IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received October 31, 2019, accepted November 19, 2019, date of publication December 6, 2019,
date of current version January 8, 2020.

Digital Object Identifier 10.1109/ACCESS.2019.2958304

The Optimal Morphological Model for Arterial Blood Pressure Wave Related Classification: Comparison of Two Types of Kernel Function Mixtures

YONGXIN CHOU[1], (Member, IEEE), PING WANG[2], AND YUFENG FENG[3]
[1]School of Electrical and Automatic Engineering, Changshu Institute of Technology, Suzhou 215500, China
[2]College of Electrical and Information Engineering, Lanzhou University of Technology, Lanzhou 730050, China
[3]Changshu No.1 People's Hospital, Changshu 215500, China

Corresponding author: Ping Wang (pingwangsky@163.com)

This work was supported in part by the Natural Science Foundation of Jiangsu Province under Grant BK20170436 and Grant BK20181033, in part by the Jiangsu Post-Doctoral Research Project, and in part by the National Natural Science Foundation of China under Grant 61901062 and Grant 61903050.

ABSTRACT The morphological modeling methods are efficient in quantifying the change of arterial blood pressure (ABP) waves. The related works focus on minimizing the modeling error but ignore the classification related modeling expression in practical applications. In this study, we explored the optimal modeling method for ABP wave related classifications. Two types of conventional models, Gaussian or Lognormal kernel function mixtures, were employed to quantitively describe the change of ABP signals, and the parameters of different models were engaged to train the different classifiers by probabilistic neural network (PNN) and random forest (RF) for identifying the ABP waves by age, gender, and whether belonging to extreme bradycardia (EB) or extreme tachycardia (ET). Then, we defined some indexes about the performance of modeling and classifications as the references to compare the different models. The ABP signals of Fantasia and 2015 PhysioNet/CinC Challenge databases were exploited as the experimental data to select the optimal model. The modeling results show that the Lognormal kernel function mixtures have a lower error in ABP wave modeling. The two-sample Kolmogorov-Smirnov test (*ks*-test) results indicate that the parameters of all models are markedly different at a highly significant level ($h = 1, p < 0.05$) between different groups. The classification results show that the classifiers based on the four-Gaussian function model have the best performance with the average Kappa coefficients (KC) of $99.160 \pm 0.123\%$, while the average KC for the classifiers of two-Lognormal function models is $97.585 \pm 0.172\%$, which means there is excessive information redundancy in the classifications by the three and four kernel functions models. Considering some other indexes such as time consumption and RAM space, the 2 Lognormal function model has more potential in practical applications.

INDEX TERMS Arterial blood pressure signal, optimal morphological model, Gaussian function, lognormal function, classification.

I. INTRODUCTION

Arterial blood pressure (ABP) signal is one of the most important physiological signals and contains abundant information about the beat rhythm and hemodynamics of the cardiovascular system. It plays an important role in detecting some cardiovascular diseases, e.g., arteriosclerosis, heart failure, coronary heart disease, and arrhythmia [1]–[3]. It is easy to record the ABP signal from our arterial network through some non-invasive methods such as photoelectric, piezoelectric, or ultrasonic sensors in some wearable or portable devices [4], so the ABP signal has the great potential in detecting and monitoring cardiovascular diseases in m-health.

The blood circulates in the cardiovascular network under the diastolic and contractile forces from the heart. It is reflected in the junctures of some arteries (thoracic and

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

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see http://creativecommons.org/licenses/by/4.0/

abdominal aorta, abdominal aorta and common iliac arteries), and is re-reflected in the junctures of distal vascular [5]. Thus, it is believed that an ABP waveform consists of a pressure wave and several reflection waves [6], and some studies show that these waves can be quantitively described by some kernel functions. Then, some relative works proposed that the ABP waveforms can be decomposed into several functions, by which we obtain a morphological model with several kernel function mixture to quantitively describe the change of the ABP waveform [7]. So far, the morphological modeling method has been used in many areas, e.g. cardiovascular disease classification and analysis [8], psychological stress analysis [9], motion state analysis [10], maternal health status assessment [11], signal compression and reconstruction [12] and motion artifacts detection and correction [6].

The morphological modeling method derives a set of parameters of the kernel functions from the measured ABP waveform. Some physiological or pathological information has been extracted from these parameters, which are employed as the features in some ABP-related classifications. Sorelli *et al.* [13], [14] extracted the crest time, stiffness index (SI), reflection index (RI) from the four-Gaussian morphological model, and trained classifiers for detecting type-1 diabetes and estimating the vascular's age. Paradkar and Chowdhury [2] drew out the augmentation index (AI), SI and RI from the two-Gaussian morphological model and detected the coronary artery disease (CAD) with the sensitivity of 85% and specificity of 78%. Liu *et al.* [3] engaged the three-Gaussian mixture to analyze the ABP signals of heart failure patients and found four model parameters, which are potential in detecting heart failure. Banerjee *et al.* [9] proposed a two-step Gaussian modeling method to assess the psychological stress and found the parameters $\boldsymbol{B}_1$, $\boldsymbol{B}_2$, $\boldsymbol{C}_1$, and $\boldsymbol{C}_2$, which are significantly changed. Li *et al.* [11] presented a three-positive-Gaussian model to analyze the three trimesters of pregnancy. He *et al.* [15] engaged the Gaussian model to detect the unobservable dicrotic notch of the pulse signal and computed the pulse transit time to estimate the arterial blood pressure. The works of these studies show that the morphological modeling method has great potential for ABP-related classifications, while we found that the expression of the morphological model varies in different studies.

Recently, the optimal expression of the morphological model has drawn much attention, and the related works focus on reducing the error between kernel function mixture and measured ABP waveform by selecting the optimal number or/and type of kernel function. The Gaussian, Lognormal, Rayleigh, Gamma [16], Double-exponential [17] and Cosine [18] functions have been employed as the kernel functions of the morphological models and their number have been tried from two to seven [12]. Sološenko *et al.* [19] concluded that the model consists of one-Lognormal and two-Gaussian functions mixture that has the minimum error for the data of atrial fibrillation after comparing the four models with different kernel function mixtures. Wang *et al.* [20] suggested that four- or five-Gaussian models have maximum accuracy. Tigges *et al.* [16], [21] proposed that we can obtain a model with an arbitrarily small error simply by increasing the number of the kernel function, while it will lead to overfitting of data and consequently to the physiologically uninterpretable solution. Recently, Liu *et al.* [17], [22] demonstrated that morphological models with three-Gaussian and three-Lognormal functions are better than that of Raleigh and double-exponential functions for healthy subjects. Obviously, there is a controversy on the ideal number and type of the kernel function. Different optimal morphological models have been concluded since the ABP waveform changed with different kinds of subjects.

