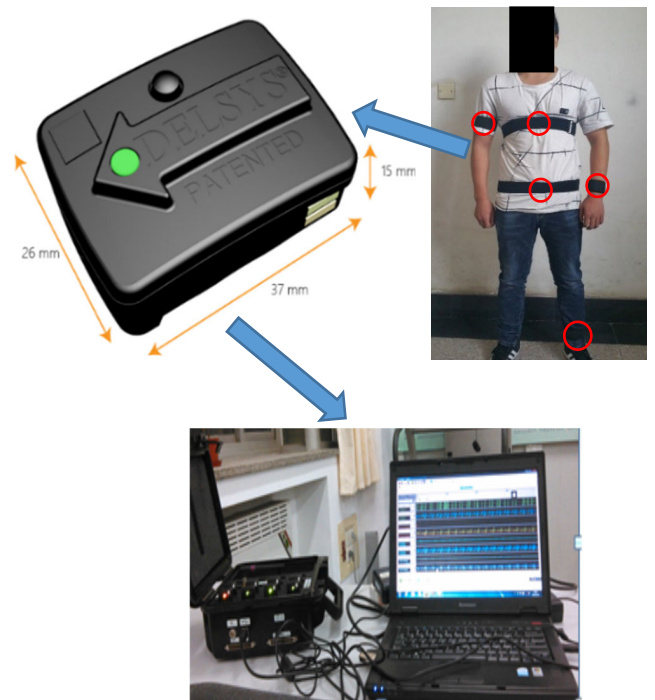


FIGURE 4. The positions of sensor nodes and graphical depiction of data collection process.

B. PERFORMANCE MEASURES

To evaluate the effectiveness of the proposed method and show its superiority over the comparative methods, we compare them in terms of the number of selected features and obtained recognition performance which will be measured by the following four measures:

The accuracy measure is used to evaluate the performance of the proposed method, which can be expressed as:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{24}$$

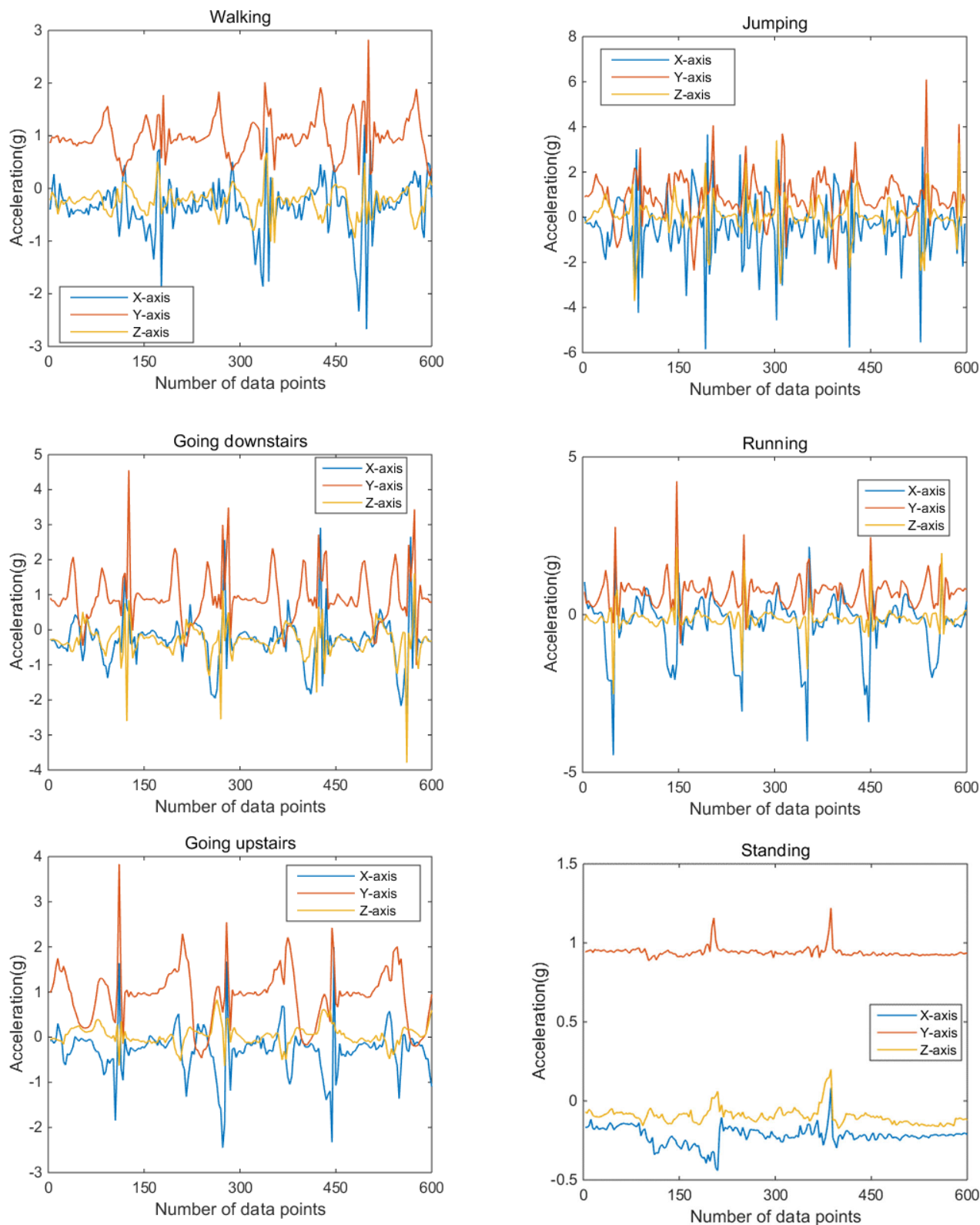


FIGURE 5. Sample accelerometer data of different activities from left ankle.

where TP , TN , FP , and FN , respectively, represent the number of true positive, true negative, false positive, and false negative outcomes in a given experiment.

