

## Research Article

# MPD-Model: A Distributed Multipreference-Driven Data Fusion Model and Its Application in a WSNs-Based Healthcare Monitoring System

Jibing Gong,<sup>1,2,3</sup> Li Cui,<sup>1</sup> Kejiang Xiao,<sup>1,3</sup> and Rui Wang<sup>1</sup>

<sup>1</sup>Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100190, China

<sup>2</sup>Department of Computer Science and Engineering, Yanshan University, Qinhuangdao 066004, China

<sup>3</sup>Graduate University of Chinese Academy of Sciences, Beijing 100190, China

Correspondence should be addressed to Jibing Gong, gongjibing@gmail.com

Received 6 September 2012; Accepted 31 October 2012

Academic Editor: Fu Xiao

Copyright © 2012 Jibing Gong et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

We first propose an MPD-Model, a novel distributed multipreference-driven data fusion model for WSNs. Here, *preferences* are looked as the core elements of collaboration mechanism in a data fusion procedure. We then present MFA, a distributed multipreference feature-level fusion algorithm based on weighted average method. Next, to implement feature extraction of wrist-pulse data, we propose FEA, a light-weight adaptive feature extraction algorithm for time series sensed data. Simultaneously, we design TFD-Pattern that is a unique human pulse pattern. Based on historical data, we propose an SVM-based algorithm for health status detection tasks. Finally, we implement the proposed methods in a real wearable healthcare monitoring system which had been previously developed in-house. We validate the proposed methods using real-world data sets with 2046 pulse samples. Experimental results show that the proposed methods outperform the baseline methods, and the proposed MPD-Model is reasonable and effective.

## 1. Introduction

The rapid development of Wireless Sensor Networks (WSNs) brings some new situations. On the one hand, WSNs applications rapidly expand from traditional fields, for example, Military and National Defense and Environmental Monitoring, to emerging civil fields such as intelligent transportation, healthcare monitoring, and smart home. On the other hand, more mobile devices (e.g., smartphones) or local network systems (e.g., Body Sensor Networks) are frequently treated as one single node of WSNs. Undoubtedly, the situations will lead to form a large-scale and complex sensor network. We think the network is a typical Internet of Things (IoT) [1] and it has some characteristics such as interactivity and sociality. Simultaneously, It emphasizes intelligence and sensing-actuating ability of WSNs applications. Consequently, sensed data generated under the situations have the following features [1–4]. (1) The data has polymorphism and heterogeneity. (2) The data is massive.

(3) The data is real time. (4) Storage locations of the data have much diversity. (5) The data has high complexity and strong relationships.

In the new situations, users play more and more importance to intelligence of sensor network systems because of the growing actual demands. But it is very difficult to implement the intelligence due to seriously limited resources on nodes and the lack of historical sensed data. This is particularly true in the healthcare monitoring applications based on WSNs. Fortunately, system detection/identification accuracies and intelligence can be greatly improved if multisource data fusion technologies are fully leveraged. However, if so, some challenges still exist below. (1) How to select and evaluate those multisource sensing parameters to fuse? (2) How to establish novel data fusion model/mechanism to improve detection accuracy? (3) How to leverage limited resources on nodes to automatically recognize complex parameter patterns?

To address these challenges, we propose an MPD-Model, a novel-distributed multi-preference-driven data fusion model for WSNs. Based on traditional data fusion architecture, MPD-Model takes regard preference information as core elements of collaboration mechanism in a data fusion procedure. It introduces large-scale historical sensed data into machine learning for data fusion, treats the massive data as real-world test set, and thus effectively ensures enough identification accuracy and high intelligence of WSNs application systems. Here, *preference* is defined as a kind of description information being different from existing nature, accuracy, and detail level information of monitoring tasks in traditional data fusion method. It can reflect subjective wishes of individuals or a group. For example, the *preference*, Task Parameter Relevance, just is a kind of importance degree information telling that the accuracies of detection tasks depend on sensing parameters selected for fusion. It has two characteristics: uncertainty and dynamics. The former means that preference information varies with both the size and the type of WSNs applications. This well embodies individual data fusion. The latter indicates that preference information can be quantified and then evaluated by formalizing them.

We systematically address the problems of data fusion and health status detection in a healthcare monitoring system based on WSNs with the following contributions.

- (i) To address the basic problem of data fusion, we propose an MPD-Model, a novel-distributed multi-preference driven data fusion model for WSNs. Based on MPD-Model, this paper conducts the individual data fusion of multisource sensing parameters in a real-world healthcare monitoring system. MPD-Model can not only solve incompleteness, inaccuracy, and uncertainty problems of preference information, but also significantly improve detection accuracy and intelligence of WSNs applications. According to MPD-Model, this paper presents several fusion algorithms to finish practical health status detection tasks.
- (ii) In MPD-Model, we design and formalize four kinds of preference information (or impact factors). Through analyzing multi-preference information, we can quantify and evaluate importance degree of different multisource sensing parameters during data fusion. We propose MFA, a distributed and unified multi-preference fusion algorithm based on weighted average method for generating complex parameter patterns that will be further used for health status detection tasks.
- (iii) To implement MPD-Model in a real wearable health status detection system [1], we propose EFA, a light-weight adaptive feature extraction algorithm for pulse data, to obtain physiological features of human wrist pulse waveforms. According to these features, we design TFD-Pattern, a novel time series pattern, and make the pattern as the input of MFA algorithm. Considering massive historical pulse

data, this paper also presents SVM-based feature-decision-making algorithm to detect human health statuses, for example, subhealth. The algorithm is simultaneously an instance of  $y_i$  element in MPD-Model.

- (iv) We validate our proposed MPD-Model and fusion algorithms using large-scale real-world healthcare data sets. Experimental results show that FEA is effective and outperforms the existing derivative-based method at an over 16% feature extraction accuracy; an average fusion ratio of 98.73% is available when running MFA. SVM-based health status detection method well support intelligence of MPD-Model via outperforming (20~30%) AR and WPT methods. Different weighted values of multi-preference parameters perform distinct diagnosis performance. It proves that our individual fusion model based on preferences and weighted average method is reasonable and effective.

## 2. Multipreference-Driven Data Fusion Model for WSNs

**2.1. MPD-Model.** In this paper, we propose an MPD-Model, a novel-distributed multi-preference-driven data fusion model for Wireless Sensor Networks, as in Figure 1. Explanations of MPD-Model are given below. Its input may include environment monitoring information, human physiological parameters, and movements/activities information.  $R_i$  ( $1 \leq i \leq n$ ) indicates the output results of signals/data processing procedure  $P_i$  on the sensor node  $S_i$ .  $R_i \blacklozenge R_j$  ( $1 \leq i \neq j \leq n$ ) states the fusion operation between any two output results.  $D_k$  ( $1 \leq k \leq m$ ) means one motoring target state or decision-making as the output of the model.  $y_i$  stands for any method, algorithm, or model for recognition or decision-makings.

The model is suitable for all kinds of WSNs applications for human sensing, particularly in healthcare monitoring, due to the following flexible features. (1)  $P_i$ , a signal/data processing procedure, may be any kind of processing method or technology that can run on the sensor nodes, and they should be distributed, light-weight, or interactive. (2) Fusion operation, namely,  $R_i \blacklozenge R_j$ , may also be any kind of calculation, algorithm, or technology. (3) The decision-making  $y_i$  may be a complicated classification method (e.g., Support Vector Machine and Time-Space Factor Graph Model), a user-defined rule, or even a simplest Threshold Value Judgement Operation such as IF statement. (4) Values of  $D_k$  depend on different WSNs applications. It can identify one monitored target state with Yes or No when  $n = 2$ . Otherwise, multiple states of one target can be recognized and output.