To sum up, the related works focus on minimizing the modeling error. However, in the classification-related applications, we are more concerned with the time consumption of the program, the amount of RAM space occupied by the variables, and the results of classification, while the modeling error is not the primary consideration. In fact, the more kernel functions in modeling, the more model parameters need to be computed, the more time consumption in classification and modeling, and the more RAM space for storing parameters. But the increase of model parameters may cause information redundancy, and we do not know whether it has a marked improvement in the performance of classifiers. Therefore, the purpose of this study is to find the optimal morphological model for ABP signal related classifications by comparing two conventional kernel function mixtures. First, the noises and interference are eliminated, and abnormal segments are wiped out from the ABP signal. The ABP signals are segmented into a series of pulse waves. Then, the ABP morphological modeling methods with two conventional kernel function mixtures and the optimal model selecting indexes are introduced in detail. The renowned international database of the Fantasia and the PhysioNet/Computing in Cardiology Challenge 2015 are employed to validate and compare the different models. Moreover, the probabilistic neural network (PNN) and random forest (RF) are engaged to classify the data by age, gender, extreme bradycardia (EB), and extreme tachycardia (ET) based on different models, and we compare the time consumption and performance of the different classification and obtain the optimal model for ABP related classification.

The remainder of this paper is organized as follows: The database and the morphological modeling method used in this study are detailed in section II. The results are presented in section III. Moreover, a thorough analysis of the results is presented in section IV. Finally, the conclusion is given in section V.

II. MATERIALS AND METHODS

A. EXPERIMENTAL DATA

The experimental data used in our work is from the international physiological database PHYSIONET. A group of data recorded from the young and old subjects is engaged in the classification by age and gender, and they are also exploited as the healthy subjects in the detection of patients

with EB and ET. Another group of data recorded from the patients with some life-threaten arrhythmias is employed in the classification by disease.

The data of healthy subjects is from the sub-database *Fantasia* (URL: https://www.physionet.org/physiobank/database/fantasia/) [23], which were donated by Harvard Medical School, Boston University, Beth Israel Hospital and Massachusetts Institute of Technology. Forty rigorously screened healthy subjects, of which 20 young and 20 old subjects with sinus rhythm were employed in the experiment. The number of male and female are the same. During data recording, all subjects were supine in the bed and watching the movie named Fantasia (Disney 1940) to keep waking. The electrocardiogram (ECG), APB and respiratory signals were acquired synchronously with the sampling frequency of 250 Hz (only the record f2y02m with 333 Hz), and the duration of each data is 120 minutes. However, only half of the subjects' ABP signals were recorded, their information is shown in Table 1.

TABLE 1. The information about the subjects in the Fantasia database.

Name of Document	Age	Sex	Name of Document	Age	Sex
f2o01m	73	Female	f2y01m	23	Female
f2o02m	75	Female	f2y02m	23	Male
f2o03m	85	Female	f2y03m	28	Female
f2o04m	70	Female	f2y04m	27	Female
f2o05m	83	Male	f2y05m	25	Female
f2o06m	70	Male	f2y06m	26	Male
f2o07m	77	Male	f2y07m	31	Male
f2o08m	71	Male	f2y08m	21	Male
f2o09m	77	Male	f2y09m	21	Female
f2o10m	73	Female	f2y10m	21	Male

The records of the patients with EB and ET are from the sub-database 2015 PhysioNet/CinC Challenge (URL: https://www.physionet.org/content/challenge-015/1.0.0/) [24], and they were recorded in the Intensive Care Unit (ICU) of four hospitals in the USA and Europe. The aim of this challenge is to improve the true alarms rate in the ICU for some life-threatening arrhythmias, e.g., EB and ET. Patients with EB whose heart rate are lower than 40 beats per minute (bpm) for 5 consecutive beats, and that of patients with ET are over 140 bpm for 17 consecutive beats, so it's easy to detect EB and ET by their definitions. However, the signal channels in monitoring devices are suffered from the movement artifact of body, the sensor movement or disconnects and other events, which generate many abnormal segments inner the signals that are used for monitoring whether the arrhythmias outburst. Thus, the ICU monitors will give the false alarm if they rely on the definitions of EB and ET, and sometimes the false alarm rate is even as high as 86% [25]. To supply a ''gold standard'' for arrhythmias detecting, a team of experts' annotators reviewed each record with alarm and labeled it either 'true alarm' or 'false alarm'. The evaluation criterion is that if two-thirds majority agreed with the annotation, and the corresponding record could be adopted by the challenge. The ECG, APB, photoplethysmogram (PPG) and respiratory signals were acquired synchronously and have been resampled to 12-bit, 250 Hz. In this study, only the 'true alarm' records were engaged in the experiments, so there are forty-five records of the EB subjects and one hundred and thirty-one records of ET subjects. While only seventeen records of EB subjects and thirty-seven records of ET subjects contain the ABP signals, and thirteen of the ET records are heavily corrupted by abnormal segments, so only forty-one records are used finally. The records name of EB subjects are b268s, b455l, b456s, b494s, b495l, b515l, b516s, b517l, b560s, b561l, b562s, b578s, b659l, b664s, b708s, b722s, and b757l, and that of ET subjects are t173l, t208s, t214s, t276s, t277l, t333l, t335l, t406s, t412s, t413l, t417l, t418s, t425l, t594s, t677l, t680s, t690s, t702s, t707l, t719l, t731l, t739l, t760s, and t777l. The record name ended by the letter 's' means the signal with a length of 5 minutes, and that ended by the letter 'l' means the signal with a length of 5.5 minutes.

B. THE MORPHOLOGICAL MODELING METHOD OF ABP WAVE

The core of the morphological modeling method is to synthesize the ABP wave with a mixture of several kernel functions. For an ABP wave $\{y(n)\}$, whose morphological model is denoted as $\{\hat{y}(n)\}$, and then we have,

$$\hat{y}(n) = \boldsymbol{f}(n, \boldsymbol{\theta}) + B(n, \boldsymbol{\psi}) \quad n \in [1, PPI] \tag{1}$$

where, $\hat{y}(n)$ is the evaluation of the n-th sample in ABP wave. n is the index of a sample and it is a positive integer. PPI means the pulse to pulse interval, and here it is also the length or the cycle of the ABP wave. $\boldsymbol{f}(\cdot)$ represents the morphological wave synthesized by several kernel functions. $B(\cdot)$ is the baseline, and $\boldsymbol{\theta}$ and $\boldsymbol{\psi}$ are the parameters vectors of this model.

In (1), we should derive the detailed expression of $\boldsymbol{f}(\cdot)$ and $B(\cdot)$. Because the trend of an ABP wave only has a slight change, the baseline can be approximated as a linear local trend [1]:

$$B(n, \boldsymbol{\psi}) = kn + b \tag{2}$$

where, k and b are the slope and the vertical intercept of baseline, respectively. Then, $\boldsymbol{\psi} = [k, \mathrm{b}]$.