Precision and recall are defined as measuring the recognition rate of records correctly classified from a class of total positive records and the recognition rate of records

correctly classified from a class of total true samples in a class, respectively.

$$\text{precision} = \frac{TP}{TP + FP} \tag{25}$$

$$\text{recall} = \frac{TP}{TP + FN} \tag{26}$$

In addition, F1 evaluation criterion is also considered. F1 is defined as the combination of precision and the recall, which is defined as follows:

$$F1 = \frac{2 \times \text{recall} \times \text{precision}}{\text{recall} + \text{precision}} \tag{27}$$

VI. EXPERIMENTAL SETUP AND RESULTS

The leave-one-out (LOO) strategy is utilized to train the classifier and recognize the activities. The validation is repeated ten times with data from each person used exactly once for testing. The final results are their average values. In the combinatorial optimization phase, we set the maximum iterations as 100 and the population size of fireflies as 30, absorption coefficient $\gamma = 1$, attractiveness $\beta_0 = 1$, and the stopping criterion was set as follows: the number of iterations reaches 100 or there is no improvement in the fitness for 10 consecutive iterations. As discussed in the former section, RBF kernel function is used in KELM classifier and parameter C is limited in the interval $[0.001, 100]$ encoded by 8 bits for each individual. Meanwhile, the parameter γ is limited in the interval $[0.1, 500]$ encoded by 16 bits for each individual. All our experiments were carried out in MATLAB 2014a using a desktop with a 3.2GHz processor and 8G memory storage.

A. THE PERFORMANCE OF THE PROPOSED METHOD ON DATA FROM DIFFERENT POSITIONS

To validate the effectiveness of the proposed HAR approach based on GTFS-BFA feature selection and combinatorial optimization, it is compared with several other HAR approaches in this study. Since the proposed feature selection method contains two phases of convergence, the number of selected features in the pre-selection phase is related to the performance of GTFS-BFA. Experimental data obtained in the pre-selection phase is also worthy of observation. Figure 6 shows the performance with sequential feature subsets in the GTFS-based filter phase. It is important to note that the results are only based on the feature set selected by GTFS.

As shown in Figure 6, too small number of features will not benefit the recognition accuracy, which is expected. Then, the accuracy increases as the number of features increases. However, when the number of features reaches a certain value, the accuracy of recognition will stop rising or even decreasing. This further proves that more features are not better, and there will be redundant features in the set that will pull down the recognition results. Therefore, the number of preselected features should be kept within a reasonable range. According to the data obtained in experiments, in the feature pre-selection phase, the highest accuracy always appears

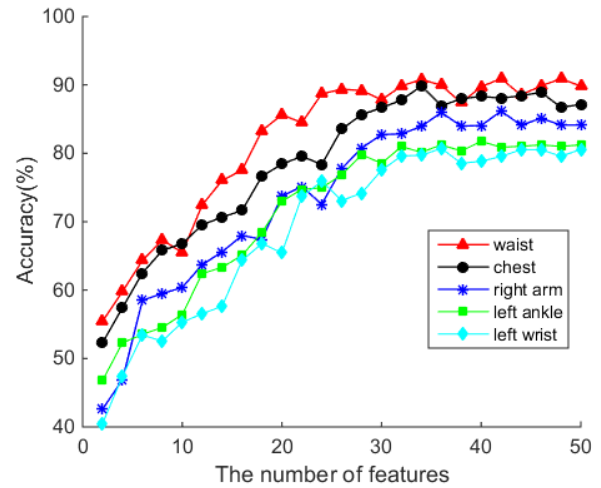


FIGURE 6. The accuracy with the sequential GTFS feature subsets.

when the feature number is between 25 and 35. Therefore, we select the 30 features with the highest rankings as the candidates for feature reselection phase. Accordingly, we set $N_p = 30$ bits in the firefly encoding to represent the state of feature in the reselection phase.

In this subsection, we show the performance of the proposed feature selection approach for HAR. In order to present an intuitive impression of the performance of the proposed combinatorial optimization method compared with original features, GTFS method and feature optimization, Figure 7 shows the performance comparison in terms of accuracy, precision, recall and F1 of these four methods on the data from five positions. In Figure 7, the codes ‘‘W’’,

