Combination of Heterogeneous Features for Wrist Pulse Blood Flow Signal Diagnosis via Multiple Kernel Learning

Lei Liu, Wangmeng Zuo, Member, IEEE, David Zhang, Fellow, IEEE, Naimin Li, and Hongzhi Zhang

Abstract—Wrist pulse signal is of great importance in the analysis of the health status and pathologic changes of a person. A number of feature extraction methods have been proposed to extract linear and nonlinear, and time and frequency features of wrist pulse signal. These features are heterogeneous in nature and are likely to contain complementary information, which highlights the need for the integration of heterogeneous features for pulse classification and diagnosis. In this paper, we propose a novel effective method to classify the wrist pulse blood flow signals by using the multiple kernel learning (MKL) algorithm to combine multiple types of features. In the proposed method, seven types of features are first extracted from the wrist pulse blood flow signals using the state-of-the-art pulse feature extraction methods, and are then fed to an efficient MKL method, SimpleMKL, to combine heterogeneous features for more effective classification. Experimental results show that the proposed method is promising in integrating multiple types of pulse features to further enhance the classification performance.

Index Terms—Feature extraction, multiple kernel learning (MKL), pulse diagnosis, wrist pulse blood flow signal.

I. INTRODUCTION

Wrist pulse signal contains rich physiological and pathologic information and is of great importance in the analysis of the health status and pathologic changes of a person. In traditional Chinese medicine (TCM), for thousands of years the practitioners use the fingers to feel the pulse fluctuations as a measure of the pulse signal, and then analyze the health condition of the person. Studies on modern medicine also show that pulse signal can also be used as signs of several cardiac diseases, such as ventricular tachycardia or atrial fibrillation.

Traditional pulse diagnosis method, however, suffers from several intrinsic limitations. First, pulse diagnosis is a skill that requires considerable training and experience. Second, the results sincerely depend on the subjective analysis of the practitioner and sometimes may be unreliable and inconsistent. To overcome these limitations, computational pulse signal diagnosis techniques have been recently studied to acquire quantitative pulse signal using various types of sensors and to obtain objective and consistent analysis results using signal processing and pattern recognition approaches [1]–[4].

Generally speaking, computerized pulse signal diagnosis involves the following three modules: data acquisition, feature extraction, and pattern classification. In the first module, pulse signals are first acquired using a pressure, photoelectric, or a Doppler ultrasound sensor, and then preprocessed for denoising, baseline wander removal [5], [6], and segmentation [7].

In the feature extraction module, a number of feature extraction methods have been proposed to extract the features from the pulse signals, which can be roughly grouped into two categories according to the usage of the frequency transform or not. In the first category, fiducial points usually are first required to be detected [8]. Spatial features, elastic similarity measures, autoregressive [9], and Gaussian mixture models [10] are then adopted to derive appropriate feature description of pulse signal. In the second category, pulse signal is first transformed to the transform domain using Fourier wavelet [11] or Hilbert–Huang transform (HHT) [8]. Energy or other statistical features are then extracted for pulse feature representation.

In the classification module, various classifiers have been developed for different classification tasks, e.g., diagnosis of diseases, pulse waveform classification, and analysis of health conditions. Several classifiers, e.g., Bayesian networks classifier [12], support vector machine (SVM) [13], and artificial neural network [14], have been adopted for pulse signal classification.

Although the feature extraction and classification methods are heterogeneous in nature, different feature representations may reflect different aspects of pulse signal, and are likely to contain complementary information for pulse diagnosis. Thus, appropriate combination of these heterogeneous features would benefit the classification performance. Moreover, it is also interesting to investigate the redundancy of the features, and what types of features would contribute more to pulse diagnosis.

In this paper, using multiple kernel learning (MKL), we proposed a framework for integrating heterogeneous pulse features to enhance the classification accuracy of pulse diagnosis. As shown in Fig. 1, the proposed framework consists of three major stages. First, we extract seven types of pulse features using the following feature extraction methods developed by our research...
group: fiducial point-based spatial features (FP), auto-regressive model (AR), time warp edit distance (TWED), Hilbert–Huang transform (HHT), approximate entropy (ApEn), wavelet packet transform (WPT), and wavelet transform (WT). Second, we design suitable kernel function (basis kernel) for each type of features i.e., we adopt Gaussian time warp edit distance kernel [15] for a single-period pulse signal, and adopt a Gaussian RBF kernel for the other types of features. Finally, an MKL algorithm, SimpleMKL [16], is used to integrate the heterogeneous features by simultaneously learning an optimal linear combination of the basis kernels and a kernel classifier. SimpleMKL [16] is an efficient algorithm recently developed for solving the MKL problem. Compared with other MKL algorithms [17], [18], SimpleMKL [16] is more efficient, especially for large-scale problems with many data points and multiple kernels. By combining the heterogeneous features of pulse signal using SimpleMKL, we expect that such a framework would further enhance the pulse signal classification accuracy.