For the expression of $\boldsymbol{f}(\cdot)$, the key is to determinate the type and the number of kernel functions. So far, the mixtures of Gaussian or Lognormal functions are verified to be the most effective in modeling ABP wave, and their number varies from two to four for different subjects [17]. Thus, in this study, we investigate the optimal ABP model expression in classification among the models mixed by two- to four- Gaussian/Lognormal functions. For Gaussian function, the expression of $\boldsymbol{f}(\cdot)$ is:

$$\begin{aligned} \boldsymbol{f}(n, \boldsymbol{\theta}) &= \sum_{l=1}^{h} g_l\,(n, a_l, b_l, c_l) \\ &= \sum_{l=1}^{h} a_l \exp[-\frac{(n/f_s - b_l)^2}{2\,(c_l)^2}] \quad n \in [1, PPI] \end{aligned} \tag{3}$$

where, h is the number of kernel functions, $h = 2$, 3 and 4. $g_l(\cdot)$ is the l-th Gaussian function in the model. f_s is the sampling frequency of the measured ABP signal. a_l, b_l, and c_l are the amplitude, the position and the width of the l-th Gaussian function, respectively.

For the Lognormal function, the expression of $\boldsymbol{f}(\cdot)$ is:

$$\begin{aligned}\boldsymbol{f}(n,\boldsymbol{\theta}) &= \sum_{l=1}^{h} LN_l(n,\alpha_l,\beta_l,\gamma_l) \\ &= \sum_{l=1}^{h} \frac{\alpha_l}{\sqrt{2\pi}\times\gamma_l\times n/f_s}\exp[-\frac{(\ln(\frac{n}{f_s\times\beta_l}))^2}{2(\gamma_l)^2}] \end{aligned} \quad (4)$$

where, $LN_l(\cdot)$ is the l-th Lognormal function in model, α_l, β_l, and γ_l are its parameters, respectively.

Therefore, the ABP wave can be quantified by the modeling parameters:

$$[\,PPI\;\; k\;\; b\;\; \{a_l\;\; b_l\;\; c_l\}_{l=1}^{h}] \quad \text{or} \quad [\,PPI\;\; k\;\; b\;\; \{\alpha_l\;\; \beta_l\;\; \gamma_l\}_{l=1}^{h}] \quad (5)$$

where, $\{\cdot\}_{l=1}^{h}$ means the parameters of different models, e.g., $\{a_l\, b_l\, c_l\}_{l=1}^{2}$ is the parameters of $\boldsymbol{f}(\cdot)$ which consists of two Gaussian functions. Thus,

$$\{a_l\quad b_l\quad c_l\}_{l=1}^{2} = [\,a_1\quad b_2\quad c_2\;a_2\quad b_2\quad c_2\,] \quad (6)$$

Then, we can obtain a vector of all ABP waves:

$$\begin{aligned}&[\boldsymbol{PPI}\quad \boldsymbol{K}\quad \boldsymbol{B}\quad \{\boldsymbol{A}_l\quad \boldsymbol{B}_l\quad \boldsymbol{C}_l\}_{l=1}^{h}] \\ &\text{or} \\ &[\boldsymbol{PPI}\quad \boldsymbol{K}\quad \boldsymbol{B}\quad \{\boldsymbol{\alpha}_l\quad \boldsymbol{\beta}_l\quad \boldsymbol{\gamma}_l\}_{l=1}^{h}] \end{aligned} \quad (7)$$

Fig. 1 depicts the process to select the optimal ABP model for classification. The process includes: 1) ABP signal pre-processing, 2) feature vector extracting, 3) model-related classification, and 4) the optimal model selecting. There are some noise, interference and abnormal segments in raw ABP signal, and the ABP signal should be split into a series of waves by the cardiac cycle. Hence, we should attenuate the noise and interference, eliminate the abnormal segments, and detect the start points of the ABP waves during ABP signal pre-processing. Then, the processed ABP waves are employed to compute the parameter of the different models as the feature vectors for different classifications. Moreover, the features between different classification are selected, and the RF and PNN are utilized to train the classifiers to distinguish the ABP waves by gender, age, health, EB and ET. Finally, some indexes are defined as the standard to assess the performance of ABP modeling and classification, by which we chose the optimal ABP morphological modeling method.

1) ABP SIGNAL PRE-PROCESSING

Three missions are assigned in the ABP signal pre-processing: noise and interference attenuating, abnormal segments detecting and ABP wave start points positioning.

The raw ABP signal is corrupted by some noise and interference, i.e., the baseline wanders, the power line interference and electromyogram (EMG). Here, a fast-mathematical-morphology filtering approach, we proposed in the previous study, is employed to attenuate these noise and interference [26], and the core of this method is to cut or fill the glitches attached at the ABP waveform by some mathematical operators between a signal and a symmetric structuring element (SSE). We can attenuate the different noise or interference from ABP signal by adjusting the length of SSE. For the power line interference and the EMG, an alternate-hybrid filter is designed, and the length of SSE is 10 samples. In addition, the length of SSE increases to 200 samples for designing another alternate-hybrid filter to estimate the baseline wanders of the ABP signal. Then, the processed signal is obtained by subtracting the baseline wanders from the result of the former filter.

Although we have obtained a cleaner signal after filtering, some abnormal segments caused by movement artifact, senor sliding or disconnects still exist in the ABP signal, they lose almost all the features of the ABP waveform and have a great influence on the accuracy of wave modeling. Thus, it is essential to detect and remove these abnormal segments from ABP signal. Here, an ASCD method is engaged to detect the abnormal segments [27].

In order to split the ABP signal into ABP waves, we should find the start points of these waves. We find the start point is the minimum between two systolic waves (inner a period). Thus, we first compute the interval between two adjacent systolic waves with a sliding window iterative method in [28], then the start point is the minimum of the interval. Here, the ABP wave is denoted as $\{x(n)\}$, then we can obtain the ***PPI*** by the first-order difference of start points.

2) FEATURE VECTOR EXTRACTING

Fig. 1 illustrates the process to extract the feature vector from measured ABP waves. First, in order to reduce the influence generated by the difference of ABP waves' amplitude, the ABP waves are normalized by:

$$y(n) = \frac{x(n) - \frac{1}{PPI}\sum_{n=1}^{PPI} x(n)}{\sqrt{\frac{1}{PPI}\sum_{n=1}^{PPI}\left(x(n) - \frac{1}{PPI}\sum_{n=1}^{PPI} x(n)\right)^2}} \quad (8)$$

where, $y(n)$ is the result of the n-th sample in an ABP wave.

In (7), the features in the vector can be divided into three parts: 1) ***PPI*** is the intervals of adjacent ABP waves, it is the results of the first order of the start points. 2) $\boldsymbol{K}$ and $\boldsymbol{B}$ are derived from the linear trend. 3) $\{\boldsymbol{A}_l\ \boldsymbol{B}_l\ \boldsymbol{C}_l\}_{l=1}^{h}$ or $\{\boldsymbol{\alpha}_l\ \boldsymbol{\beta}_l\ \boldsymbol{\gamma}_l\}_{l=1}^{h}$ are estimated from the measured ABP waves by modeling method. Here, according to the expression of the linear trend as (2), we can compute k and b by:

$$k = \frac{y(PPI) - y(1)}{PPI - 1} \quad (9)$$

$$b = y(1) - k \quad (10)$$

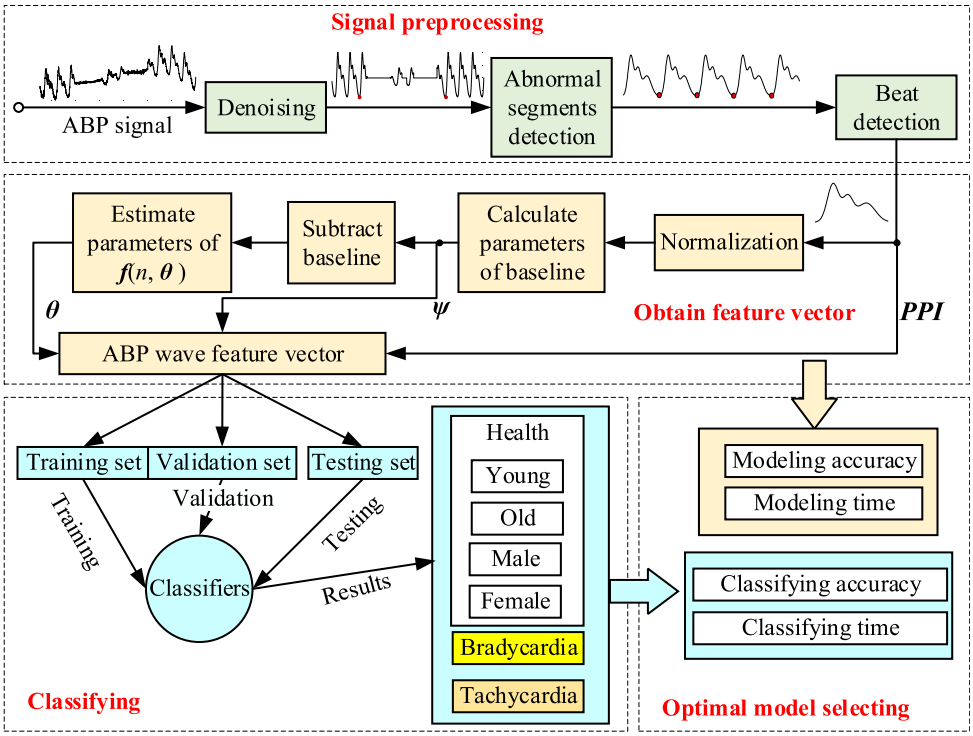


FIGURE 1. Process of optimal model selecting.

where, $y(PPI)$ and $y(1)$ are the end and the start of an ABP wave, respectively.

For the models in (3) and (4), the ABP waves should obey the Dirichlet and Neuman boundary conditions by [1], [29], and the curve obtained by subtracting the linear trend from the ABP wave can approximately meet these two conditions. Then:

$$z(n) = y(n) - kn - b \tag{11}$$

where, $z(n)$ is the result we obtained and can be employed to estimate the parameters of the models in (3) and (4).

For the model parameters estimation, a nonlinear curve fitting method is engaged. The fitting error is defined as:

$$E(\boldsymbol{\theta}) = \frac{1}{PPI}\sum_{n=1}^{PPI}(z(n) - \boldsymbol{f}(n, \boldsymbol{\theta}))^2 \tag{12}$$

Then, we solve (12) by:

$$\begin{aligned} &\min_{\boldsymbol{\theta}} \ E(\boldsymbol{\theta}) \\ &\text{subject to } \boldsymbol{F}\{\boldsymbol{f}(n, \boldsymbol{\theta}), \boldsymbol{\theta}\} \ \text{且} \ lb \le \boldsymbol{\theta} \le ub \end{aligned} \tag{13}$$

where, $\boldsymbol{F}\{\boldsymbol{f}(\boldsymbol{n}, \boldsymbol{\theta}), \boldsymbol{\theta}\}$ is the constraint condition. $lb \le \boldsymbol{\theta} \le ub$ is the boundary condition, lb and ub are the lower and upper bounds of parameters. Here, the nonlinear least-squares method is engaged to solve (13). Some details we used are in Table 2.

TABLE 2. Parameters setting of curve fitting.

Parameters	Values
lb	$[\min(z(n)), 0, 0]$ #
ub	$[\max(z(n)), PPI(i), PPI(i)]$
$\boldsymbol{F}\{\boldsymbol{f}(n, \boldsymbol{\theta}), \boldsymbol{\theta}\}$	$b_l < b_{l+1}, 0 < \alpha_l, \beta_l, \gamma_l < 1$
Initial values	mean($\boldsymbol{\theta}'$)
Optimization algorithm	'Trust-Region'
Maximum number of iterations	500
Iterative termination condition	$\Delta E(\boldsymbol{\theta}) \leqslant 0.001$ or $\Delta\boldsymbol{\theta} \leqslant 0.001$

#min($z(n)$) and max($z(n)$) are the minimum and maximum of an ABP wave, respectively. mean($\boldsymbol{\theta}'$) is the mean of $\boldsymbol{\theta}'$ corresponding to the average parameters of the previous ABP wave's model.

3) MODEL-RELATED CLASSIFICATION

The PNN and RF are utilized to design the classifiers, which are trained and tested by the extracted feature vector. The neural network toolbox and the 'randomforest-matlab' toolbox (available at https://code.google.com/archive/p/randomforest-matlab/) are engaged to achieve the PNN and the RF classifiers in MATLAB, respectively.

PNN is a kind of radial basis network suitable for classification [30]. Here, the function 'newpnn($\boldsymbol{P}$, $\boldsymbol{T}$, *spread*)' in the toolbox is utilized to create a two-layer network. $\boldsymbol{P}$ is the input vectors. $\boldsymbol{T}$ is the target class vectors. *spread* is the spread speed of the radial basis functions (RBF). The first layer consists of RBF neurons and their weighted inputs are computed with Euclidean distance weight function. The second layer has competitive transfer function neurons and their weighted

inputs are calculated with Dot product weight function. The spread speed we used is 0.02.

RF adds an additional layer of randomness to Bagging on the basis of the decision tree method, and it utilizes the Gini index to compute the weight. The function 'classRF_train(***P***, ***T***, *ntree*, *mtry*, *extra_options*)' in the toolbox [31] is engaged to train RF classifiers. ***P*** and ***T*** are the input vector and the target class vector, respectively. *ntree* is the number of trees grown and here it is set to 30. *extra_options* are some other options to control RF and their initial values we used are the default. *mtry* is the number of predictors sampled for splitting at each node and can be calculated by the number of features:

$$mtry = \left\lfloor \sqrt{h \times 3 + 3} \right\rfloor \tag{14}$$

where, $\lfloor \cdot \rfloor$ means the rounding down operation. The result of $h \times 3 + 3$ is the number of futures.

In this study, we classified the ABP waves by gender, age, health, EB and ET based on the feature vectors. In *Fantasia* database, there are 79310 ABP waves from healthy subjects in total, in which that of 41170 belong to young subjects, 38140 belong to old subjects, 42955 belong to females and 36355 belong to males. In *2015 PhysioNet/CinC Challenge* database, we extracted 4595 ABP waves from EB subjects and 12836 ABP waves from ET subjects. Thus, there are 96741 ABP waves in total. They were divided into a training set, a validation set, and a testing set, and their corresponding numbers for different classifications are shown in Table 3. Here, the cross-validation is engaged to improve the performance of classifiers.

TABLE 3. The sub-datasets of different classifications.