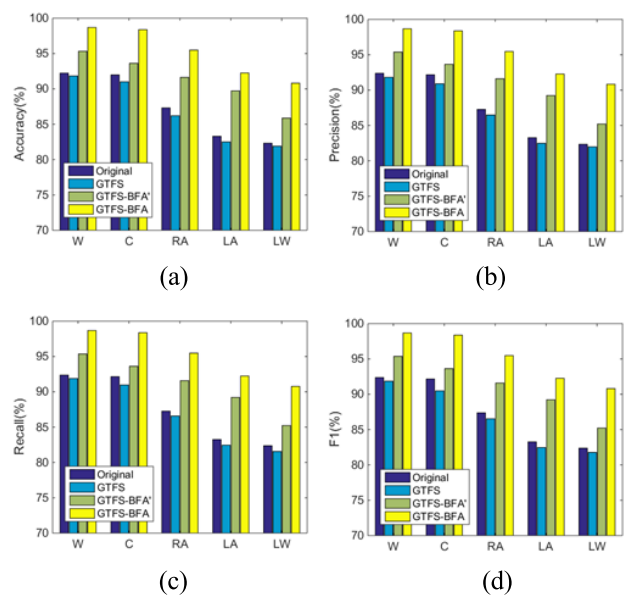


FIGURE 7. Performance comparison of different feature selection methods. (a) Accuracy comparison. (b) Precision comparison. (c) Recall comparison. (d) F1 comparison.

“C”, “RA”, “LA” and “LW” represent the body position waist, chest, right arm, left ankle and left wrist, respectively and GTFS-BFA’ means that only feature optimization is performed without parameter optimization. Generally, we can come to these conclusions: (1) the positions of the waist and chest are the ideal sensor wearing positions. (2) The GTFS-based filter method can achieve similar recognition performance to the original features but with fewer features. (3) The proposed combinational optimization method has superiority over the only optimization of features in achieving better performance. (4) The proposed combinational optimization GTFS-BFA have the best performance compared with the comparative methods, which demonstrates the effectiveness of GTFS-BFA.

Additionally, in order to gain a better insight into the activity recognition problem and the proposed feature selection method, the corresponding confusion matrixes are constructed, which are shown in Tables 2 to 6, respectively. According to the results, we can observe that the proposed method can distinguish dynamic activity (walking, running, jumping, going upstairs and downstairs) from static activity (standing) with a high accuracy. For example, for the position of waist and left ankle, only 3 and 4 dynamic activity samples are misrecognized as static activity (standing) respectively. Furthermore, for the position of waist and chest, the proposed method makes 4 and 5 recognition errors out of the 591 test samples respectively when recognizing the static activity (standing).

B. COMPARISON WITH OTHER STATE-OF-THE-ART APPROACHES

1) PERFORMANCE COMPARISON OF FEATURE OPTIMIZATION

To further demonstrate the effectiveness of the proposed GTFS-BFA approach for HAR, it is compared with other five existing state-of-the-art approaches including MBACO [41],

TABLE 2. Confusion matrix for the proposed method on the data from the waist.

| | W | R | GU | GD | J | S |
|----|-----|-----|-----|-----|-----|-----|
| W | 612 | 2 | 2 | 1 | 1 | 1 |
| R | 3 | 581 | 4 | 2 | 1 | 0 |
| GU | 2 | 3 | 546 | 4 | 1 | 1 |
| GD | 0 | 1 | 3 | 571 | 3 | 1 |
| J | 0 | 1 | 3 | 2 | 581 | 0 |
| S | 1 | 1 | 1 | 1 | 0 | 587 |

TABLE 3. Confusion matrix for the proposed method on the data from the chest.

| | W | R | GU | GD | J | S |
|----|-----|-----|-----|-----|-----|-----|
| W | 608 | 3 | 2 | 3 | 1 | 2 |
| R | 4 | 578 | 4 | 3 | 2 | 0 |
| GU | 3 | 4 | 539 | 4 | 6 | 1 |
| GD | 0 | 1 | 2 | 573 | 2 | 1 |
| J | 0 | 1 | 1 | 2 | 583 | 0 |
| S | 1 | 2 | 1 | 1 | 0 | 586 |

TABLE 4. Confusion matrix for the proposed method on the data from the right arm.

| | W | R | GU | GD | J | S |
|----|-----|-----|-----|-----|-----|-----|
| W | 587 | 12 | 7 | 6 | 4 | 3 |
| R | 4 | 554 | 12 | 14 | 6 | 1 |
| GU | 2 | 6 | 527 | 19 | 1 | 2 |
| GD | 1 | 4 | 21 | 542 | 11 | 0 |
| J | 1 | 2 | 6 | 6 | 572 | 0 |
| S | 2 | 2 | 2 | 1 | 1 | 583 |

TABLE 5. Confusion matrix for the proposed method on the data from the left ankle.

| | W | R | GU | GD | J | S |
|----|-----|-----|-----|-----|-----|-----|
| W | 562 | 27 | 18 | 7 | 4 | 1 |
| R | 25 | 537 | 19 | 7 | 3 | 0 |
| GU | 8 | 16 | 504 | 24 | 5 | 0 |
| GD | 4 | 11 | 24 | 523 | 14 | 3 |
| J | 2 | 4 | 21 | 12 | 548 | 0 |
| S | 3 | 1 | 3 | 5 | 2 | 577 |

TABLE 6. Confusion matrix for the proposed method on the data from the left wrist.

| | W | R | GU | GD | J | S |
|----|-----|-----|-----|-----|-----|-----|
| W | 574 | 20 | 14 | 7 | 3 | 1 |
| R | 27 | 523 | 19 | 14 | 6 | 2 |
| GU | 7 | 18 | 502 | 21 | 9 | 0 |
| GD | 5 | 21 | 27 | 507 | 17 | 2 |
| J | 3 | 17 | 19 | 23 | 523 | 2 |
| S | 3 | 3 | 3 | 5 | 6 | 571 |