MPD-Model views intelligence and historical sensed data as core elements and key features. This is reasonable and scientific because they are the development tendency of data fusion model. There are several reasons as follows. (1) The two features are strongly required by user demands of WSNs applications in the new situations. Another point to

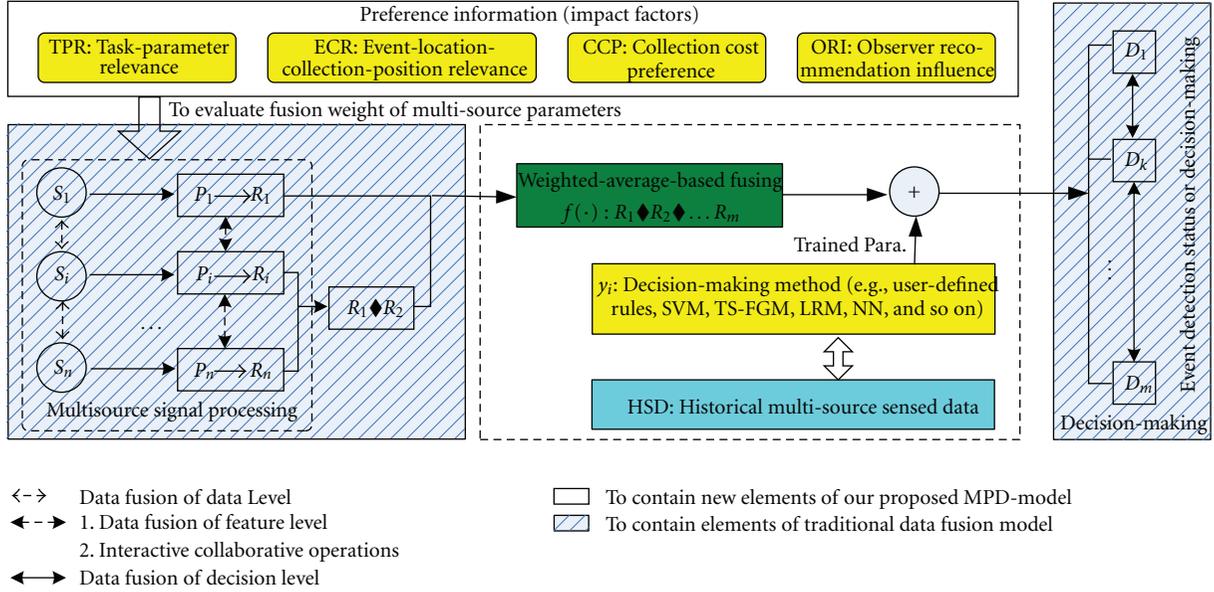


FIGURE 1: MPD-Model: a distributed multipreference driven data fusion model for Wireless Sensor Networks.

illustrate is that characteristics of sensed data generated in WSNs have dramatically varied. (2) Historical sensed data is paying more and more important role in a data fusion procedure. Thus, the role should be clearly embodied in a novel data fusion model, or fusion performance conducted by intelligent methods based on analyzing historical data will be seriously weakened. (3) With rapid expanding of WSNs applications from traditional domains to civil fields, they have dramatically increased demands on the intelligence. (4) Data processing ways depend on data types. But only data fusion model can effectively solve the requirement when considering the intelligence. This model should be driven by physical model of WSNs, and aiming at the intelligence of WSNs applications, this model should combine all kinds of data processing technologies and establish a unified architecture or a novel fusion mechanism to efficiently implement users' requirements.

## 2.2. Formalization of Preferences' Information/Impact Factors.

Influences on detection decision-making results of sensing parameters depend on specific WSNs applications. Thus, we propose an individual data fusion model for WSNs. The model first computes fusion weights of different impact factors according to one WSN application's own features, and then combines and fuses these factors using weighted average method. Finally, more reasonable and accurate event detection results can be obtained. In the model, we innovatively design four kinds of ubiquitous preferences (impact factors or constraints) for data fusion in WSNs. They can clearly describe uniqueness or features of every WSNs application. They are formalized and described below.

**2.2.1. Task-Parameter Relevance, TPR.** The impact factor measures the relevance degree between sensing parameters and task types of event detection. For instance, although

some parameters, for example, temperature, humidity, air composition, and so on, are often combined together to detecting fire event, temperature parameter will obviously have more influences on fire event detection than others. TPR is designed from the perspective of the event detection types and computed using the following formula:

$$t\_score(P, E) = \frac{\vec{q}_P \cdot \vec{q}_E}{\|\vec{q}_P\| \|\vec{q}_E\|}, \quad (1)$$

where  $\vec{q}_E$  is the vector of event types and  $\vec{q}_P$  is the vector of sensing parameters for detection or monitoring task.

### 2.2.2. Event-Location-Collection-Position Relevance (ECR).

The impact factor reflects the relevance degree between the event location and the sensing parameter collection position. As we all know, those parameters that are collected from event locations will have more influences on event detection performance. For example, in the fire detection, temperature parameters at fire location will have higher reference value than the ones at other locations. ECR is computed according to formula (2)

$$e\_score(P, E) = \begin{cases} \#P's \text{ es of } E's \text{ location} & \text{if } E's \text{ location} \in La, \\ 0 & \text{if } E's \text{ location} \notin La, \end{cases} \quad (2)$$

where es indicates event and  $La = \{l_1, l_2, \dots, l_m\}$  means the set of position where sensing parameters are collected from.

**2.2.3. Observer Recommendation Influence (ORI).** It indicates one parameter's importance degree in the eyes of observers or domain experts during event detection. It is

well known that different experts will emphasize different sensing parameters in one same event detection task. Thus, the preference, that experts assign to one sensing parameter, will be very the important information for fusion. For example, environmental experts can obtain more accurate detection results of Algae outbreak by combining water quality, PH value, and water temperature parameter together, and ordinary observers can directly do it by using the value of Algae content information. ORI is calculated using the following formulas:

$$o\_score(O, E) = \begin{cases} 1 - spw(O, E) & \text{if } spw(O, E) < 1, \\ 1 & \text{if } spw(O, E) \geq 1, \end{cases} \quad (3)$$

$$o\_score(P, E) = \frac{\sum_{P's \text{ observers}} o\_score(O, E)}{\#P's \text{ recommenders}},$$

where  $O$  denotes an observer or a domain expert,  $E$  an event, and  $spw$  means the weight of shortest path between  $O$  and  $E$ . The term “observers” is equivalent with “recommenders.”

**2.2.4. Collection Cost Preference (CCP).** It measures collection cost of sensing parameters selected in data fusion of WSNs applications in order to design more reasonable detection solution. For example, we can easily detect Algae outbreak event using Algae content sensors. However, for most environmental managers, it is impossible to do it by this way due to the technology difficulty or the high-cost solution. But according to suggestions of environmental experts, they can combine water quality, water temperature, water visibility, and PH value together to detect the same event. These kinds of information can be obtained using sensors or devices with high quality and low price. In this case, collection cost of sensing parameters has undoubtedly effect on the detection precision of Algae outbreak event.

Computing CCP needs to map all kinds of information from both sensing parameters and collecting sensors/devices into one same number interval so that comparisons might be made at one same order of magnitude. Therefore, Sigmoid function is used to map all inputs of real number field into the interval  $[0, 1]$ , and then rank fraction function is applied to map ranking information into the interval  $[0, 1]$ . They are defined as in (4), respectively,

$$\text{sigmoid} : S(t) = \frac{1}{1 + e^{-t}},$$

$$\text{rank\_score} : R(r, c) \quad (4)$$

$$= 1 - \log_{10} \left( 1 + 9 \times \frac{r-1}{c-1} \right), \quad c > 1,$$

where  $r$  indicates the rank of a sensing parameter or a device, and  $c$  indicates the number of the ranks. We first adjust input errors of devices and then carry out formatting and normalization processes. At last, combining sensing parameters' information, we compute  $t\_rank$ —the rank of technology maturity degree of sensors/devices,  $t\_count$ —the number of technology maturity degrees,  $p\_rank$ —the rank of popularity degree of devices,  $p\_count$ —the number of popularity degrees,  $d\_rank$ —the rank of difficulty grade/degree for

collecting sensing parameters,  $d\_count$ —the number of difficulty grades, and  $gpa\_score$ —the percentage transferred from the price of devices. Thus, we can compute collection cost score of a sensor/device for a parameter using formula (5), the final formula for computing CCP impact weight is given, as in (6):

$$c\_score(P, I) = S(R(I's \ t\_rank, t\_count) + (I's \ p\_rank, p\_count) - R(P's \ d\_rank, d\_count) - I's \ gpa\_score), \quad (5)$$

$$c\_score(P, E) = \frac{\sum_{P's \text{ devices for } E} c\_score(P, I)}{\#P's \text{ devices for } E}, \quad (6)$$

where  $d\_count = n$ ,  $t\_count = m$ ,  $p\_count = k$ ,  $P$  means a kind of sensing parameter,  $I$  indicates a sensor/device, and  $E$  means a monitoring task or an event.