Classification[$]	Training Set	Validation Set	Testing Set	Total	Labels
Y vs O	56310	8000	15000	79310	1 vs 2
M vs F	56310	8000	15000	79310	1 vs 2
Ht vs EB	60905	8000	15000	83905	1 vs 2
Ht vs ET	69146	8000	15000	92146	1 vs 2
EB vs ET	13431	2000	2000	17431	1 vs 2
Ht vs EB vs ET	73741	8000	15000	96741	1 vs 2 vs 3

[$]The 'Y', 'O', 'M', 'F' and 'Ht' are the abbreviations of young, old, male, female and healthy, respectively.

4) THE OPTIMAL MODEL SELECTING

The aim of this study is to find the optimal morphological modeling mixture for ABP-related classification, so it is important to employ some indexes as the standards to assess the performance of the ABP modeling method and the classifiers. Here, the modeling accuracy and modeling time-consumption are calculated to analyze the results of modeling. The classification performance and time-consumption are used to evaluate the results of different classifiers. The time-consumption of the modeling and classification is counted by the functions (tic and toc) in MATLAB.

The root mean square error (RMSE) is utilized to compute the modeling accuracy:

$$RMSE = \sqrt{\frac{1}{PPI}\sum_{n=1}^{PPI}\left(y(n) - \hat{y}(n)\right)^2} \tag{15}$$

Then, the specificity (Sp), sensitivity (Se), accuracy (Ac) and kappa coefficient (KC) are exploited to assess the performance of classifiers. Among them, the definitions of *Sp*, *Se,* and *Ac* are [32]:

$$Sp = \frac{TN}{FP + TN} \times 100\% \tag{16}$$

$$Se = \frac{TP}{TP + FN} \times 100\% \tag{17}$$

$$Ac = \frac{TP + TN}{TP + FP + FN + TN} \times 100\% \tag{18}$$

where, *TP*, *FP*, *FN,* and *TN* are the true positive, false positive, false negative, true negative of the classification results, respectively. Actually, *TP* is the number of ABP waves belong to AA and are classified as AA. *FP* is the number of ABP waves belong to BB and are classified as AA. *FN* is the number of ABP waves belong to AA and are classified as BB. *TN* is the number of ABP waves belong to BB and are detected as BB. Here, AA and BB replace the categories that the ABP waves should belong to. AA is young while BB is old in the classification by age. AA is male while BB is female in the classification by gender. AA is a patient while BB is healthy in the classification of healthy subjects and patients. AA is the patients with EB and BB is that of ET in the classification of patients.

In addition, the KC is engaged to measure the average performance of classifiers, and its definition is [33]:

$$KC = \frac{p_0 - p_e}{1 - p_e} \tag{19}$$

$$p_0 = \frac{\sum_{t=1}^{r} q_{tt}}{m} \tag{20}$$

$$p_e = \frac{\sum_{t=1}^{r} (q_{t+} \times q_{+t})}{m^2} \tag{21}$$

where, p_0 and p_e are the parameters derived from the confusion matrix of classification results. p_0 is the ratio of the true classification, and p_e is the ratio of false classification. r is the number of the rows in the confusion matrix. q_{tt} replaces the diagonal element in the confusion matrix. q_{t+} is the sum of the elements on the line t, and q_{+t} is the sum of the elements on the column t. *KC* is in the range of $[-1, 1]$. The closer the KC value of the classification result is to 1, and the better the classification result is acquired. Moreover, we assess the classification result of each category by KC_t, which is defined as:

$$KC_t = \frac{p_{tt} - p_{t+}p_{+t}}{p_{+t} - p_{t+}p_{+t}} \tag{22}$$

$$p_{tt} = \frac{q_{tt}}{m} \tag{23}$$

$$p_{t+} = \frac{q_{t+}}{m} \tag{24}$$

$$p_{+t} = \frac{q_{+t}}{m} \tag{25}$$

TABLE 4. (a) The modeling results with 2 Gaussian functions. (b) The modeling results with 3 Gaussian functions. (c) The modeling results with 4 Gaussian functions.

(a)

P *	Y	O	F	M	Ht	EB	ET
A_1	3.138±0.382	2.772±0.293	2.998±0.391	2.919±0.381	2.962±0.388	2.504±0.774	2.607±0.624
B_1	0.128±0.014	0.184±0.027	0.154±0.040	0.156±0.028	0.155±0.035	0.246±0.153	0.128±0.040
C_1	0.048±0.008	0.076±0.014	0.060±0.020	0.062±0.016	0.061±0.018	0.121±0.129	0.046±0.013
A_2	1.584±0.278	1.334±0.376	1.397±0.354	1.543±0.332	1.464±0.352	1.461±0.735	1.124±0.593
B_2	0.373±0.037	0.434±0.056	0.393±0.062	0.413±0.046	0.402±0.056	0.382±0.134	0.276±0.457
C_2	0.199±0.037	0.207±0.043	0.191±0.034	0.217±0.043	0.203±0.041	0.173±0.097	0.098±0.845
K	0.007±0.174	0.004±0.200	0.006±0.191	0.005±0.182	0.006±0.187	0.006±0.200	-0.064±1.104
B	-1.231±0.180	-1.182±0.203	-1.201±0.200	-1.215±0.185	-1.207±0.193	-1.090±0.200	-1.009±0.434
PPI	0.931±0.133	1.028±0.181	0.917±0.134	1.051±0.168	0.978±0.165	1.068±0.469	0.541±0.860
RMSE	0.214±0.037	0.192±0.067	0.197±0.060	0.211±0.048	0.203±0.055	0.144±0.044	0.138±0.056
DL	38343.507	39219.460	39368.852	38194.115	77562.967	4905.800	6949.124
Time	1651.567	1715.277	1794.346	1572.498	3366.844	226.516	623.499

*The letter P means the model parameters or indexes used to evaluate the modeling results. The data in all Tables are expressed as 'mean±std'. *DL* is the abbreviation of data length.

(b)

P	Y	O	F	M	Ht	EB	ET
A_1	2.530±0.912	2.333±0.586	2.424±0.785	2.449±0.772	2.435±0.779	1.748±0.722	1.908±0.777
B_1	0.148±0.051	0.176±0.056	0.155±0.055	0.168±0.054	0.161±0.055	0.188±0.114	0.107±0.034
C_1	0.048±0.014	0.056±0.021	0.050±0.018	0.053±0.018	0.052±0.018	0.068±0.045	0.034±0.011
A_2	1.986±0.912	1.869±0.546	1.946±0.802	1.910±0.709	1.930±0.761	2.196±0.751	1.910±0.585
B_2	0.251±0.200	0.247±0.090	0.248±0.109	0.250±0.200	0.249±0.157	0.268±0.207	0.158±0.056
C_2	0.064±0.030	0.076±0.042	0.077±0.041	0.062±0.028	0.070±0.036	0.096±0.089	0.045±0.018
A_3	1.456±0.256	1.270±0.389	1.335±0.354	1.404±0.317	1.366±0.339	0.909±0.343	0.891±0.544
B_3	0.400±0.085	0.451±0.120	0.401±0.131	0.452±0.055	0.425±0.106	0.511±0.216	0.302±0.078
C_3	0.162±0.056	0.192±0.067	0.154±0.064	0.203±0.050	0.176±0.063	0.195±0.100	0.082±0.040
RMSE	0.121±0.033	0.097±0.032	0.102±0.030	0.118±0.037	0.110±0.035	0.078±0.039	**0.057±0.042**
Time	2242.379	2318.670	2439.056	2121.993	4561.049	281.393	778.523