ReliefF-BPSO [42], ReliefF-GA [43], CMIM-GA [44], and MPI-FA [38]. All these heuristic methods are the binary searching algorithms. In order to demonstrate the performances of each algorithm empirically, the same population size and the same number of iterations are set for these binary searching algorithms, other parameters are set as described in their respective references. For ReliefF-BPSO, ReliefF-GA, CMIM-GA and MPI-FA, the filter phase selects the same number of features with the proposed GTFS-BFA. The MBACO method selects features from the original feature set. The four performance measures and the number of the selected features (d) are utilized to show the performance of the six methods. In this section, we only verify the performance of these methods in optimizing features without considering the combinational optimization of the features and parameters. Tables 7 to 11 present the performance comparison of different methods for the data from five body positions, respectively.

As can be seen from the results, the proposed GTFS-BFA method has the best optimization ability in feature selection for HAR, its average performance is the best for data from the five positions, and the number of the selected features (d) is in the range from 10 to 16, which is obviously fewer than other methods. For the data from the waist, the proposed GTFS-BFA achieves 95.32% accuracy with average 10 features, which is better than the other five algorithms. In particular, the MBACO method selects the most features

TABLE 7. Performance comparison of feature optimization for different methods using data from the waist.

| Method | Acc | Pre | Rec | F1 | <i>d</i> |
|-------------------|--------|--------|--------|--------|----------|
| MBACO | 92.43% | 92.46% | 92.45% | 92.45% | 30 |
| ReliefF-BPSO | 93.68% | 93.71% | 93.69% | 93.70% | 16 |
| ReliefF-GA | 93.14% | 93.15% | 94.14% | 93.64% | 18 |
| CMIM-GA | 93.76% | 93.78% | 93.77% | 93.77% | 17 |
| MPI-FA | 94.43% | 94.46% | 94.44% | 94.45% | 13 |
| Proposed GTFS-BFA | 95.32% | 95.38% | 95.36% | 95.37% | 10 |

TABLE 8. Performance comparison of feature optimization for different methods using data from the chest.

| Method | Acc | Pre | Rec | F1 | <i>d</i> |
|-------------------|--------|--------|--------|--------|----------|
| MBACO | 91.52% | 91.58% | 91.53% | 91.55% | 32 |
| ReliefF-BPSO | 92.58% | 92.59% | 92.58% | 92.58% | 16 |
| ReliefF-GA | 92.15% | 92.17% | 92.16% | 92.16% | 20 |
| CMIM-GA | 92.43% | 92.44% | 92.43% | 92.43% | 17 |
| MPI-FA | 93.18% | 93.19% | 93.18% | 93.18% | 13 |
| Proposed GTFS-BFA | 93.62% | 93.65% | 93.62% | 93.63% | 12 |

TABLE 9. Performance comparison of feature optimization for different methods using data from the right arm.

| Method | Acc | Pre | Rec | F1 | <i>d</i> |
|-------------------|--------|--------|--------|--------|----------|
| MBACO | 87.48% | 87.46% | 87.45% | 87.45% | 35 |
| ReliefF-BPSO | 89.77% | 89.75% | 89.74% | 89.74% | 17 |
| ReliefF-GA | 89.45% | 89.42% | 89.39% | 89.40% | 19 |
| CMIM-GA | 90.23% | 90.21% | 90.18% | 90.19% | 18 |
| MPI-FA | 90.82% | 90.81% | 90.78% | 90.79% | 14 |
| Proposed GTFS-BFA | 91.63% | 91.61% | 91.58% | 91.59% | 13 |

compared with other methods but the performance is very poor. This demonstrates that combining filter and wrapper phase helps to improve the performance of feature selection.

TABLE 10. Performance comparison of feature optimization for different methods using data from the left ankle.

| Method | Acc | Pre | Rec | F1 | <i>d</i> |
|-------------------|--------|--------|--------|--------|----------|
| MBACO | 84.47% | 84.49% | 84.46% | 84.47% | 33 |
| ReliefF-BPSO | 86.92% | 86.94% | 86.93% | 86.93% | 18 |
| ReliefF-GA | 86.07% | 86.08% | 86.05% | 86.06% | 21 |
| CMIM-GA | 86.38% | 86.39% | 86.34% | 86.36% | 20 |
| MPI-FA | 88.64% | 88.65% | 88.63% | 88.64% | 16 |
| Proposed GTFS-BFA | 89.72% | 89.75% | 89.73% | 89.73% | 15 |

TABLE 11. Performance comparison of feature optimization for different methods using data from the left wrist.

| Method | Acc | Pre | Rec | F1 | <i>d</i> |
|-------------------|--------|--------|--------|--------|----------|
| MBACO | 82.27% | 82.28% | 82.27% | 82.27% | 33 |
| ReliefF-BPSO | 83.29% | 83.29% | 83.28% | 83.28% | 20 |
| ReliefF-GA | 82.74% | 82.75% | 82.74% | 82.74% | 21 |
| CMIM-GA | 83.28% | 83.29% | 83.27% | 83.28% | 19 |
| MPI-FA | 84.72% | 84.74% | 84.72% | 84.73% | 18 |
| Proposed GTFS-BFA | 85.87% | 85.89% | 85.86% | 85.87% | 16 |

In addition, it can be seen from the results that the MPI-FA and the proposed GTFS-BFA select fewer features compared with BPSO and GA based methods when the number of features in the filter phase is the same. This demonstrates that FA has better optimization capability for features in HAR compared with BPSO and GA.