**2.3. Historical Sensed Data in MPD-Model.** Though lots of WSN-based applications had employed machine learning methods to conduct recognition in WSNs [1, 2], to the best of our knowledge, few work definitely introduce Historical Sensed Data (or Historical Data), shortly HSD, as the core component of fusion model [6]. In our opinion, HSD should be involved in data fusion model design due to the following facts.

- (i) In WSNs, long-term continuous monitoring will inevitably produce massive data. The importance of these data is that it not only provides query services for users, but also lays a solid foundation of decision-making for future data mining.
- (ii) In WSNs, most parameters, such as pulse, ECG, and physical image, have complex data patterns. Accordingly, diverse machine learning methods are needed to train, learn, and classify these patterns. Therefore, HSD should be undoubtedly involved in data fusion procedure of WSNs systems.
- (iii) Parameter types often determine what decision-making methods should be employed to finish health status detection task. Specifically, in the case of unattended monitoring, only some simple data processing method can be employed.

In the case of human sensing, complex factors such as users' social relationships might be considered and accordingly call for more complicated data processing methods (e.g., Time-Space Factor Graph Model).

To illustrate the role of HSD in data fusion tasks, we conduct comparative diagnosis experiments based on AR [7] and WPT [8] methods to recognize the five kinds of human health statuses. SVM and user-defined rules are employed as decision-making methods. The comparison results, as in Figure 2, show that SVM-based methods with HSD often have higher recognition accuracies (30~40%) than user-defined rules without HSD.

### 3. Health Status Detection

Chinese Pulse Diagnosis Theory (CPDT) [9] has been used in clinical diagnosis for thousands of years and has been proved to be valuable. According to CPDT, wrist-pulse data contains the abundant physiological and pathological parameter information of human body. Pulse information has different patterns which can identify different kinds of people with a specific health status or symptom/disease. Thus, sensed pulse data can be leveraged to effectively detect human health status.

**3.1. TFD-Pattern: A Novel Time Series Pattern for Health Status Detection.** In this paper, a unique kind of pulse pattern is designed based on CPDT [9] and our previous researches on pulse diagnosis [1, 5]. We first compute 90 time-domain feature variables based the five feature points (see Figure 3). For example, the feature variable  $T_1/T$  can be obtained through leveraging the  $T$ -point and the  $P$ -point. Then we compute 36 frequency-domain feature variables based on the discrete Fourier transform formula, as in (7), and the discrete power spectral density function, as in (8). Finally we obtain 6 feature variables for pressure value. Each of them is the average value of all the real-time external pulse pressures in one pulse waveform data. Actually, these above computations are just suboperations of  $f(\cdot)$  in Figure 1:

$$X(k) = \sum_{n=0}^{N-1} x(n)e^{-j(2\pi/N)kn}, \quad (7)$$

$$\sum_{n=0}^{N-1} x^2(n) = \frac{1}{N} \sum_{k=0}^{N-1} |X(k)|^2. \quad (8)$$

We merge the three types of feature variables into one 132-dimensional pulse pattern that stands for a wrist-pulse waveform and call it time-frequency-domain Pattern, shortly TFD-Pattern, as in Figure 4.

**3.2. SVM-Based Feature-Decision-Making Algorithms for Health Status Detection.** Since the SVM performs well on problems with low training set sizes, we conduct SVM-based identification experiments on these pulse data sample groups with different human health statuses. Based on decision function of SVM theory, as in (9), we propose a health status detection algorithm which runs on the central server:

$$f(x) = \text{sgn} \left( \sum_{i=1}^k \alpha_i^* y_i K(E_i, x) + b^* \right), \quad (9)$$

where  $k$  is the number of pulse data samples;  $E$  is a set of feature vectors.  $K$  might be one of Linear, Quadratic, Polynomial, RBF, and MLP kernel functions.  $\alpha_i y_i$  is weight vector and  $\sum_{i=1}^n \alpha_i y_i = 0$ .

It needs to be noted that the algorithm was designed just according to our proposed MPD-Model and is corresponding to  $\alpha_i y_i$  in the model. Another characteristic of this algorithm is that it is based on our designed TFD-Pattern which is totally different from other pulse diagnosis (or health status detection) methods [7, 8].

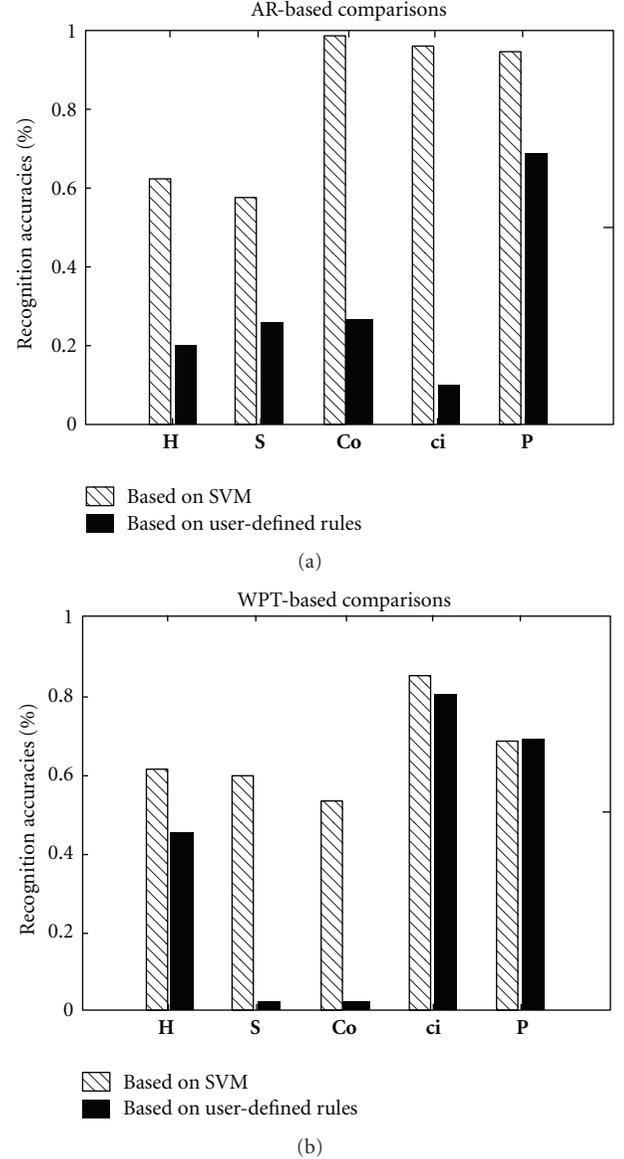


FIGURE 2: Contribution and importance of Historical Sensed Data (HSD) for data fusion (user-defined rules are designed according to expert's prior knowledge. At the  $x$ -axis, **H** shows the comparison results of identifying the Hypertensive from the Normal, **S**—Subhealthy versus Normal, **Co**—Coronary Heart disease versus Normal, **ci**—Cirrhosis disease versus Normal, and **P**—Pregnant versus Normal female.)