(c)

P	Y	O	F	M	Ht	EB	ET
A_1	2.677±0.759	2.237±0.746	2.304±0.820	2.656±0.693	2.465±0.784	1.869±0.643	1.538±0.685
B_1	0.166±0.055	0.180±0.106	0.197±0.102	0.144±0.037	0.173±0.084	0.149±0.090	0.111±0.044
C_1	0.056±0.017	0.051±0.018	0.058±0.021	0.048±0.011	0.053±0.018	0.051±0.035	0.030±0.011
A_2	2.230±0.970	1.886±0.739	2.127±0.938	1.992±0.809	2.065±0.884	1.463±0.708	1.642±0.806
B_2	0.226±0.140	0.289±0.137	0.209±0.110	0.313±0.154	0.256±0.142	0.342±0.275	0.129±0.081
C_2	0.059±0.022	0.081±0.046	0.059±0.020	0.082±0.048	0.069±0.037	0.108±0.126	0.035±0.020
A_3	1.247±0.606	1.614±0.937	1.370±0.877	1.486±0.704	1.423±0.804	1.606±0.853	1.288±0.691
B_3	0.395±0.121	0.411±0.173	0.395±0.125	0.412±0.172	0.403±0.148	0.341±0.157	0.218±0.096
C_3	0.076±0.026	0.102±0.070	0.072±0.025	0.108±0.069	0.088±0.054	0.098±0.049	0.046±0.021
A_4	0.769±0.252	0.822±0.351	0.756±0.281	0.839±0.324	0.794±0.304	0.757±0.620	0.803±0.548
B_4	0.573±0.114	0.556±0.108	0.562±0.104	0.568±0.120	0.565±0.111	0.531±0.180	0.306±0.090
C_4	0.149±0.048	0.130±0.062	0.144±0.049	0.135±0.063	0.140±0.056	0.146±0.080	0.072±0.038
RMSE	0.068±0.018	**0.044±0.018**	0.058±0.026	0.053±0.015	0.056±0.022	0.042±0.021	**0.037±0.038**
Time	2670.839	2413.373	2758.277	2325.935	5084.212	332.540	905.445

The experiments were implemented via MATLAB 2016a, which is installed at a laptop with an Intel(R) Core (TM) i7-6700HQ CPU at 2.6 GHz clock speed, 16 GB installed memory, and 64-bit windows-7 operating system.

III. RESULTS

A. RESULTS OF ABP MORPHOLOGICAL MODELING

The proposed method is engaged to compute the ABP morphological models of the experimental data in different groups, and the results are shown in Table 4(a) to Table 5(c). We only present the values of K, B, and *PPI* in Table 4(a), because they are no changes among the models with different kernel function mixtures. The variable *DL* means the sum of ABP waves' lengths of one group, and the *time* is the time consumed by modeling all ABP waves of one group.

Table 4(a) to Table 4(c) are the modeling results of two-, three- and four-Gaussian functions, respectively. We can infer from the tables that the accuracies of models increase as the number of kernel functions grows from two to four. The range of *RMSE* are 0.138 ± 0.056 to 0.214 ± 0.037, 0.057 ± 0.042 to 0.121 ± 0.033, and 0.037 ± 0.038 to 0.068 ± 0.018 for different groups, and they diminish as the growth of kernel functions' number. The mean *RMSEs* of the different mixtures are 0.192 ± 0.060, 0.101 ± 0.040, and 0.0528 ± 0.025, and it decreases markedly as the more functions used in models. Fig. 2 is the modeling results of an ABP wave we chose randomly. The (a) to (c) indicate that the difference

TABLE 5. (a) The modeling results with 2 Lognormal functions. (b) The modeling results with 3 Lognormal Functions. (c) The modeling results with 4 Lognormal Functions.

(a)

P	Y	O	F	M	Ht	EB	ET
α_1	0.645±0.130	0.719±0.355	0.676±0.299	0.685±0.221	0.680±0.266	0.681±0.475	0.326±0.226
β_1	0.155±0.031	0.252±0.097	0.187±0.057	0.220±0.108	0.202±0.086	0.305±0.226	0.140±0.158
γ_1	0.512±0.074	0.479±0.148	0.500±0.118	0.492±0.115	0.496±0.117	0.486±0.177	0.424±0.104
α_2	0.515±0.220	0.520±0.376	0.446±0.281	0.601±0.312	0.517±0.305	0.511±0.534	0.224±0.203
β_2	0.476±0.046	0.475±0.160	0.479±0.110	0.472±0.121	0.476±0.115	0.454±0.250	0.276±0.482
γ_2	0.336±0.100	0.408±0.202	0.355±0.164	0.389±0.157	0.371±0.162	0.401±0.247	0.294±0.150
RMSE	0.122±0.034	0.115±0.039	0.121±0.032	0.117±0.041	0.119±0.036	0.088±0.030	0.110±0.065
Time	1371.209	1568.833	1580.329	1359.713	2940.042	212.986	652.246

(b)

P	Y	O	F	M	Ht	EB	ET
α_1	0.598±0.182	0.587±0.319	0.609±0.242	0.574±0.273	0.593±0.257	0.606±0.397	0.244±0.211
β_1	0.150±0.025	0.235±0.062	0.186±0.070	0.197±0.053	0.191±0.063	0.248±0.138	0.141±0.354
γ_1	0.492±0.089	0.495±0.208	0.530±0.131	0.450±0.175	0.493±0.158	0.528±0.154	0.392±0.123
α_2	0.305±0.242	0.312±0.275	0.331±0.228	0.280±0.288	0.308±0.258	0.279±0.315	0.187±0.176
β_2	0.439±0.112	0.327±0.160	0.397±0.133	0.371±0.164	0.385±0.148	0.400±0.281	0.195±0.144
γ_2	0.303±0.208	0.293±0.159	0.335±0.188	0.254±0.172	0.298±0.186	0.296±0.122	0.294±0.129
α_3	0.270±0.248	0.335±0.299	0.186±0.217	0.438±0.275	0.302±0.276	0.290±0.361	0.121±0.161
β_3	0.467±0.089	0.500±0.128	0.442±0.123	0.530±0.068	0.483±0.111	0.491±0.240	0.316±0.631
γ_3	0.230±0.148	0.278±0.135	0.190±0.120	0.327±0.135	0.253±0.144	0.288±0.163	0.220±0.122
RMSE	0.077±0.017	0.084±0.033	0.082±0.031	0.079±0.020	0.081±0.027	0.050±0.026	**0.063±0.056**
Time	1856.703	2264.538	2235.124	1886.117	4121.241	287.122	876.066

(c)