Recent studies have verified the effectiveness of Doppler ultrasonic blood flow signal for computerized pulse diagnosis [9], [10], [19]. In this paper, we first evaluate the proposed method using our Doppler wrist blood flow signal dataset. Naturally, the proposed method can be directly extended to the classification of pressure or photoelectric pulse signals and other pulse classification tasks. To verify this, using the pressure pulse signals, we further test the proposed method for pulse waveform classification.

The reminder of this paper is organized as follows. Section II describes the seven feature extraction and matching methods for pulse signals. Section III introduces the SimpleMKL algorithm for learning the linear combination of basis kernels and the kernel classifier. Section IV provides the experimental results. Finally, Section V ends this paper with a few concluding remarks.

II. PULSE SIGNAL FEATURE EXTRACTION

Before pulse signal feature extraction, a preprocessing step usually is required for denoising, baseline wander removal, and segmentation. As shown in Fig. 2(a), the raw data acquired by the Doppler ultrasonic device are in the form of Doppler spectrogram, where the envelope corresponds to the blood flow velocity signal. In the preprocessing step, we first detect the maximum velocity envelope of the Doppler spectrogram to extract wrist pulse blood flow signal. Using the 7-level “db6” wavelet transform, we remove the baseline wander by suppressing the seventh-level “db6” wavelet approximation coefficients, and reduce the first-level wavelet detail coefficients for denoising. Fig. 2(b) shows an example of the wrist pulse blood flow signal after drift wander and noise removal. Considering that pulse signals are quasi-periodic signals, we finally adopt an automatic method to locate the onset of each period, as shown in Fig. 2(b).

Pulse signal feature extraction approaches can be roughly grouped into two categories: nontransform-based and transform-based methods. In this section, we describe seven feature extraction methods, which involve three nontransform-based methods, i.e., fiducial point-based features, AR model, and time series matching, and four transform-based methods, i.e., HHT, ApEn, WPT, and WT.

A. Nontransform-Based Feature Extraction

In this section, we introduce three nontransform-based pulse signal feature extraction methods, i.e., fiducial point-based features, AR model, and time series matching.

1) Fiducial Point-Based Features: As shown in Fig. 3, there are several kinds of fiducial points in pulse signal, i.e., onset, peak of primary wave, dicrotic notch, and peak of secondary wave [7], [8]. By far, spatial features have been extracted based on the location and amplitude of the fiducial points, and the shape between fiducial points [8], [20], [21]. As a summary, Table I lists ten fiducial point-based features. For the meaning of each feature, refer to the description in Table I and the illustration in Fig. 3.
et al. introduces a parameter \( \lambda \) to penalize the \( \varepsilon_g(t) \) and \( \varepsilon_f(t) \) terms is then given by
\[
\varepsilon_g(t) = g(t) - \sum_{i=1}^{p} a_i g(t - i).
\]
We further calculate two features, the mean and the standard deviation of \( \varepsilon_g(t) \), to characterize the diagnostic parameters of the pulse signal.