### 4. Distributed Multipreference-Driven Feature-Level Fusion Algorithms Based on Weighted Average Method

In terms of  $f(\cdot) : R_1 \diamond R_2 \diamond \dots \diamond R_m$  in MPD-Model, we present MFA, a novel distributed and unified multipreference fusion algorithm based on weighted average method. MFA first computes multi-preference parameter's impact weights of detecting health statuses using probabilities of four impact factors mentioned above, as in (10), and

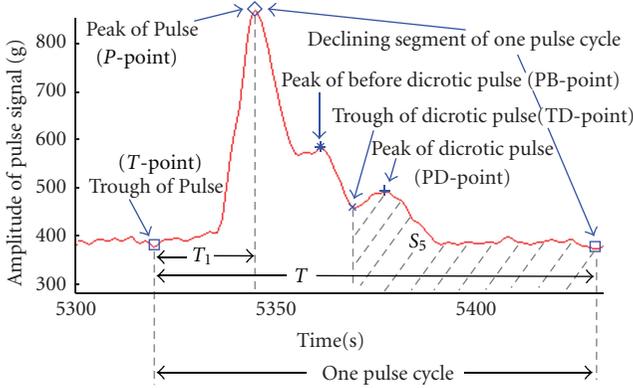


FIGURE 3: The five most important physiological feature points of human pulse waveform data [5].

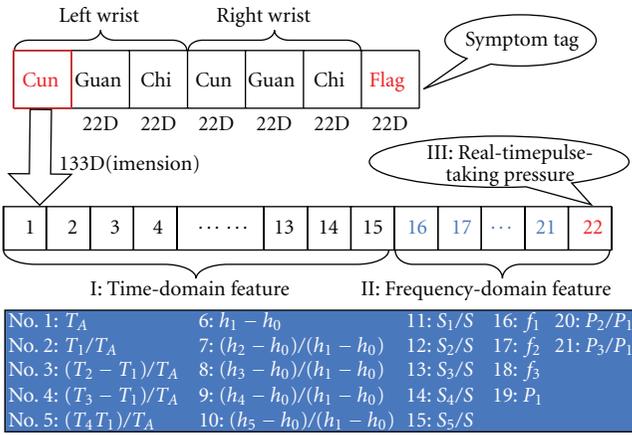


FIGURE 4: Illustration of our designed TFD-Pattern.

then fuses all physiological parameters to generate parameter patterns for further health status detection decision-making, as in (11):

$$\begin{aligned}
 & \text{impact\_weight}(P, D) \\
 &= (t\_score(P, D) \times t\_weight \\
 &+ m\_score(P, D) \times m\_weight \\
 &+ e\_score(P, D) \times e\_weight \\
 &+ c\_score(P, D) \times c\_weight) / \\
 & (t\_weight + m\_weight + e\_weight + c\_weight), \\
 & \text{Fusion}(P_1, P_2, \dots, P_m) = \bigcup_{i=1}^m f(P_i) \diamond \text{impact\_weight}(P_i, D),
 \end{aligned} \tag{11}$$

where the initial values of all weights are set to 1 and  $\sum_{i=1}^m \text{impact\_weight}(P_i, D) = 1$ . With receiving feedbacks from users, the model would use feedback data to optimize these weight values.

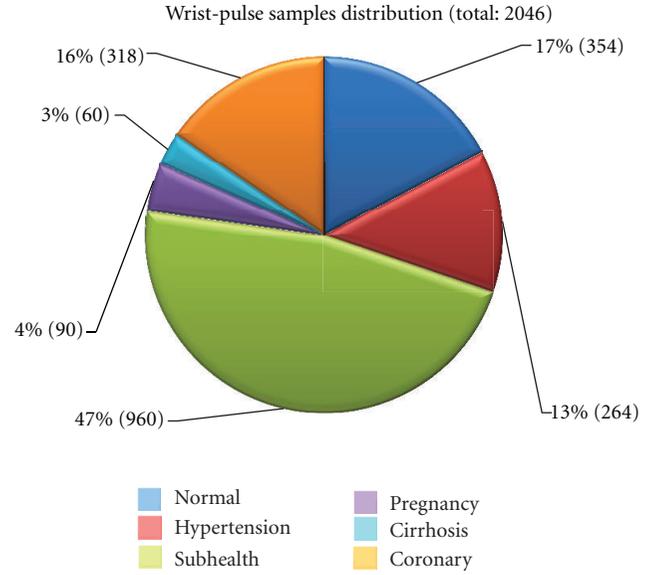


FIGURE 5: Wrist-pulse samples distribution in the real-world health care data sets.

## 5. Experimental Results

**5.1. Experimental Setup.** In our previous work, we had developed a wearable healthcare monitoring sensor network system [1] for health status detection. This system helps us to collect the real-world health care data set in a large scale clinic experiments which happened at the Institute of Computing Technology (ICT), Chinese Academy of Sciences (CAS) in 2009 and 2010 Hi-tech Fair of China in Shenzhen City, respectively. Wrist-pulse data samples distribution is shown in Figure 5.

We validate the effectiveness of the proposed methods comparing with the following baseline methods.

**5.1.1. Derivative-Based Feature Extraction Method [11].** In this method, an automated derivative-based time-domain feature extraction method on wrist pulse waveform is proposed to extract magnitude-type feature  $h_i$  ( $i = 1$  to 5). Here, the set of  $h_i$  is equal to the five feature points (see Figure 3) that are employed in our work.

**5.1.2. AR Pulse Diagnosis Method [7].** In this health status detection method, an autoregressive- (AR-) based method is proposed to extract the pulse signal features. The mean and variance of the prediction error are calculated and selected as features. The selected features are then taken as inputs to a support vector machine (SVM) for diseases classification.

**5.1.3. WPT Pulse Diagnosis Method [8].** This health status detection made pulse waveform feature extraction based on optimal Wavelet Packet Transform (WPT). Subband energies contained in the best basis were extracted as features and SVM classifiers were trained, differentiating cholecystitis patients and nephritic syndrome patients from normal people.

We evaluate our method from the following four angels.

- (1) *Precision*. We evaluate the proposed method in terms of Extraction Precision and Diagnosis Precision comparing to corresponding baseline methods.
- (2) *Fusion Ratio*. It states the ratio of the fused file size versus the original file size. The former file is used to save multi-preference parameter data. Fusion ratio is computed by the formula:  $(1 - \text{fused size/original size}) * 100\%$ .
- (3) *CPU Time*. It indicates how long learning procedures would take for different SVM\\_-based methods. It is employed to measure real time of these comparative methods.
- (4) *Significant Difference*. It can tell users what feature variables in TFD-Pattern can much help in revealing the human symptom condition, and then, it would prove that our proposed method is effective if any valid feature variable with significant difference exists.

**5.2. Feature Extraction.** According to CPDT [9], we must first extract the five most important time-domain feature points (see Figure 3) before running MFA algorithm in our health status detection. Thus, we propose FEA, an adaptive feature extraction algorithm for pulse data. It can be considered as an instance of  $P_i \rightarrow R_i$  ( $1 \leq i \leq n$ ) in MPD-Model. Specifically, we formalize the slope variability as a trend decision function, as in (12). Then, the final desirable feature points can be selected from the candidate ones according to some adaptive parameters. The most important adaptive parameter Threshold\_of\_PULSE is used to find the PB-point, the TD-point, and the PD-point, and it is computed by (13). Another adaptive parameter Amp\_pulse indicates the mean value of all pulse cycles in one pulse waveform. It can be computed by (14):

$$f(i) = \begin{cases} 0 & \text{Other,} \\ 1 & s(i) < 0, s(i+1) < 0, \\ 2 & s(i) > 0, s(i+1) < 0, \\ -2 & s(i) < 0, s(i+1) > 0, \end{cases} \quad (12)$$

$$\text{Threshold\_of\_PULSE} = \frac{(1/Fs + m)}{\text{Amp\_pulse}} \quad (13)$$

$$\text{Amp\_pulse} = \frac{1}{\text{Peak\_Num}} \sum_{k=1}^{\text{Peak\_Num}-1} (T\text{-point}(k+1) - T\text{-point}(k)). \quad (14)$$

**5.2.1. Effectiveness of Adaptive Parameters.** To illustrate the effectiveness of Threshold\_of\_PULSE, we run FEA algorithm on normal pulse data with 50 Hz, 100 Hz, and 1000 Hz sampling frequencies, respectively. The results are shown in Figure 6. From this figure, we can see that all the beacon features points can be accurately extracted.