P	Y	O	F	M	Ht	EB	ET
α_1	0.306±0.283	0.442±0.266	0.307±0.242	0.447±0.309	0.371±0.283	0.312±0.240	0.246±0.162
β_1	0.138±0.027	0.230±0.104	0.164±0.055	0.204±0.111	0.182±0.088	0.303±0.257	0.129±0.044
γ_1	0.372±0.167	0.425±0.155	0.397±0.177	0.398±0.147	0.397±0.164	0.362±0.188	0.404±0.119
α_2	0.228±0.258	0.277±0.236	0.221±0.237	0.288±0.257	0.252±0.249	0.306±0.344	0.104±0.091
β_2	0.405±0.135	0.384±0.134	0.409±0.124	0.379±0.145	0.395±0.135	0.442±0.254	0.276±0.116
γ_2	0.220±0.170	0.309±0.227	0.258±0.227	0.268±0.174	0.263±0.205	0.304±0.152	0.202±0.116
α_3	0.413±0.304	0.252±0.195	0.287±0.238	0.394±0.293	0.336±0.270	0.243±0.247	0.104±0.084
β_3	0.278±0.157	0.319±0.174	0.273±0.164	0.327±0.166	0.298±0.167	0.321±0.257	0.285±0.127
γ_3	0.406±0.236	0.284±0.151	0.324±0.197	0.374±0.219	0.347±0.209	0.339±0.193	0.219±0.128
α_4	0.224±0.195	0.269±0.246	0.318±0.251	0.160±0.141	0.245±0.222	0.310±0.341	0.092±0.085
β_4	0.543±0.151	0.396±0.180	0.432±0.128	0.521±0.219	0.473±0.181	0.382±0.198	0.271±0.162
γ_4	0.246±0.164	0.303±0.218	0.344±0.223	0.191±0.106	0.274±0.194	0.354±0.202	0.263±0.153
RMSE	0.054±0.019	**0.051±0.021**	0.054±0.022	0.051±0.018	0.053±0.020	0.035±0.026	**0.051±0.063**
Time	2725.979	2735.770	3027.950	2433.799	5461.749	363.965	984.510

between a raw ABP wave and its model decreases with the increase of kernel functions. We can conclude the same result from (a1) to (c1), which are the corresponding residual error of the figures above them, respectively.

The accuracies of the models with two- to four-Lognormal functions also increase with the growth of the kernel functions' number. The mean ***RMSEs*** are 0.116 ± 0.042, 0.077 ± 0.033, and 0.051 ± 0.030, respectively. We can obtain the same results from (e) to (f) and (e1) to (f1) in Fig. 2. The conclusion is the same as that of the modeling results of the Gaussian function.

However, we found that the Lognormal function is more suitable for ABP wave modeling than the Gaussian function. The mean ***RMSEs*** of Gaussian-function-based models are 0.192 ± 0.060, 0.101 ± 0.040, and 0.0528 ± 0.025, while that of Lognormal-function-based models are just 0.116 ± 0.042, 0.077 ± 0.033, and 0.051 ± 0.030 under the same number of kernel function. The ***RMSEs*** of all groups in Table 4(a) are larger than that in Table 5(a). The same conclusion we would obtain by comparing Table 4(b) with Table 5(b), and Table 4(c) with 5(c) except some singularities we bolded them up in tables. In Fig. 2, the ABP waveform has a long tail after the high systolic wave, and we found that the Lognormal function also has long tail after its peak, while the Gaussian function is bilaterally symmetric. Thus, we can easily infer that why the Lognormal-function-based models are more precise than the models constructed by Gaussian functions.

Table 4(a) illustrates that the length of all ABP waves form the healthy, EB and ET subjects is 89417.89 seconds (s). The modeling time of two-, three- and four-Gaussian-based models are 4216.859 s, 5620.965 s, and 6322.197 s and that of Lognormal-function-based models are 3805.274 s, 5284.429 s, and 6810.224 s. For the models with the same kernel function, the computational load increases as the growth of the kernel functions' number. The reason is that the number of model parameters increases as the more kernel function

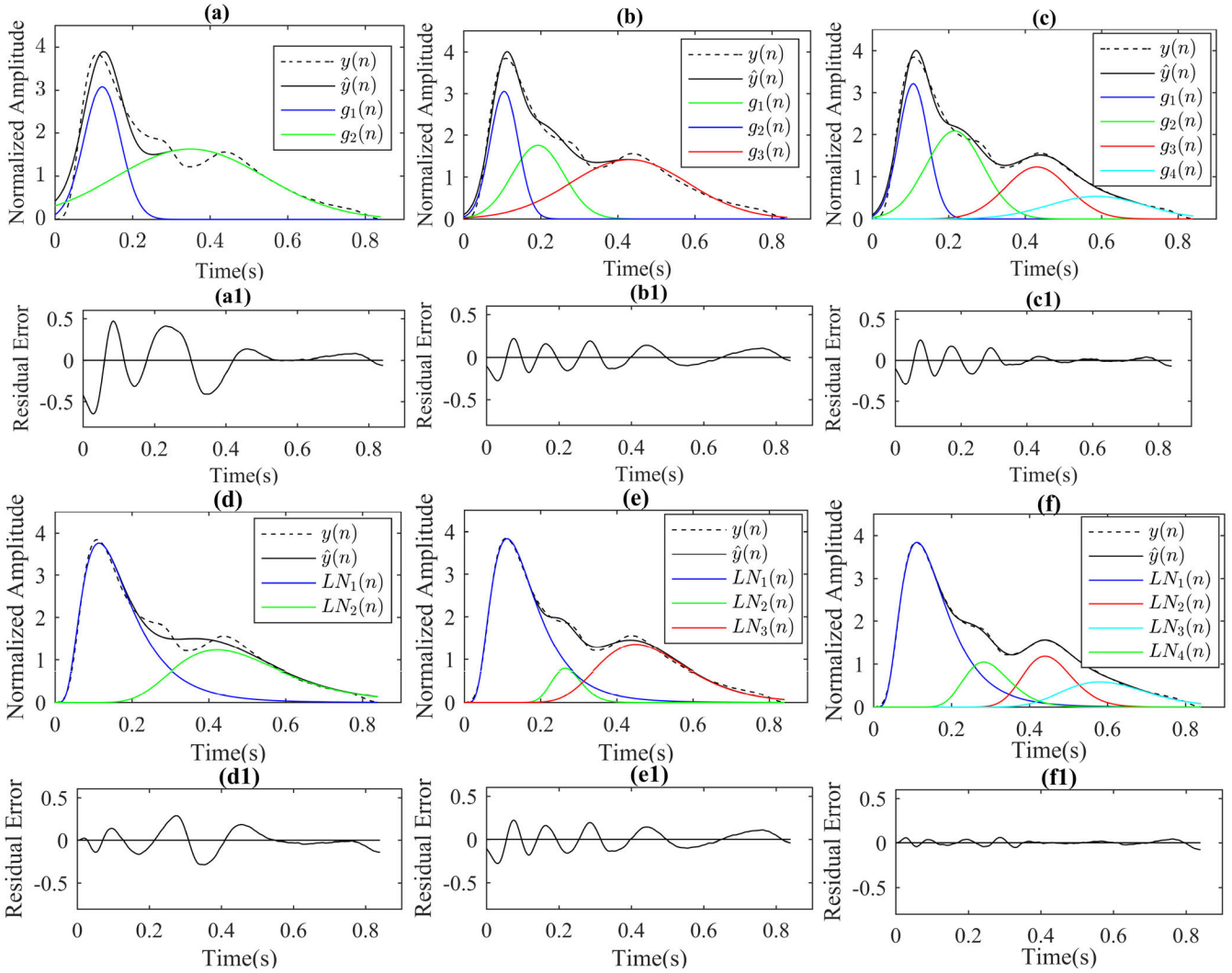


FIGURE 2. An example of ABP wave modeling using different kernel functions mixtures. (a) two Gaussian functions, (b) three Gaussian functions, (c) four Gaussian functions, (d) two Lognormal functions, (e) three Lognormal functions, and (f) four Lognormal functions. (a1) to (f1) are the residual errors corresponding to (a) to (f), respectively.