2) PERFORMANCE COMPARISON OF COMBINATIONAL OPTIMIZATION

In this subsection, we analyze the performance of combinational optimization of the proposed and comparative methods for HAR. This helps us to understand the superiority of combination optimization compared to feature optimization and verify the effectiveness of the proposed method for HAR. To demonstrate the performances empirically, the parameters of all the methods are set as the same as the previous section. Tables 12 to 16 respectively show the performance comparison of different methods for the data from five body positions when combinational optimizing the feature and classifier is considered.

TABLE 12. Performance comparison of combinational optimization for different methods using data from the waist.

| Method | Acc | Pre | Rec | F1 | <i>d</i> |
|-------------------|--------|--------|--------|--------|----------|
| MBACO | 92.95% | 92.94% | 92.94% | 92.94% | 31 |
| ReliefF-BPSO | 94.72% | 94.71% | 94.72% | 94.71% | 17 |
| ReliefF-GA | 94.28% | 94.28% | 94.27% | 94.27% | 18 |
| CMIM-GA | 95.64% | 95.65% | 95.64% | 95.69% | 17 |
| MPI-FA | 97.68% | 97.69% | 97.68% | 97.68% | 12 |
| Proposed GTFS-BFA | 98.69% | 98.68% | 98.68% | 98.68% | 11 |

TABLE 13. Performance comparison of combinational optimization for different methods using data from the chest.

| Method | Acc | Pre | Rec | F1 | <i>d</i> |
|-------------------|--------|--------|--------|--------|----------|
| MBACO | 92.12% | 92.12% | 92.12% | 92.12% | 32 |
| ReliefF-BPSO | 94.07% | 94.08% | 94.05% | 94.06% | 15 |
| ReliefF-GA | 93.67% | 93.69% | 93.65% | 93.67% | 18 |
| CMIM-GA | 94.23% | 94.25% | 94.22% | 94.23% | 17 |
| MPI-FA | 96.32% | 96.34% | 96.31% | 96.32% | 13 |
| Proposed GTFS-BFA | 98.38% | 98.38% | 98.37% | 98.37% | 12 |

It can be seen from Tables 12 to 16, the proposed GTFS-BFA also outperforms the other methods when combinational optimization is considered. In addition, all the methods have improved performance compared with the results of only performing feature optimization, which demonstrates the combinational optimization for optimal parameters and feature subset at the same time can improve the performance and robustness of the HAR system. For example, the average recognition accuracy using the proposed GTFS-BFA has reached 98.69%, 98.38%, 95.49%, 92.25% and 90.81% for the data from five positions, respectively, which are all higher than using GTFS-BFA for feature optimization. Moreover, the number of features utilized by the proposed GTFS-BFA for combinational optimization will not increase significantly. For example, with the same number of features, the proposed GTFS-BFA achieves 98.38% and 95.49% accuracy on data from the chest and right arm, which is obviously better than 93.62% and 91.63% accuracy when only feature

TABLE 14. Performance comparison of combinational optimization for different methods using data from the right arm.

| Method | Acc | Pre | Rec | F1 | <i>d</i> |
|-------------------|--------|--------|--------|--------|----------|
| MBACO | 89.28% | 89.26% | 89.27% | 89.26% | 34 |
| ReliefF-BPSO | 91.36% | 91.35% | 91.36% | 91.35% | 18 |
| ReliefF-GA | 90.36% | 90.35% | 90.36% | 90.35% | 21 |
| CMIM-GA | 93.63% | 93.61% | 93.62% | 93.61% | 17 |
| MPI-FA | 94.27% | 94.25% | 94.26% | 94.25% | 14 |
| Proposed GTFS-BFA | 95.49% | 95.46% | 95.48% | 95.47% | 13 |

TABLE 15. Performance comparison of combinational optimization for different methods using data from the left ankle.

| Method | Acc | Pre | Rec | F1 | <i>d</i> |
|-------------------|--------|--------|--------|--------|----------|
| MBACO | 86.84% | 86.87% | 86.83% | 86.85% | 33 |
| ReliefF-BPSO | 88.91% | 88.93% | 88.91% | 88.92% | 18 |
| ReliefF-GA | 88.37% | 88.39% | 88.36% | 88.37% | 22 |
| CMIM-GA | 90.14% | 90.15% | 90.13% | 90.14% | 20 |
| MPI-FA | 91.38% | 91.39% | 91.36% | 91.37% | 17 |
| Proposed GTFS-BFA | 92.25% | 92.28% | 92.24% | 92.26% | 15 |

TABLE 16. Performance comparison of combinational optimization for different methods using data from the left wrist.

| Method | Acc | Pre | Rec | F1 | <i>d</i> |
|-------------------|--------|--------|--------|--------|----------|
| MBACO | 82.97% | 82.99% | 82.95% | 82.97% | 32 |
| ReliefF-BPSO | 84.97% | 84.98% | 84.96% | 84.97% | 21 |
| ReliefF-GA | 84.37% | 84.39% | 84.35% | 84.37% | 23 |
| CMIM-GA | 85.49% | 85.51% | 85.48% | 85.49% | 22 |
| MPI-FA | 87.25% | 87.26% | 87.24% | 87.25% | 17 |
| Proposed GTFS-BFA | 90.81% | 90.83% | 90.77% | 90.8% | 14 |

optimization is performed. In brief, the proposed GTFS-BFA has superior optimization ability for features and parameters of HAR.