3) Time Series Matching: Wrist pulse signals actually are time series data, and thus, time series matching algorithms can be used for pulse signal classification. To date, various time series matching methods have been developed, e.g., edit distance, dynamic time warping [22], edit distance with real penalty [23], and TWED [24]. In [25], Liu et al. applied TWED for the diagnosis of wrist blood flow signals. TWED satisfies the triangular inequality property and is an elastic metric for time series matching. Let \( A \) and \( B \) be two time series \( A = \{a_1', \ldots, a_m'\} \) and \( B = \{b_1', \ldots, b_m'\} \) with \( a_i' = (a_i, t_{a_i}) \) and \( b_i' = (b_i, t_{b_i}) \), where \( a_i \) and \( t_{a_i} \) is the ith sample value and the corresponding time label. TWED introduces a stiffness parameter \( \nu \) to control the elasticity of the distance between the elements \( a_i' \) and \( b_i' \), \( d(a_i', b_i') = d(a_i, b_i) + \nu \cdot \left( |t_{a_i} - t_{b_i}| \right) \), and introduces a parameter \( \lambda \) to penalize the delete operation. Then the TWED \( \delta_{\lambda,\nu} \) of \( A \) and \( B \) is recursively defined as
\[
\delta_{\lambda,\nu}(A^p_i, B^q_i) = \min \left\{ \begin{array}{l}
\delta_{\lambda,\nu}(A^{p-1}_i, B^q_i) + \Gamma(a'_p \to \Lambda) \quad \text{delete}_A \\
\delta_{\lambda,\nu}(A^{p-1}_i, B^{q-1}_i) + \Gamma(a'_p \to b'_q) \quad \text{match} \\
\delta_{\lambda,\nu}(A^{p-1}_i, B^{q-1}_i) + \Gamma(\Lambda \to b'_q) \quad \text{delete}_B
\end{array} \right.
\]
with
\[
\Gamma(a'_p \to \Lambda) = d(a_p, a_{p-1}) + \nu \cdot (t_{a_p} - t_{a_{p-1}}) + \lambda
\]
\[
\Gamma(a'_p \to b'_q) = d(a_p, b_q) + d(a_{p-1}, b_{q-1}) + \nu \cdot (|t_{a_p} - t_{b_q}|)
\]
\[
\Gamma(\Lambda \to b'_q) = d(b_q, b_{q-1}) + \nu \cdot (t_{b_q} - t_{b_{q-1}}) + \lambda
\]
where \( \delta_t \) denotes a time subseries with discrete time index varying between 1 and \( p \). The TWED \( \delta_{\lambda,\nu} \) can be calculated efficiently using dynamical programming. To further improve

2) AR Model: In [9], Chen et al. proposed an AR model-based feature extraction method. In the training stage, an AR model is constructed from the reference signal obtained by pulse signals from healthy persons. During feature extraction, given a test wrist pulse signal, one can compute the residual error to the AR model, and then further calculate the mean and the standard deviation of the residual error as features of the test signal. According to [1], the residual error feature is disease sensitive and is promising in distinguishing healthy persons from patients with specific diseases.

In the following, we describe the procedure of the AR model-based method in more detail. First, each pulse signal \( f(t) \) in the training set is normalized by
\[
\hat{f}(t) = \frac{f(t) - m_f}{\delta_f}
\]
where \( m_f \) and \( \delta_f \) are the mean and the standard deviation of the original signal \( f(t) \), respectively. The reference signal \( \hat{f}(t) \) is defined as the average of all the normalized pulse signals of healthy persons in the training set.

In a AR model, the current observation \( f(t) \) can be predicted as a linear function of the previous observations, \( f(t-1), f(t-2), \ldots, f(t-p) \). Given the reference signal \( \hat{f}(t) \), the AR model with \( p \) terms is then given by
\[
f(t) = \sum_{i=1}^{p} a_i f(t-i) + \varepsilon_f(t)
\]
where \( a_i \) is the ith AR coefficient and \( \varepsilon_f(t) \) denotes the modeling error. Using the reference signal \( \hat{f}(t) \), all the model parameters can be estimated, where the coefficients \( a_i \) are obtained using the least-squares method and the order \( p \) is determined by the Akaike information criteria.

For a given wrist pulse signal of a person with unknown healthy status \( g(t) \), its residual error can be calculated as
\[
\varepsilon_g(t) = g(t) - \sum_{i=1}^{p} a_i g(t - i).
\]
the efficiency, one can incorporate the Sakoe–Chiba band to prune the paths required to be considered [25].

B. Transform-Based Feature Extraction

In this section, we introduce four transform-based pulse signal feature extraction methods, i.e., HHT, ApEn, WPT, and WT.

1) HHT: In [11], Zhang et al. proposed an HHT-based method for wrist blood flow signal feature extraction. Let \( g(t) \) be a pulse signal. First, empirical mode decomposition EMD [26] is used to decompose \( g(t) \) into a series of intrinsic mode functions (IMFs), \( \text{IMF}_n(t) \), and a residue \( r(t) \). For simplicity, the residue \( r(t) \) is treated as the last IMF, resulting in

\[
g(t) = \sum_{n=1}^{N} \text{IMF}_n(t)
\]

where \( N \) is the number of IMFs. Then, Hilbert transform [26] of \( \text{IMF}_n(t) \) is defined as

\[
Y_n(t) = \frac{1}{\pi} P \int_{-\infty}^{\infty} \frac{\text{IMF}_n(\tau)}{t - \tau} d\tau
\]