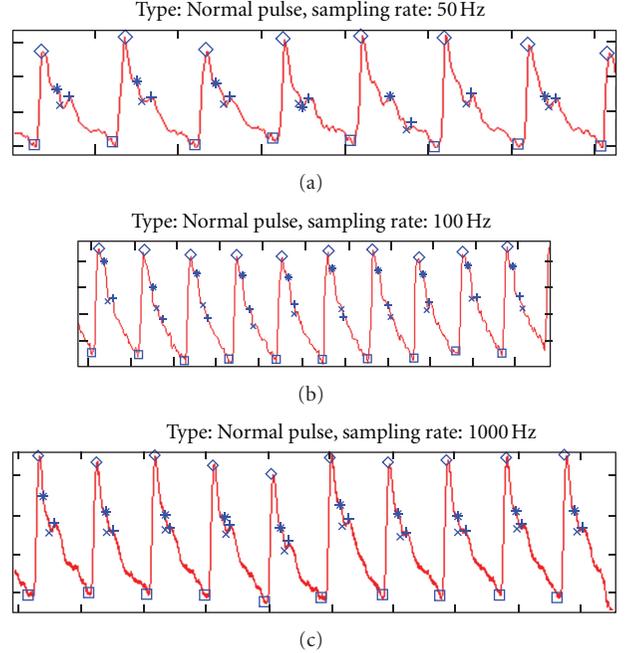


FIGURE 6: Effectiveness of the adaptive parameter Threshold\_of\_PULSE.

Similarly, according to our prior knowledge, we may set  $\text{Amp\_pulse} = T/2$  when the irregular slippery pulse is processed, and  $\text{Amp\_pulse} = T/3$  when the irregular normal pulse is dealt with.  $T$  indicates the length value of pulse waveform cycle. The results in Figure 7 indicate that FEA algorithm also well performs on different wrist-pulse conditions.

**5.2.2. Performance Evaluation of Feature Extraction.** We compare the proposed FEA algorithm with a typical derivative-based method [11] via extracting these five feature points on 354 healthy people's wrist-pulse samples. Table 1 shows comparison results of extraction accuracies of 10 random examples.

From Table 1, FEA outperforms (+16%) the derivative-based method due to the following. (1) Though the derivative-based method is also light-weight, it is vulnerable to heavy noise from human body's activities. (2) The FEA not only is light-weight and provides great support to real-time pulse signal processing tasks, but also makes full use of many adaptive parameters based on automatic computing or prior knowledge.

**5.3. Fusion Effectiveness.** In our experiments, different types of data at each processing stage would be saved into the respective files. Considering the existing information fusion estimator, for example, sample-based estimator using measurements [12], we calculate data fusion\reduction ratios by comparing file sizes before and after processing sensed pulse data. Reduction degree of these files size directly shows the performance of MFA algorithm. Table 2 displays MFA's data fusion ratios. In this table, the first column indicates

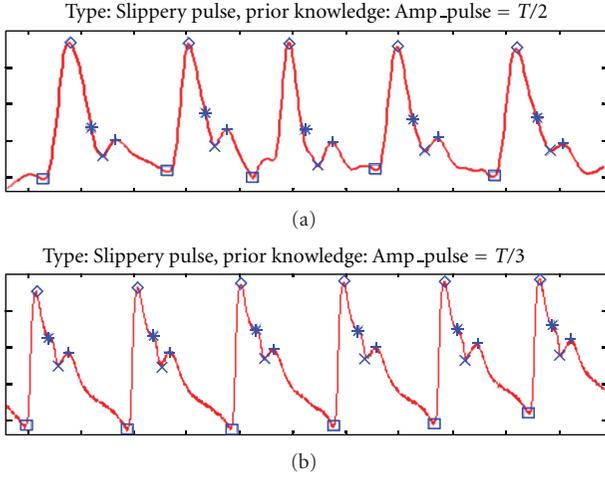


FIGURE 7: Effectiveness of prior knowledge: Amp\_pulse.

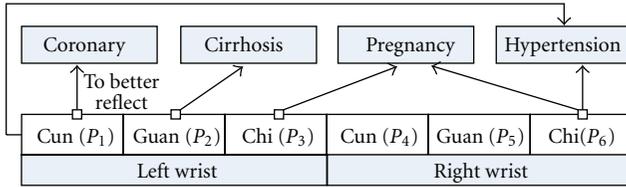


FIGURE 8: Wrist-pulse from diverse sources has different effects on revealing health status conditions according to Traditional Chinese Medicine [10].

TABLE 1: Comparison results of FEA versus the derivative-based method.

Number	ID	Derivative-based method	FEA
...	...	...	...
26	A19_709A	71%	86%
27	A20_809C	70%	84%
28	A22_737E	67%	90%
29	A37_719E	72%	82%
30	A66_524G	69%	89%
31	B20_735E	74%	89%
32	B9_0724C	64%	84%
33	B10_723E	70%	78%
34	B18_706K	73%	89%
35	B19_706G	70%	88%
...	...	...	...
Average		70%	86%

the number of one wrist-pulse ample, the second column stands for the sizes of these files before running MFA algorithm, the third column shows the sizes of these files after running MFA algorithm, and the final one gives data fusion ratios. The results show MFA algorithm has excellent fusion effect for wrist-pulse data.

**5.4. Health Status Detection Performance.** We make evaluations from the following aspects.

TABLE 2: Data fusion ratios of MFA algorithm.

Data file name	Original size (KB)	MFA	Ratios
...	...	..	...
NORMAL0613B_4.01	1491	14	99.06%
hBP_0612A_0.01	658	6	99.08%
hBP_AO0709B_1.01	192	3	98.44%
NORMAL0612B_0.01	271	5	98.15%
NORMAL0709A_5.02	858	8	99.07%
hBP_AO0709_1.01	119	2	98.32%
NORMAL0715A_1.01	1058	11	98.96%
...	...	..	...
Average			98.73%

TABLE 3: Different weights of multi-preference parameters for health status detection.

	$P_1$	$P_2$	$P_3$	$P_4$	$P_5$	$P_6$
Weights solution-I of impact factors						
H	0.5	0	0	0	0	0.5
Co	1	0	0	0	0	0
Ci	1	0	0	0	0	0
P	0	0	0.5	0	0	0.5
Weights solution-II of impact factors						
H	0	0.25	0.25	0.25	0.25	0
Co	0	0.2	0.2	0.2	0.2	0
Ci	0	0.2	0.2	0.2	0.2	0
P	0.25	0.25	0	0.25	0.25	0
Weights solution-III of impact factors						
H	0.3	0.1	0.1	0.1	0.1	0.3
Co	0.7	0.06	0.06	0.06	0.06	0.06
Ci	0.8	0.04	0.04	0.04	0.04	0.04
P	0.02	0.02	0.4	0.02	0.02	0.4

TABLE 4: Feature variables with significant difference.

Feature variable number	Left/right wrist	Cun, Guan, or Chi	Feature variable name
Number 22	Left	Cun	Real-time external pulse pressure (EPP)
No. 59	Left	Chi	$S_5/S$

**5.4.1. Impact Factor Weight Validation.** According to Chinese Pulse Diagnosis Theory (CPDT) [9, 10], we investigate that what wrist-pulse parameters from diverse sources (i.e., Cun, Guan and Chi on left and right human wrists [13]) have great impacts on revealing disease/symptom information. Based on this investigation (see Figure 8), we further compute different weight values that are listed in the 2–6 columns of Table 3. According to formula (1), we detect four kinds of health statuses or diseases/symptoms: hypertension (H), coronary (Co), cirrhosis (Ci), and pregnancy (P). These weights are selected and grouped into three weight solutions shown in Table 3. According to these weight solutions we conduct comparative experiments and give the results shown