VII. CONCLUSION

An effective and robust HAR framework, which can conduct the feature selection and classifier optimization synchronously, is proposed to achieve better recognition performance. The proposed framework is composed of a GTFS-based feature pre-selection phase and an FA-based combinational optimization phase. The filter-based GTFS is firstly employed to eliminate irrelevant and redundant features, to form a reduced input subset. The wrapper-based combinational optimization phase is then applied to reselect the feature subset and find out the optimal classifier parameters synchronously. Several experiments with data from five body positions have been conducted to analyze the effectiveness of the proposed method. Experimental results have shown that the HAR performance of the proposed GTFS-BFA based approach is superior to other well-established counterparts. Therefore, the proposed GTFS-BFA method can be a good alternative for feature selection in HAR. The solution proposed in this paper has the following applications: first, the proposed HAR framework with accurate activity recognition benefits the design and development of human-centric applications such as assisted living system, rehabilitation system and fall detection system, etc. Secondly, we tested the effectiveness of the proposed feature selection method GTFS-BFA in identifying six daily activities, but it is actually a general function that can be applied to other situations, such as other classification and regression problems.

In future work, since the inertial sensor fusion provides the mechanism to estimate orientation and rotation of movement, we plan to combine various sensors to identify more kinds of activities. For example, the combination of gyroscope, magnetometer and acceleration can be used to distinguish the activity of similar pattern such as upstairs, downstairs and cycling. Another future work is to apply the proposed GTFS-BFA approach in other related areas and compares it with other effective feature selection methods.