where \( P \) denotes the Cauchy principal value [26]. Using \( \text{IMF}_n(t) \) and \( Y_n(t) \), one can define a complex analytic signal \( Z_n(t) \) as

\[
Z_n(t) = \text{IMF}_n(t) + iY_n(t) = a_n(t)e^{i\phi_n(t)}
\]

where \( a_n(t) \) and \( \phi_n(t) \) are defined as:

\[
a_n(t) = \sqrt{(\text{IMF}_n(t))^2 + (Y_n(t))^2}
\]

\[
\phi_n(t) = \arctan \left( \frac{Y_n(t)}{\text{IMF}_n(t)} \right)
\]

are the instantaneous amplitude and phase of \( Z_n(t) \), respectively. Furthermore, the frequency \( f_n(t) \) of \( Z_n(t) \) is defined as

\[
f_n(t) = \frac{1}{2\pi} \frac{d\phi_n(t)}{dt}.
\]

Finally, given \( a_n(t) \), \( f_n(t) \), and \( \text{IMF}_n(t) \), we calculate three kinds of features, i.e., the average amplitude \( \bar{a}_n \), the average frequency \( \bar{f}_n \), and the energy \( E_n \):

\[
\bar{a}_n = \frac{\sum_{t=1}^{m} a_n(t)}{m}
\]

\[
\bar{f}_n = \frac{\sum_{t=1}^{m} a_n(t)f_n(t)}{\sum_{t=1}^{m} a_n(t)}
\]

\[
P_n = \frac{\sum_{t=1}^{m} |\text{IMF}_n(t)|^2}{\sqrt{\sum_{n=1}^{m} \sum_{t=1}^{m} |\text{IMF}_n(t)|^2}}
\]

where \( m \) is the length of \( g(t) \). However, the number of IMFs \( N \) differs for different signals. To obtain features with fixed dimension, we empirically observe \( N \geq 5 \) and thus only use the first five lower order IMFs to derive a 15-D feature vector.

2) ApEn: It is a measure of the complexity and predictability of a time series [27], and thus can be used to describe the nonlinear characteristics of pulse signal [28]. We choose the pattern length \( m = 2 \) and the measure of similarity \( r = 25\% \). By treating a pulse signal as a time series \( A = [a_1, a_2, \ldots, a_n] \), a pattern \( P_m(i) \) is defined as a subsequence \([a_i, \ldots, a_{i+m-1}] \) of length \( m \) beginning at location \( i \). Two patterns \( P_m(i) \) and \( P_m(j) \) are similar if the following equation holds:

\[
|a_{i+k} - a_{j+k}| \leq r, \quad k = 0, \ldots, m - 1.
\]

Let \( n_{i,m}(r) \) denote the number of patterns of length \( m \) which are similar to \( P_m(i) \). We then define ApEn as

\[
\text{ApEn}(m, r, A) = \phi^{m+1}(r) - \phi^m(r)
\]

where

\[
\phi^m(r) = \frac{1}{n-m+1} \sum_{i=1}^{n-m+1} \ln C_i^m(r)
\]

\[
C_i^m(r) = \frac{n_{i,m}(r)}{n-m+1}.
\]

Different from [28], we calculate ApEn in the transform domain, where EMD is first used to decompose a pulse signal into seven IMFs. ApEn is then calculated for each IMF to derive a 5-D feature vector of the pulse signal.

3) WPT: In [11], pulse signal \( f(t) \) is decomposed as follows:

\[
\begin{align*}
\psi_0(t) &= f(t) \\
\psi_1^{i=1}(t) &= \sum_k H(k-2t)\psi_1^{i-1}(t) \\
\psi_1^{i=2}(t) &= \sum_k G(k-2t)\psi_1^{i-2}(t)
\end{align*}
\]

where \( i = 1, 2, \ldots, 2^j \), \( H(k) \) and \( G(k) \) are the low-pass and high-pass filters, respectively, and \( \psi_1^{i}(t) \) is the coefficient of the decomposed subband. Here, we choose the “db3” wavelet and the decomposition level of 5. Using the Shannon entropy criterion, we obtain the optimal WT decomposition tree together with the corresponding coefficients of the pulse signal. For each subband of the optimal decomposition tree, the energy of coefficients is computed as follows:

\[
E_j^i = \sum_n |p_j^i(t)|^2.
\]

Then we use energies as the features of the pulse signal.