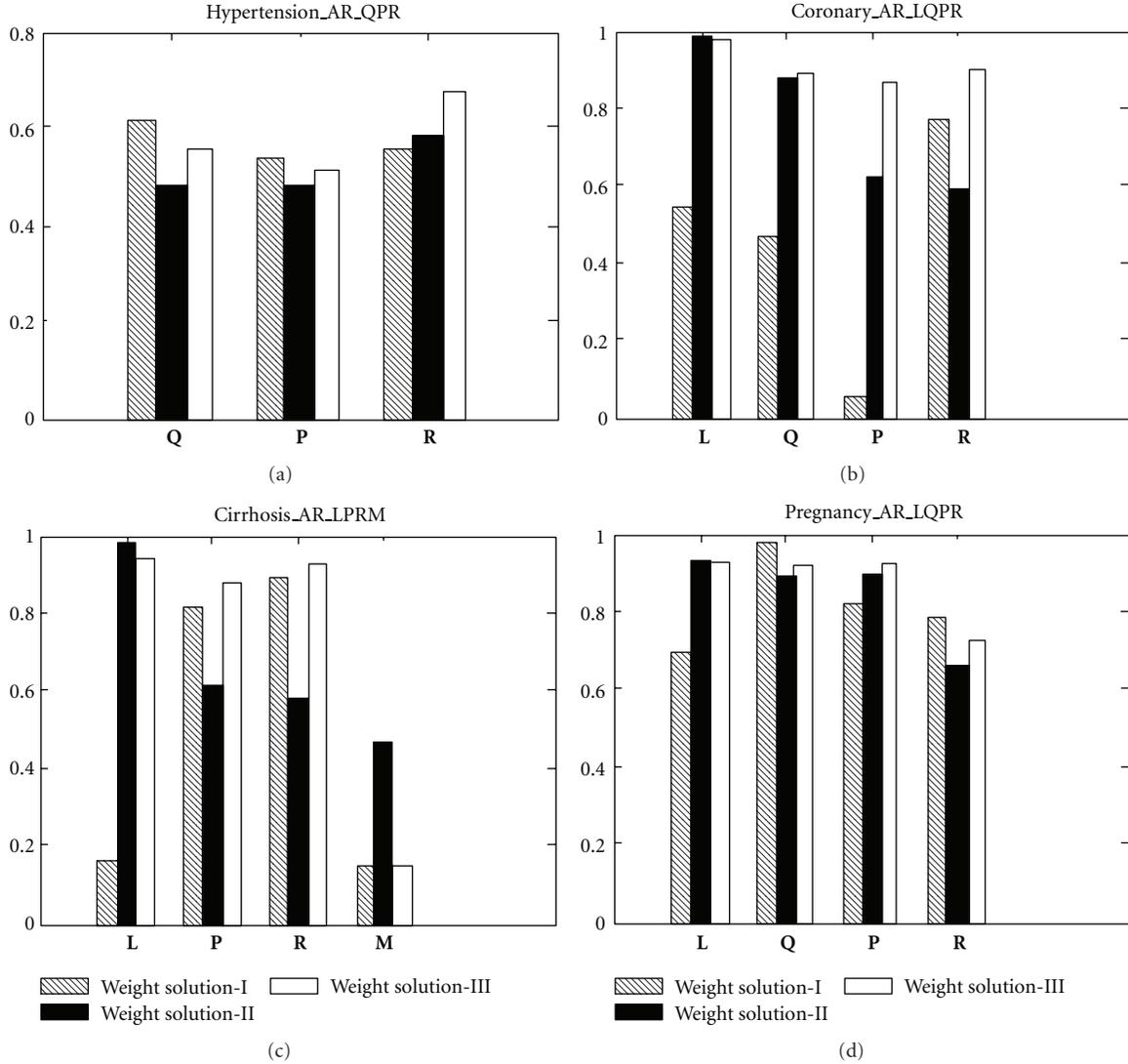


FIGURE 9: AR-based comparison results of different weights solutions for multi-preferences parameters (characters L, Q, P, R, and M stand for Linear, Quadratic, Polynomial, RBE, and MLP kernel function of SVM, respectively. The title of every subfigure tells which health status had recognized).

in Figures 9 and 10. From the two figures, we can see that the greater weight values of related impact factor ( $P_i$ ) become the higher health status detection precisions have. This proves that our multi-preferences-driven data fusion-based impact factor evaluation can well work.

**5.4.2. Comparative Evaluation.** In this paper, we compare with the baseline methods, AR [7] and WPT [8] methods, to validate the effectiveness of the proposed MPD-Model. The two methods and ours all focus on feature extraction of wrist-pulse data and are based on SVM to distinguish patients with different diseases from normal people. The difference is that distinct wrist-pulse pattern is employed to recognize. Thus, we can get solid comparison results shown in Figure 11. These comparative experiments are

conducted on the real healthcare data sets mentioned above (see Figure 6).

CPU time of ours, AR, and WPT health status detection method is approximately 17.3121 s, 164.7448 s and 42.0289 s, respectively. Among all comparative experiments, our SVM-based health status detection method takes the least running time and thus it is proved to be light-weight and more suitable to fit wearable healthcare monitoring system based on WSNs. As for hypertension, coronary, cirrhosis, and pregnancy symptoms, our method outperforms (average 15~30%) AR and WPT methods. Thus, our method is more effective to detect human health status. However, identification accuracy of our method is a little lower than the two methods when detecting subhealth status. There are two reasons to explain it: (1) low accuracies of feature extraction for pulse waveform data will cause that it is hard

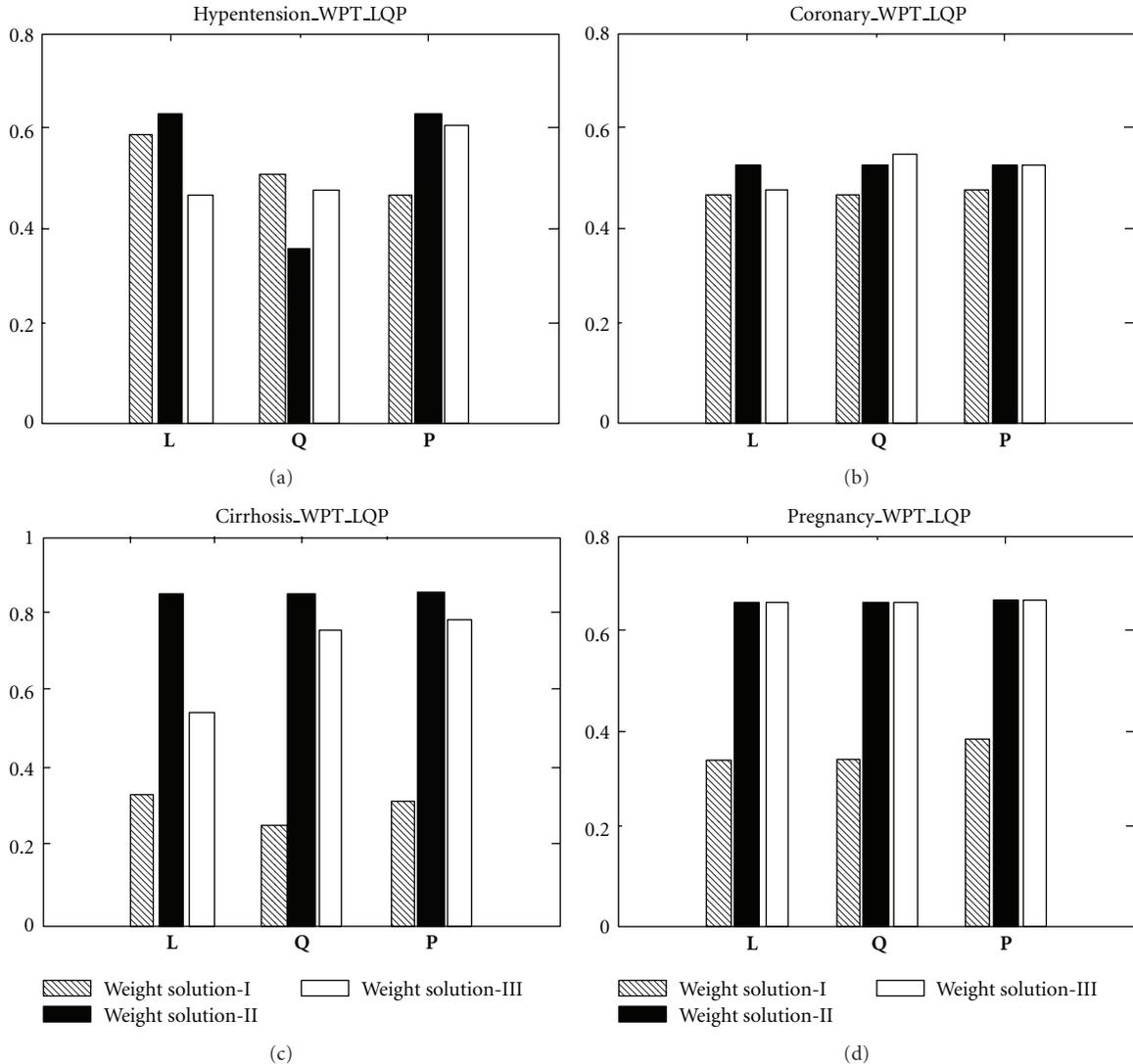


FIGURE 10: WPT-based comparison results of different weight solutions for multipreference-parameters (explanations of this figure have been given in Figure 9).

to clearly distinguish some pulse patterns from others. But AR and WPT methods have no such problem just because algorithms of computing feature vector are totally different from ours. (2) Generally, it is very difficult to identify people with subhealth status from the normal ones since such patients often have not clear disease characteristics.