REFERENCES

- [1] Y. Wang, S. Cang, and H. Yu, "A survey on wearable sensor modality centred human activity recognition in health care," *Expert Syst. Appl.*, vol. 137, pp. 167–190, Dec. 2019.
- [2] L. Chen, R. Li, H. Zhang, L. Tian, and N. Chen, "Intelligent fall detection method based on accelerometer data from a wrist-worn smart watch," *Measurement*, vol. 140, pp. 215–226, Jul. 2019.
- [3] T. Wang, Z. Wang, D. Zhang, T. Gu, H. Ni, J. Jia, X. Zhou, and J. Lv, "Recognizing parkinsonian gait pattern by exploiting fine-grained movement function features," *ACM Trans. Intell. Syst. Technol.*, vol. 8, no. 1, pp. 1–22, Oct. 2016.
- [4] R. Mooney, G. Corley, A. Godfrey, L. Quinlan, and G. ÓLaighin, "Inertial sensor technology for elite swimming performance analysis: A systematic review," *Sensors*, vol. 16, no. 1, p. 18, Dec. 2015.
- [5] Z. Wang, M. Guo, and C. Zhao, "Badminton stroke recognition based on body sensor networks," *IEEE Trans. Human-Mach. Syst.*, vol. 46, no. 5, pp. 769–775, Oct. 2016.
- [6] R. Gravina, P. Alinia, H. Ghasemzadeh, and G. Fortino, "Multi-sensor fusion in body sensor networks: State-of-the-art and research challenges," *Inf. Fusion*, vol. 35, pp. 68–80, May 2017.
- [7] M. Cornacchia, K. Ozcan, Y. Zheng, and S. Velipasalar, "A survey on activity detection and classification using wearable sensors," *IEEE Sensors J.*, vol. 17, no. 2, pp. 386–403, Jan. 2017.
- [8] O. D. Lara and M. A. Labrador, "A survey on human activity recognition using wearable sensors," *IEEE Commun. Surveys Tuts.*, vol. 15, no. 3, pp. 1192–1209, Jul. 2013.
- [9] F. S. Ayachi, H. P. Nguyen, E. G. de Brugiere, P. Boissy, and C. Duval, "The use of empirical mode decomposition-based algorithm and inertial measurement units to auto-detect daily living activities of healthy adults," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 24, no. 10, pp. 1060–1070, Oct. 2016.
- [10] Z. Wang, D. Wu, J. Chen, A. Ghoneim, and M. A. Hossain, "A triaxial accelerometer-based human activity recognition via EEMD-based features and game-theory-based feature selection," *IEEE Sensors J.*, vol. 16, no. 9, pp. 3198–3207, May 2016.
- [11] G. Chandrashekar and F. Sahin, "A survey on feature selection methods," *Comput. Elect. Eng.*, vol. 40, no. 1, pp. 16–28, Jan. 2014.
- [12] S. Chernbumroong, S. Cang, A. Atkins, and H. Yu, "Elderly activities recognition and classification for applications in assisted living," *Expert Syst. Appl.*, vol. 40, no. 5, pp. 1662–1674, Apr. 2013.
- [13] E. Zdravevski, P. Lameski, V. Trajkovik, A. Kulakov, I. Chorbev, R. Goleva, N. Pombo, and N. Garcia, "Improving activity recognition accuracy in ambient-assisted living systems by automated feature engineering," *IEEE Access*, vol. 5, pp. 5262–5280, 2017.
- [14] X. Sun, Y. Liu, J. Li, J. Zhu, X. Liu, and H. Chen, "Using cooperative game theory to optimize the feature selection problem," *Neurocomputing*, vol. 97, no. 15, pp. 86–93, 2012.
- [15] M. Cinque, C. Esposito, and A. Pecchia, "Security log analysis in critical industrial systems exploiting game theoretic feature selection and evidence combination," *IEEE Trans. Ind. Informat.*, vol. 16, no. 6, pp. 3871–3880, Jun. 2020.
- [16] D. Zhao, C. Huang, Y. Wei, F. Yu, M. Wang, and H. Chen, "An effective computational model for bankruptcy prediction using kernel extreme learning machine approach," *Comput. Econ.*, vol. 49, no. 2, pp. 325–341, 2017.
- [17] M. Wang, H. Chen, H. Li, Z. Cai, X. Zhao, C. Tong, J. Li, and X. Xu, "Grey wolf optimization evolving kernel extreme learning machine: Application to bankruptcy prediction," *Eng. Appl. Artif. Intell.*, vol. 63, pp. 54–68, Aug. 2017.
- [18] Y. Tain, J. Zhang, L. Chen, Y. Geng, and X. Wang, "Single wearable accelerometer-based human activity recognition via kernel discriminant analysis and QPSO-KELM classifier," *IEEE Access*, vol. 7, pp. 109216–109227, 2019.
- [19] F. Mohanty, S. Rup, B. Dash, B. Majhi, and M. N. S. Swamy, "An improved scheme for digital mammogram classification using weighted chaotic salp swarm algorithm-based kernel extreme learning machine," *Appl. Soft Comput.*, vol. 91, Jun. 2020, Art. no. 106266.
- [20] X. S. Yang, "Firefly algorithms for multimodal optimization," in *Stochastic Algorithms: Foundations and Applications*, O. Watanabe and T. Zeugmann, Eds. Berlin, Germany: Springer-Verlag, 2009, pp. 169–178.
- [21] K. Chandrasekaran and S. P. Simon, "Optimal deviation based firefly algorithm tuned fuzzy design for multi-objective UCP," *IEEE Trans. Power Syst.*, vol. 28, no. 1, pp. 460–471, Feb. 2013.
- [22] X.-S. Yang, S. S. S. Hosseini, and A. H. Gandomi, "Firefly algorithm for solving non-convex economic dispatch problems with valve loading effect," *Appl. Soft Comput.*, vol. 12, no. 3, pp. 1180–1186, 2012.
- [23] A. M. Khan, Y.-K. Lee, S. Y. Lee, and T.-S. Kim, "A triaxial accelerometer-based physical-activity recognition via augmented-signal features and a hierarchical recognizer," *IEEE Trans. Inf. Technol. Biomed.*, vol. 14, no. 5, pp. 1166–1172, Sep. 2010.
- [24] Y. Tian, X. Wang, L. Chen, and Z. Liu, "Wearable sensor-based human activity recognition via two-layer diversity-enhanced multiclassifier recognition method," *Sensors*, vol. 19, no. 9, p. 2039, Apr. 2019.
- [25] H. Ghasemzadeh, N. Amini, R. Saeedi, and M. Sarrafzadeh, "Power-aware computing in wearable sensor networks: An optimal feature selection," *IEEE Trans. Mobile Comput.*, vol. 14, no. 4, pp. 800–812, Apr. 2015.
- [26] Z. Chen, Q. Zhu, C. S. Yeng, and L. Zhang, "Robust human activity recognition using smartphone sensors via CT-PCA and online SVM," *IEEE Trans. Ind. Appl.*, vol. 13, no. 6, pp. 3070–3080, Dec. 2017.
- [27] A. Wang, G. Chen, J. Yang, S. Zhao, and C.-Y. Chang, "A comparative study on human activity recognition using inertial sensors in a smartphone," *IEEE Sensors J.*, vol. 16, no. 11, pp. 4566–4578, Jun. 2016.
- [28] S. Althloothi, M. H. Mahoor, X. Zhang, and R. M. Voyles, "Human activity recognition using multi-features and multiple kernel learning," *Pattern Recognit.*, vol. 47, no. 5, pp. 1800–1812, May 2014.