4) WT: According to [11], we use WT to decompose pulse signal and extract the energy feature of each subband. Using the 7-level “db6” WT, we decompose the pulse signal into seven subbands, one coarse subband, and seven detailed subbands. Finally, we define the energies of coefficients as follows:

\[
\begin{align*}
E_{cA_7} &= \sum_{k=1}^{L_{cA_7}} cA_7^2(k) \\
E_{cD_i} &= \sum_{k=1}^{L_{cD_i}} cD_i^2(k), \quad i = 1, \ldots, 7
\end{align*}
\]

where \( cA_7 \) and \( cD_i \) are the coarse and the \( i \)th detailed wavelet subbands, respectively, and \( L_{cA7} \) and \( L_{cDi} \) denote the length of \( cA7 \) and \( cDi \), respectively.

III. PULSE SIGNAL CLASSIFICATION BASED ON MKL

In this section, we proposed an MKL framework to integrate heterogeneous features extracted from the pulse signal. First, we choose suitable kernel function for each feature extraction
or matching method, resulting in kernel-based representation of features or matching methods. Second, we use a recently proposed MKL algorithm, SimpleMKL [16], to learn a linear combination of kernels together with an SVM classifier to integrate heterogeneous features for enhanced classification accuracy.

A. Kernel Functions

Normalization usually is required to address the difference in distribution of heterogeneous features extracted by different methods. For the TWED method, we normalize each time series to have zero mean with standard deviation of 1. For the features extracted by other methods, each feature is normalized to have zero mean with standard deviation of 1.

Kernel functions are used to implicitly embedding features or matching methods into a high- or indefinite-dimensional feature space [29]. By far, kernel classifiers, e.g., SVM, have been widely adopted in many classification applications [30]–[32]. For different feature extraction methods, the feature vectors vary in feature dimension. Moreover, TWED should be regarded as a matcher rather than a feature extraction method. Thus, the design of appropriate kernel function [33] for each feature extraction or matching method is essential for building kernel classifier.

For different features extracted from pulse signal, we adopt two kinds of kernel functions. By referring to [15], we can represent TWED by means of a Gaussian TWED kernel function

\[
K_{GTWED}(A, B) = \exp \left( -\frac{\delta^2 A, B}{2\sigma^2} \right)
\]

(23)

where \( A \) and \( B \) are two time series, and \( \sigma \) is the kernel parameter. Besides, the Gaussian RBF kernel is adopted for the features extracted by other feature extraction methods

\[
K_{GRBF}(x, y) = \exp \left( -\frac{\|x - y\|^2}{2\sigma^2} \right)
\]

(24)

where \( x \) and \( y \) are two feature vectors, and \( \sigma \) is the kernel parameter.

As a summary, Table II lists the dimension of features extracted by each method. If we use one kernel function to represent the features extracted by each method, we have seven kernel functions. Except TWED, the number of features extracted by the other methods is 55 in total. More aggressively, we can construct four kernels for each of these 55 features with four different value of \( \sigma \). To automatically select the kernel parameter, for each feature, we construct four Gaussian RBF kernel functions with \( \sigma = 10, 15, 20, \) and 25. For TWED, we construct seven Gaussian TWED kernel functions with \( \sigma = 5, 10, 15, 20, 25, 30, \) and 35. Finally, we use MKL to adaptively learn the optimal kernel parameter or linear combination. Taking both Gaussian RBF kernels and Gaussian TWED kernels into account, we construct 227 basis kernels in total.

B. SimpleMKL

Compared with other MKL algorithms, e.g., the quadratically constrained quadratical programming method by Lanckriet et al. [29] and the semi-infinite linear programming method by Sonnenburg et al. [18], SimpleMKL [16] is much more efficient for large-scale classification problems with many data points and multiple kernels. Thus, we adopt SimpleMKL to integrate the heterogeneous features extracted from pulse signal.

Given \( M \) basis kernels \( K_m(x, y) (m = 1, \ldots, M) \), MKL intends to learn an optimal combination of basis kernels

\[
K(x, y) = \sum_{m=1}^{M} d_m K_m(x, y), \text{ subject to } d_m \geq 0,
\]

(25)

\[
\sum_{m=1}^{M} d_m = 1
\]

(26)

where \( d_m \) denotes the weight of kernel \( K_m(x, y) \), \( l \) denote the \( i \)th support vector, \( l \) is the number of support vectors, and \( \alpha_i^* \) and \( b^* \) are coefficients of SVM. If the basis kernels \( K_m(x, y) \) satisfy the Mercer criterion, one can easily verify that \( K(x, y) \) also satisfies the Mercer criterion.