**5.4.3. Effectiveness Evaluation.** At last, feature variables with significant difference in the two groups, as in Table 4, are obtained by conducting  $T$  statistical analysis method.

As Figure 3 shows, feature variable  $S_5/S$  indicates the ratio of TD-point– $T$ -point area to the whole area of one pulse cycle. It reveals clear physiological meaning. That is, patient's reflected wave of blood flow in one heartbeat cycle are stronger than the healthy. This is may be one reason that blood pressure values of a hypertension people are always high. Another feature variable, real-time External

Pulse Pressure (EPP), indicates that real-time blood pressure values recorded by pulse-sensor nodes have distinct effects on the patients. It is more reasonable than  $S_5/S$  to explain such the pathological phenomenon that patient's blood pressure is higher than healthy people's.

## 6. Related Work

WSN-based healthcare solutions become a hot research topic. It has been well developed through the research work in activity recognition [14], physiological data gathering [15], pattern recognition [16], and so forth. Though lots of existing work collect blood pulse data for a healthcare-monitoring purpose, few work study systematically how to fuse parameters from diverse sources based on impact factors.

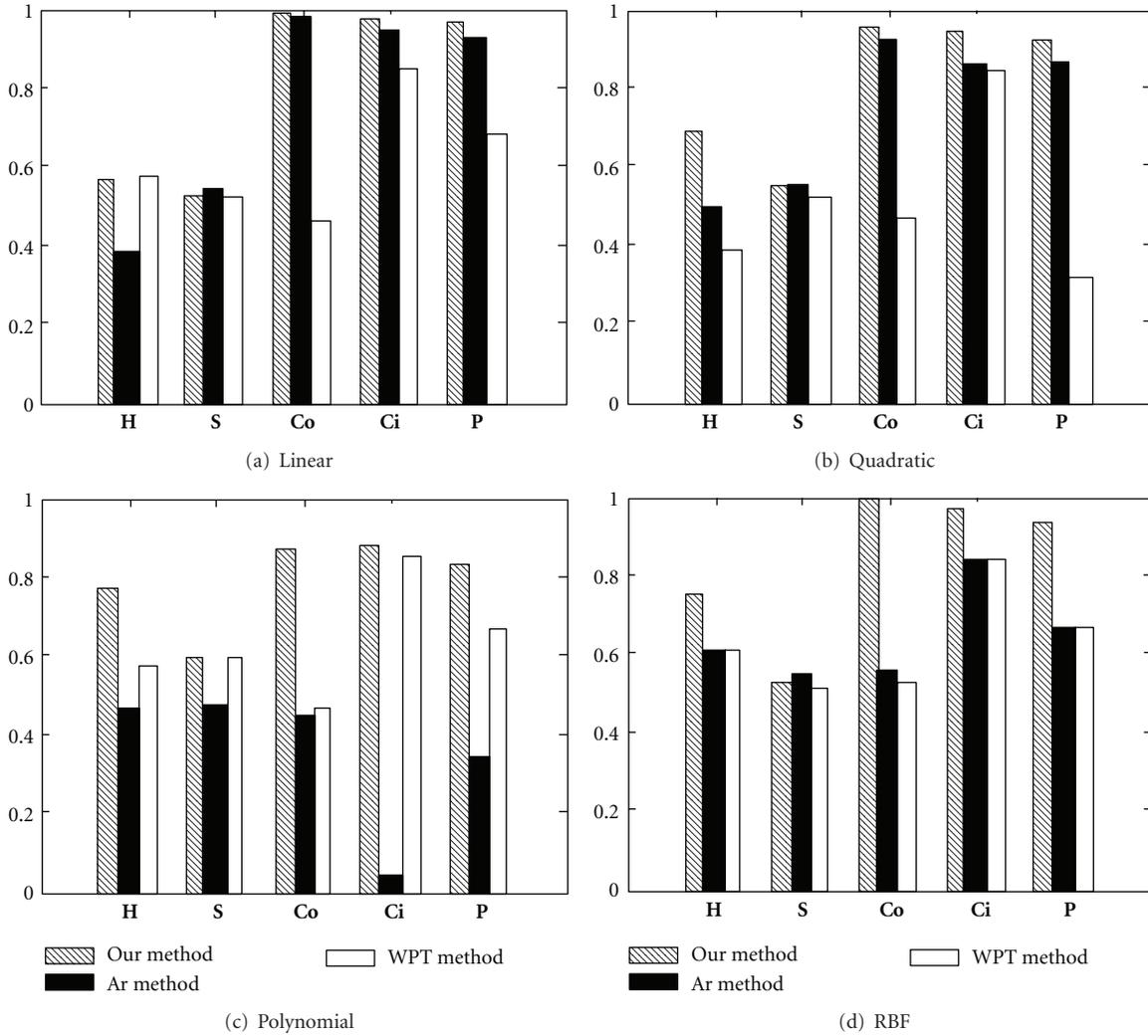


FIGURE 11: Comparison with AR and WPT methods (Part (a), (b), (c) and (d) show SVM-based diagnosis experiments with Linear, Quadratic, Polynomial and RBF kernel function, respectively. The y-axis indicates Precision (%) of health status or disease/symptom identification, and the explanations of the x-axis has already been given in Figure 2.).

The five most important feature points (see Figure 3) considered as the basis of the objectification of pulse study by most researchers [17]. Currently, most extraction methods of time domain features are derivative based [11]. This kind of method is easy to be implemented on sensor nodes, but is just suitable to regular pulse waveform with high Signal-to-Noise Ratio (SNR). To address the problem of limited resources, feature extraction in our work focuses on sensed pulse data with low SNR. The proposed algorithm is lightweight and adaptive. Much work focus on pulse detection and classification. For example, [18] proposed a Lempel-Ziv complexity analysis-based approach to classify seven pulse patterns that exhibit different rhythms. Besides, Chinese Pulse Diagnosis Theory (CPDT) is also objectified in order to pulse acquisition, feature extraction of pulse, and pulse diagnosis [17]. As a powerful method for pattern classification with low training set sizes, SVM is often leveraged

to diagnose pulse signals/patterns that are computerized using Modified Auto-Regressive Models [7] or spatial and spectrum features [19].

Though there already exist some data fusion methods, for example, Crash Fault Correction [20], Privacy Protection [21], position-based Aggregator node election in WSNs [22, 23], and environmental Monitoring [24], they are not suitable for Wearable Sensor Networks due to complicated sensed data patterns and high intelligence requirements [6]. Reference [25] also focus on multisource fusion model, but employing Chi-square distribution with  $n$  degrees of freedom method instead of weighted average method. Reference [12] discusses information fusion estimation problems and provides useful estimation methods for our proposed data fusion model. A continuous classification of multisource sensed data in [26, 27] inspired us to introduce historical data into designing multisource-driven data fusion model.