- [29] Y. Chen and Z. Wang, "A hierarchical method for human concurrent activity recognition using miniature inertial sensors," *Sensor Rev.*, vol. 37, no. 1, pp. 101–109, Jan. 2017.
- [30] F. Sikder and D. Sarkar, "Log-sum distance measures and its application to human-activity monitoring and recognition using data from motion sensors," *IEEE Sensors J.*, vol. 17, no. 14, pp. 4520–4533, Jul. 2017.
- [31] A. Grünerbl, A. Muaremi, V. Osmani, G. Bahle, S. Ohler, G. Troster, O. Mayora, C. Haring, and P. Lukowicz, "Smartphone-based recognition of states and state changes in bipolar disorder patients," *IEEE J. Biomed. Health Inform.*, vol. 19, no. 1, pp. 140–148, Jan. 2015.
- [32] J. Hannink, T. Kautz, C. F. Pasluosta, K.-G. Gaßmann, J. Klucken, and B. M. Eskofier, "Sensor-based gait parameter extraction with deep convolutional neural networks," *IEEE J. Biomed. Health Inform.*, vol. 21, no. 1, pp. 85–93, Jan. 2017.
- [33] F. Ordóñez and D. Roggen, "Deep convolutional and LSTM recurrent neural networks for multimodal wearable activity recognition," *Sensors*, vol. 16, no. 1, p. 115, Jan. 2016.
- [34] F. Rustam, A. A. Reshi, I. Ashraf, A. Mehmood, S. Ullah, D. M. Khan, and G. S. Choi, "Sensor-based human activity recognition using deep stacked multilayered perceptron model," *IEEE Access*, vol. 8, pp. 218898–218910, 2020.
- [35] A. Jurek, C. Nugent, Y. Bi, and S. Wu, "Clustering-based ensemble learning for activity recognition in smart homes," *Sensors*, vol. 14, no. 7, pp. 12285–12304, Jul. 2014.
- [36] M. Woźniak, M. Graña, and E. Corchado, "A survey of multiple classifier systems as hybrid systems," *Inf. Fusion*, vol. 16, pp. 3–17, Mar. 2014.
- [37] Z. Chen, C. Jiang, and L. Xie, "A novel ensemble ELM for human activity recognition using smartphone sensors," *IEEE Trans. Ind. Informat.*, vol. 15, no. 5, pp. 2691–2699, May 2019.
- [38] A. K. Shukla, P. Singh, and M. Vardhan, "A hybrid framework for optimal feature subset selection," *J. Intell. Fuzzy Syst.*, vol. 36, no. 3, pp. 2247–2259, Mar. 2019.
- [39] L. Zhang, K. Mistry, C. P. Lim, and S. C. Neoh, "Feature selection using firefly optimization for classification and regression models," *Decis. Support Syst.*, vol. 106, pp. 64–85, Feb. 2018.
- [40] G.-B. Huang, Q.-Y. Zhu, and C.-K. Siew, "Extreme learning machine: Theory and applications," *Neurocomputing*, vol. 70, nos. 1–3, pp. 489–501, 2006.
- [41] M. Wang, Y. Wan, Z. Ye, and X. Lai, "Remote sensing image classification based on the optimal support vector machine and modified binary coded ant colony optimization algorithm," *Inf. Sci.*, vol. 402, pp. 50–68, Sep. 2017.
- [42] X. Zhang, Q. Zhang, M. Chen, Y. Sun, X. Qin, and H. Li, "A two-stage feature selection and intelligent fault diagnosis method for rotating machinery using hybrid filter and wrapper method," *Neurocomputing*, vol. 275, pp. 2426–2439, Jan. 2018.
- [43] J. Z. Li, Q. G. Zhang, K. Wang, and J. Y. Wang, "Optimal dissolved gas ratios selected by genetic algorithm for power transformer fault diagnosis based on support vector machine," *IEEE Trans. Dielectr. Electr. Insul.*, vol. 23, no. 2, pp. 1198–1206, Apr. 2016.



YIMING TIAN received the B.S. degree in automation from Hebei University of Technology, Tianjin, China, in 2012, the M.S. degree in control science and control engineering from the University of Science and Technology Liaoning, Anshan, China, in 2015, and the Ph.D. degree in the control science and control engineering from Hebei University of Technology, in 2020.

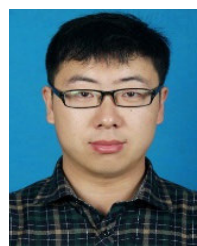
He is currently a Lecturer with the College of Information Engineering, Tianjin University of Commerce. His research interests include neural networks, wearable sensors, pattern recognition, and bioinformatics.



JIE ZHANG (Senior Member, IEEE) received the B.Sc. degree in control engineering from Hebei University of Technology, Tianjin, China, in 1986, and the Ph.D. degree in control engineering from the City, University of London, London, in 1991.

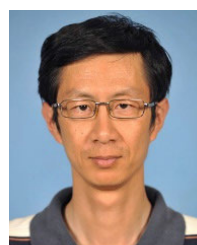
He is currently a Reader with the School of Engineering, Newcastle University, Newcastle upon Tyne, U.K. His research interests include neural networks, neuro-fuzzy systems, intelligent control systems, genetic algorithms, optimal control of batch processes, and multivariate statistical process control. He has published over 180 articles in international journals, books, and conferences.

Dr. Zhang is a member of the IEEE Control Systems Society and the IEEE Computational Intelligence Society. He served as a Reviewer for many prestigious international journals, including *IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS*, *IEEE TRANSACTIONS ON FUZZY SYSTEMS*, *Neural Networks*, *Automatica*, *Chemical Engineering Science*, and *AICHe Journal*.



LIPENG LI received the master's degree from Hebei University of Technology, in 2015.

He is currently an Experimenter with the College of Information Engineering, Tianjin University of Commerce. His research interests include the application of Internet of Things, pattern recognition, intelligent mobile robot, and artificial intelligence control.



ZUOJUN LIU received the M.S. degree in control science and control engineering from Hebei University of Technology, Tianjin, China, and the Ph.D. degree in control science and control engineering from Nankai University, Tianjin.

He is currently a Professor and a Ph.D. Supervisor with the School of Artificial Intelligence and Data Science, Hebei University of Technology. His research interests include rehabilitation aids, intelligent robots, and intelligent control theory and application.

• • •