In MKL, the coefficients and weights can be simultaneously learned by solving the following convex optimization problem:

\[
\min_{\alpha, b, \xi, d} \frac{1}{2} \sum_i \alpha_i K(\cdot, x_i) + C \sum_i \xi_i \\
\text{s.t. } y_i \sum \alpha_i K(x, x_i) + y_i b \geq 1 - \xi_i \\
\xi_i \geq 0 \\
\sum_m d_m = 1, d_m \geq 0
\]

(27)

to enforce fast MKL learning. SimpleMKL adopts an equivalent constrained optimization of (27):

\[
\min_{d} J(d), \text{ s.t. } \sum_m d_m = 1, d_m \geq 0 \forall m
\]

(28)

where

\[
J(d) = \min_{\alpha, b, \xi, d} \frac{1}{2} \sum_i \alpha_i K(\cdot, x_i) + C \sum_i \xi_i \\
\text{s.t. } y_i \sum \alpha_i K(x, x_i) + y_i b \geq 1 - \xi_i \\
\xi_i \geq 0
\]

(29)
One can use a state-of-the-art SVM solver to solve the problem to obtain the optimal solutions of $\alpha_i^*$ and $b^*$. Actually, this problem can be solved more efficiently by using the “warm-start” strategy [16]. The function $J(d)$ can be rewritten as

$$ J(d) = -\frac{1}{2} \sum_{i,j} \alpha_i^* \alpha_j^* y_i y_j \sum_{m=1}^M d_m K_m(x_i, x_j) + \sum_i \alpha_i^* . $$

(30)

Using (30), one can obtain the partial derivative of $J(d)$ with respect to $d_m$:

$$ \frac{\partial J}{\partial d_m} = -\frac{1}{2} \sum_{i,j} \alpha_i^* \alpha_j^* y_i y_j K_m(x_i, x_j). $$

(31)

By taking into account the equality constraint $\sum_m d_m = 1$, the reduced gradient of $J(d)$ is represented as

$$ \begin{cases} 
[\nabla_{red} J]_m = \frac{\partial J}{\partial d_m} - \frac{\partial J}{\partial d_\mu}, & \forall m \neq \mu \\
[\nabla_{red} J]_\mu = \sum_{m \neq \mu} \frac{\partial J}{\partial d_m} - \frac{\partial J}{\partial d_\mu}, & \text{else}
\end{cases} $$

(32)

where $\mu$ is chosen to be the index of the largest component of vector $d$ for the sake of the numerical stability. Furthermore, to satisfy the non-negativity constraint $d_m \geq 0$, rather than directly use $-\nabla_{red} J$, the descent direction should be modified for updating $d$:

$$ D_m = \begin{cases} 
0, & \text{if } d_m = 0 \text{ and } \frac{\partial J}{\partial d_m} - \frac{\partial J}{\partial d_\mu} > 0 \\
-\frac{\partial J}{\partial d_m} + \frac{\partial J}{\partial d_\mu}, & \text{if } d_m > 0 \text{ and } m \neq \mu \\
\sum_{\nu \neq \mu, d_\nu > 0} \left( \frac{\partial J}{\partial d_\nu} - \frac{\partial J}{\partial d_\mu} \right), & \text{for } m = \mu.
\end{cases} $$

(33)

Given the descent direction $D_m$, SimpleMKL uses the greedy and line search methods to further enhance the efficiency. For more details on the implementation of SimpleMKL, refer to [16, Algorithm 1].

IV. EXPERIMENTAL RESULTS AND DISCUSSION

The proposed method is implemented in MATLAB. All the experiments are carried out on a computer with a Core 2 Quad Q6600 processor running at 2.40 GHz.

A. Classification Experimental of Wrist Blood Flow Signal

In this subsection, we evaluate the classification performance of the proposed MKL framework using our wrist blood flow signal dataset, and compare MKL with the individual classifiers and other classifier fusion approaches.

Using EDAN’s CBS 2000 Transcranial Doppler Flow Analyzer, we construct a wrist blood flow signal dataset of 302 samples. Specifically, the samples in dataset are grouped into four categories, which include 95 samples of healthy persons, 36 samples of patients with sugar diabetes (SD), 50 with nephropathy (N), and 121 with gastrointestinal diseases (GD). The healthy persons are chosen from the staff and students from the Harbin Institute of Technology who have been diagnosed as healthy persons in their yearly physical examination. The patients are collected from Harbin Binghua Hospital where the diseases are diagnosed by the doctors according to the clinical data. The sampling frequency of CBS 2000 is 110 Hz. For each subject, only the pulse signal of the left hand is acquired and we select a stable segment of 1200 points for subsequent feature extraction and classification. In addition, in our dataset, 205 persons are male. The age distribution and the number of subjects of each category are summarized in Table III.