## 7. Conclusion

In this paper, we try to systematically investigate the problems of data fusion and health status detection and propose MPD-Model, a novel-distributed multi-preference-driven data fusion model for WSNs. Based on MPD-Model, we present MFA algorithm (an instance of  $f(\cdot)$  in MPD-Model), EFA algorithm ( $P_i$ ), and SVM-based health status detection method ( $y_i$ ). We also design TFD-Pattern, a novel time serials pattern according to CPDT. It is a kind of human pulse pattern as well as is both EFA's output and MFA's input. Experimental results show that they all outperform baseline methods with the feature extraction accuracy of +86% and the system diagnosis accuracy of +75%, and our proposed MPD-Model is reasonable and effective.

## Acknowledgments

This paper is supported in part by the National Basic Research Program of China (973 Program) under Grant no. 2011CB302803, the "Strategic Priority Research Program" of the Chinese Academy of Sciences under Grant no. XDA060307000, the National Natural Science Foundation of China under Grant no. 61003292, and the Key Project of Natural Science Foundation of Jiangsu Province under Grant no. BK201100.

## References

- [1] J. J. Zhang, R. Wang, S. L. Lu et al., "EASiCare: design and implementation of a portable chinese pulse-wave retrieval system," in *Proceedings of the 9th ACM Conference on Embedded Networked Sensor Systems, (SenSys '11)*, pp. 149–161, Seattle, Wash, USA, 2011.
- [2] M. Keally, G. Zhou, G. L. Xing et al., "PBN: towards practical activity recognition using smartphone-based body sensor networks," in *Proceedings of the 9th ACM Conference on Embedded Networked Sensor Systems, (SenSys '11)*, pp. 246–259, Seattle, Wash, USA, 2011.
- [3] T. Park, J. Lee, I. Hwang, C. Yoo, L. Nachman, and J. Song, "Demo: e-gesture—a collaborative architecture for energy-efficient gesture recognition with hand-worn sensor and mobile devices," in *Proceedings of the 9th ACM Conference on Embedded Networked Sensor Systems, (SenSys '11)*, pp. 359–360, Seattle, WA, USA, 2011.
- [4] E. Ertin, N. Stohs, S. Kumar et al., "AutoSense: unobtrusively wearable sensor suite for inferring the onset, causality, and consequences of stress in the field," in *Proceedings of the 9th ACM Conference on Embedded Networked Sensor Systems, (SenSys '11)*, pp. 274–287, Seattle, Wash, USA, 2011.
- [5] J. B. Gong, S. L. Lu, R. Wang, and L. Cui, "PDhms: pulse diagnosis via wearable healthcare sensor network," in *Proceedings of the IEEE International Conference on Communications, (ICC '11)*, Kyoto, Japan, 2011.
- [6] E. F. Nakamura, A. A. F. Loureiro, and A. C. Frery, "Information fusion for wireless sensor networks: methods, models, and classifications," *ACM Computing Surveys*, vol. 39, no. 3, Article ID 1267073, pp. A9/1–A9/55, 2007.
- [7] Y. Chen, L. Zhang, D. Zhang, and D. Y. Zhang, "Computerized wrist pulse signal diagnosis using modified auto-regressive models," *Journal of Medical Systems*, vol. 35, no. 3, pp. 321–328, 2011.
- [8] Q. L. Guo, K. Q. Wang, D. Y. Zhang, and N. M. Li, "A wavelet packet based pulse waveform analysis for cholecystitis and nephrotic syndrome diagnosis," in *Proceedings of the International Conference on Wavelet Analysis and Pattern Recognition, (ICWAPR '08)*, pp. 513–517, August 2008.
- [9] L. Xu, D. Zhang, and W. Kong, "Wavelet-based Cascaded Adaptive Filter for Removing Baseline Drift in Pulse Waveforms," *IEEE Transaction on Biomedical Engineering*, vol. 52, no. 11, pp. 1973–1975, 2005.
- [10] Z. F. Fei, *Contemporary Sphygmology in Traditional Chinese Medicine*, People's Medical Publishing House, Beijing, China, 2003.
- [11] C. Xia, Y. Li, J. Yan et al., "Wrist pulse waveform feature extraction and dimension reduction with feature variability analysis," in *Proceedings of the 2nd International Conference on Bioinformatics and Biomedical Engineering, (ICBBE '08)*, pp. 2048–2051, May 2008.
- [12] R. N. Madan and N. S. V. Rao, "A perspective on information fusion problems," *International Journal of Distributed Sensor Networks*, vol. 5, no. 1, 4 pages, 2009.
- [13] Y. Chen, L. Zhang, D. Zhang, and D. Y. Zhang, "Wrist pulse signal diagnosis using modified Gaussian models and Fuzzy C-Means classification," *Medical Engineering and Physics*, vol. 31, no. 10, pp. 1283–1289, 2009.
- [14] T. Gu, Z. Wu, X. Tao, H. K. Pung, and J. Lu, "epSICAR: an emerging patterns based approach to sequential, interleaved and concurrent activity recognition," in *Proceedings of the 7th Annual IEEE International Conference on Pervasive Computing and Communications, (PerCom '09)*, pp. 1–9, Galveston, Tex, USA, March 2009.
- [15] R. Cristescu, B. Beferull-Lozano, and M. Vetterli, "On network correlated data gathering," in *Proceedings of the 23rd Annual Joint Conference of the IEEE Computer and Communications Societies, (INFOCOM '04)*, pp. 2571–2582, March 2004.
- [16] J. M. Shapiro, "Embedded image coding using zerotrees of wavelet coefficients," *IEEE Transactions on Signal Processing*, vol. 41, no. 12, pp. 3445–3462, 1993.
- [17] L. Xu, K. Wang, D. Zhang et al., "Objectifying researches on traditionalchinese pulse diagnosis," *Informatica Medica Slovenica*, pp. 56–63, 2003.
- [18] L. Xu, D. Zhang, K. Wang, and L. Wang, "Arrhythmic pulses detection using Lempel-Ziv complexity analysis," *EURASIP Journal on Applied Signal Processing*, vol. 2006, Article ID 018268, 12 pages, 2006.
- [19] D. Y. Zhang, W. Zuo, D. Zhang et al., "Wrist blood flow signal-based computerized pulse diagnosis using spatial and spectrum features," *Journal of Biomedical Science and Engineering*, vol. 4, no. 3, pp. 361–366, 2010.
- [20] B. Balasubramanian and V. K. Garg, "Fused data structures for handling multiple faults in distributed systems," in *Proceedings of the International Conference on Distributed Computing Systems, (ICDCS '11)*, pp. 677–688, 2011.
- [21] W. B. He, X. Liu, H. Nguyen, and K. Nahrstedt, "A cluster-based protocol to enforce integrity and preserve privacy in data aggregation," in *Proceedings of the 29th IEEE International Conference on Distributed Computing Systems Workshops, (ICDCS '09)*, pp. 14–19, 2009.
- [22] L. Buttyán and P. Schaffer, "Position-based aggregator node election in wireless sensor networks," *International Journal of Distributed Sensor Networks*, vol. 2010, Article ID 679205, 15 pages, 2010.

- [23] A. Mpitziopoulos, D. Gavalas, C. Konstantopoulos, and G. Pantziou, "CBID: a scalable method for distributed data aggregation in WSNs," *International Journal of Distributed Sensor Networks*, vol. 2010, Article ID 206517, 13 pages, 2010.
- [24] S. Dauwe, T. Van Renterghem, D. Botteldooren, and B. Dhoedt, "Multiagent-based data fusion in environmental monitoring networks," *International Journal of Distributed Sensor Networks*, vol. 2012, Article ID 324935, 15 pages, 2012.
- [25] J. A. Benediktsson and I. Kanellopoulos, "Classification of multisource and hyperspectral data based on decision fusion," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 37, no. 3, pp. 1367–1377, 1999.
- [26] G. Xing, J. Wang, K. Shen, Q. Huang, X. Jia, and H. C. So, "Mobility-assisted spatiotemporal detection in wireless sensor networks," in *Proceedings of the 28th International Conference on Distributed Computing Systems, (ICDCS '08)*, pp. 103–110, July 2008.
- [27] B. Waske and J. A. Benediktsson, "Fusion of support vector machines for classification of multisensor data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 45, no. 12, pp. 3858–3866, 2007.



Hindawi

Submit your manuscripts at  
<http://www.hindawi.com>