In order to verify the effectiveness of the proposed MKL method, using the wrist blood flow signal dataset, we test the classification methods for classifying healthy persons and patients with three kinds of diseases. Specifically, we adopt the tenfold cross-validation method to evaluate the diagnostic accuracy of the proposed MKL method, where the proposed method achieves the classification accuracy of 66.89%.

We further compare the proposed method with SVM with individual feature extractor. For each of the seven feature extraction methods, we constructed an individual SVM classifier [13], resulting in seven SVM classifiers: SVM-FP, SVM-AR, SVM-HHT, SVM-ApEn, SVM-WPT, and SVM-WT; and then adopted the tenfold cross-validation method to assess the classification accuracy. Tables IV and V list the classification accuracy and false positive rate of different methods, respectively. From Tables IV and V, one can see that the proposed method could obtain much higher classification accuracy and lower false positive rate than any individual classifier, which verifies that the integration of the heterogeneous features would enhance the classification accuracy and reduce false positive rate. Besides, we adopt the McNemar test [36] to evaluate the statistical significance of the difference between the proposed method and SVM-TWED. The result shows that the statistic value of the McNemar test is 20.04, which indicates that the performance difference is statistically significant at $\alpha = 0.05$.

The results are confirmed and can be further verified by the comparison of the classification accuracy obtained using the proposed method and other conventional classification combination methods, e.g., major vote rule (MVR) and Bayes sum rule (BSR). From Tables IV and V, one can see that the proposed method is superior to these classification combination methods. Table VI lists the classification time of these methods. Again, the proposed method can achieve a proper balance between the classification accuracy and the computational cost, and is more computationally efficient than the conventional combination methods. Moreover, we also compare the proposed method

<table>
<thead>
<tr>
<th>Diseases</th>
<th>Age 0–20</th>
<th>21–40</th>
<th>41–60</th>
<th>&gt;61</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Healthy</td>
<td>2</td>
<td>89</td>
<td>4</td>
<td>0</td>
<td>95</td>
</tr>
<tr>
<td>SD</td>
<td>0</td>
<td>3</td>
<td>20</td>
<td>13</td>
<td>36</td>
</tr>
<tr>
<td>Nephropathy</td>
<td>10</td>
<td>20</td>
<td>9</td>
<td>11</td>
<td>50</td>
</tr>
<tr>
<td>GD</td>
<td>15</td>
<td>53</td>
<td>46</td>
<td>7</td>
<td>121</td>
</tr>
</tbody>
</table>

TABLE II

SAMPLE DISTRIBUTION OF OUR DATASET

TABLE III

DISTRIBUTION OF OUR DATASET
with another pulse signal classification method, nearest neighbor classifier using TWED (1NN-TWED) [24], and the proposed method can also achieve higher classification accuracy than 1NN-TWED.

The wrist blood flow signal dataset is class imbalanced, where the number of samples for different classes is different. Thus, we list the confusion matrices of the proposed method in Table VII. From Table VII, the proposed method can achieve comparable classification accuracy for each class.

Finally, we show the weight of each feature extraction method in Fig. 4. From Fig. 4, the weights of the ApEn and AR methods are zero in the final combined kernel. Specifically, most features have nonzero weights which only include a small number 33 of kernels in the combined kernel. Since only part of the basis kernels and feature extraction methods are needed in the classification stage, it is reasonable that the proposed method would be more computationally efficient than conventional combination methods.

### B. Other Pulse Classification Application

In this section, in order to verify this method we could extend to other pulse classification tasks using the pressure or photoelectric pulse signals; we test the proposed method for pulse waveform classification using the pressure pulse signal dataset [37]. Specifically, we selected 800 samples of five typical pulse patterns, moderate, smooth, taut, unsmooth, and hollow, where the number of samples of each pulse pattern is 160. The confusion matrix and the classification results are shown in Tables VIII and IX, respectively. One can see that the proposed method can achieve high accuracy for this pulse classification task.
V. CONCLUSION

In this paper, we propose an MKL framework for integrating heterogeneous features extracted from pulse signals. By designing appropriate kernel function for the features extracted by each feature extraction method, MKL provides a flexible way to combining information from different feature extraction and matching methods. We adopt six pulse feature extraction methods, i.e., HHT, ApEn, WPT, WT, AR, and fiducial point-based method, and one pulse matching method, i.e., TWED. In our MKL framework, we use 7 Gaussian TWED kernels for the representation of the TWED method and 220 Gaussian RBF kernels for the representation of features extracted by the other methods. Finally, we adopt the SimpleMKL algorithm to integrate the heterogeneous features of pulse signals.

We first evaluate the classification performance of the proposed MKL framework on our wrist blood flow signal dataset. Experimental results show that compared with other classification fusion approaches and individual SVM classifiers, the proposed MKL framework is very effective in enhancing the classification accuracy of pulse signal. To further verify the proposed method, we test the proposed method for pulse waveform classification. In our future work, we will collect more pulse signal samples, and incorporate MKL with cost-sensitive SVM [34, 38] to address the inherent imbalance problem in pulse signal classification.

REFERENCES

Lei Liu received the B.S. degree in electronic information engineering and the M.S. degree in control theory and control engineering from the Harbin Engineering University, Harbin, China, in 2005 and 2008, respectively. He is currently working toward the Ph.D. degree in the Department of Computing, Harbin Institute of Technology, Harbin.

From September 2009 to December 2009, he was a Research Assistant in the Department of Computing, Hong Kong Polytechnic University. His research interests include medical diagnosis, pattern recognition, machine learning, and bioinformatics.

Wangmeng Zuo (M’09) received the Ph.D. degree in computer application technology from the Harbin Institute of Technology, Harbin, China, in 2007.

From July 2004 to December 2004, from November 2005 to August 2006, and from July 2007 to February 2008, he was a Research Assistant in the Department of Computing, Hong Kong Polytechnic University. From August 2009 to February 2010, he was a Visiting Professor in Microsoft Research Asia. He is currently an Associate Professor in the School of Computer Science and Technology, Harbin Institute of Technology. His current research interests include sparse representation, biometrics, pattern recognition, and computer vision.

Dr. Zuo is an Associate Editor of the IET Biometrics.

David Zhang (F’09) received the Graduate degree in computer science from Peking University, Beijing, China. He received the M.Sc. degree in computer science and the Ph.D. degree from the Harbin Institute of Technology (HIT), Harbin, China, in 1982 and 1985, respectively, and also the Ph.D. degree in electrical and computer engineering from the University of Waterloo, Ontario, Canada, in 1994.

From 1986 to 1988, he was a Postdoctoral Fellow at Tsinghua University, Beijing, and then an Associate Professor at the Academia Sinica, Beijing. He is currently a Head in the Department of Computing, and a Chair Professor at the Hong Kong Polytechnic University, Kowloon, Hong Kong, where he is the Founding Director of the Biometrics Technology Centre (UGC/CRC) supported by the Hong Kong SAR Government in 1998. He was also a Visiting Chair Professor in Tsinghua University, and an Adjunct Professor in Peking University, Shanghai Jiao Tong University, HIT, and the University of Waterloo. He is the author of more than ten books and 200 journal papers. He is a Senior Research Fellow of Croucher.

Dr. Zhang is the Founder and Editor-in-Chief of International Journal of Image and Graphics, Book Editor of Springer International Series on Biometrics, an Organizer of the International Conference on Biometrics Authentication, and an Associate Editor of more than ten international journals, including several IEEE Transactions and Pattern Recognition. He is a Distinguished Speaker of the IEEE Computer Society and a Fellow of the International Association of Pattern Recognition.

Naimin Li received the Graduate degree from the Department of Medical Treatment, Traditional Chinese Medicine (TCM), Shenyang, China, in 1962.

He is currently a Fellow Professor at the Harbin Institute of Technology, Harbin, China. Since 1965, he has been involved in the study and exploration of the application of tongue diagnosis in modern disease diagnosis and treatment, in the fields of internal medicine, pediatrics, epidemiology, and gynecology. In 1970 and 1972, he was a Leader of Experts Group of TCM by the Chinese Government in the Middle Europe and Albania. In 1989, he started the first tongue image laboratory in China. He has authored or coauthored more than 250 papers and 14 books around his research areas.

Hongzhi Zhang received the M.E. degree in material science and engineering and the Ph.D. degree in computing and biomedical engineering from the Harbin Institute of Technology, Harbin, China, in 2001 and in 2007, respectively.

He is currently an Associate Professor in the School of Computer Science and Technology, Harbin Institute of Technology. His research interests include image processing, pattern recognition, data mining, and biomedical engineering of Chinese medicine.