employed by (3) and (4), and there are 9, 12 and 15 parameters for the models with 2, 3 and 4 kernel functions. The time consumed by the models with 2 and 3 Gaussian functions are less than the models with 2 and 3 Lognormal functions, while the model with 4 Lognormal functions costs more time than the model with 4 Gaussian functions.

B. THE CLASSIFICATION RESULTS OF DIFFERENT MODELS

In this study, since some model parameters do not obey the normal distributions, the two-sample Kolmogorov-Smirnov test (*ks*-test) is utilized to examine the different levels of the same parameter in different models, and the results are depicted in Table 6(a) to Table 7(c). Then, we select the parameters with markedly change to train the classifiers. The null hypothesis is defined as the parameters of two different group obey the same statistical probability distribution, and h is denoted to express the hypothesis result under the probability p. h is 1 when the statistic test rejects the null hypothesis less than the 5 % significance level, and 0 otherwise. Most studies refer to statistically significant as $p < 0.05$ and statistically highly significant as $p < 0.001$. Because the parameters $\boldsymbol{K}$, $\boldsymbol{B}$, and $\boldsymbol{PPI}$ are unchanged among different models, we just listed the results in Table 6(a). The *ks*-test results show that all the model parameters between two different groups are markedly different at a highly significant level ($h = 1, p < 0.001$), and most of them are in an extremely significant level ($h = 1, p = 0$). Therefore, all the model parameters are employed as the feature vector to train the classifier in this study.

Then, the feature vectors of different groups are exploited to train the classifiers with PNN and RF, the ABP waves are classified by age, gender, health, EB and ET. Each classifier is trained 100 times to eliminate the influence from the different input samples by randomly changing the samples in training,

TABLE 6. (a) The results of *ks*-test for the 2-Gaussians models' parameters of different subjects. '*h*' and '*p*' are the hypothesis and probability of *ks*-test. (b) The results of *ks*-test for the 3-Gaussians models' parameters. (c) The results of *ks*-test for the 4-Gaussians models' parameters.

(a)

P	Y vs O		F vs M		Ht vs Bc		Ht vs Tc		Tc vs Bc	
	h	*p*	*h*	*p*	*h*	*p*	*h*	*p*	*h*	*p*
A_1	1	0	1	3.4×10^{-139}	1	0	1	0	1	2.5×10^{-56}
B_1	1	0	1	0	1	0	1	0	1	0
C_1	1	0	1	0	1	0	1	0	1	0
A_2	1	0	1	0	1	0	1	0	1	0
B_2	1	0	1	0	1	1.1×10^{-248}	1	0	1	0
C_2	1	0	1	0	1	0	1	0	1	0
K	1	6.5×10^{-307}	1	5.8×10^{-7}	1	1.6×10^{-4}	1	0	1	1.5×10^{-223}
B	1	0	1	1.6×10^{-90}	1	0	1	0	1	2.7×10^{-214}
PPI	1	0	1	0	1	9.6×10^{-279}	1	0	1	0

(b)

P	Y vs O		F vs M		Ht vs Bc		Ht vs Tc		Tc vs Bc	
	h	*p*	*h*	*p*	*h*	*p*	*h*	*p*	*h*	*p*
A_1	1	0	1	2.2×10^{-116}	1	0	1	0	1	1.1×10^{-87}
B_1	1	0	1	0	1	1.6×10^{-90}	1	0	1	0
C_1	1	0	1	0	1	2.1×10^{-238}	1	0	1	0
A_2	1	0	1	3.4×10^{-313}	1	2.9×10^{-321}	1	2.8×10^{-259}	1	1.3×10^{-146}
B_2	1	0	1	0	1	1.7×10^{-151}	1	0	1	0
C_2	1	0	1	0	1	6.3×10^{-209}	1	0	1	0
A_3	1	0	1	2.9×10^{-310}	1	0	1	0	1	3.8×10^{-167}
B_3	1	0	1	0	1	1.5×10^{-165}	1	0	1	0
C_3	1	0	1	0	1	1.9×10^{-79}	1	0	1	0

(c)

P	Y vs O		F vs M		Ht vs Bc		Ht vs Tc		Tc vs Bc	
	h	*p*	*h*	*p*	*h*	*p*	*h*	*p*	*h*	*p*
A_1	1	0	1	0	1	0	1	0	1	9.5×10^{-235}
B_1	1	0	1	0	1	0	1	0	1	0
C_1	1	0	1	0	1	1.9×10^{-322}	1	0	1	0
A_2	1	0	1	0	1	0	1	0	1	1.8×10^{-205}
B_2	1	0	1	0	1	1.2×10^{-170}	1	0	1	0
C_2	1	0	1	0	1	1.1×10^{-162}	1	0	1	0
A_3	1	0	1	0	1	2.0×10^{-84}	1	1.6×10^{-173}	1	1.4×10^{-222}
B_3	1	0	1	0	1	0	1	0	1	0
C_3	1	0	1	0	1	3.4×10^{-260}	1	0	1	0
A_4	1	0	1	2.7×10^{-225}	1	0	1	2.0×10^{-228}	1	2.9×10^{-82}
B_4	1	0	1	2.6×10^{-301}	1	3.0×10^{-162}	1	0	1	0
C_4	1	0	1	0	1	5.5×10^{-165}	1	0	1	0

validation, and testing sets, and the results are shown in Table 8(a) to Table 8(f).

Table 8(a) shows the classifying results of the young and old subjects, the '2G', '3G', '4G', '2L', '3L' and '4L' are the abbreviations of the ABP models with two Gaussian functions mixtures, three Gaussian functions mixtures, four Gaussian functions mixtures, two Lognormal functions mixtures, three Lognormal functions mixtures, and four Lognormal functions mixtures, respectively. The results are shown as the 'mean ± standard deviation' of 100 train results.

For the classifying approaches, the results of RF-based classifiers are better than that of PNN-based classifiers. The *KC* of RF-based classifiers for 2G to 4L models are 98.335 ± 0.119%, 99.425 ± 0.090%, 99.761 ± 0.054%, 98.534 ± 0.093%, 99.602 ± 0.066%, and 99.618 ± 0.067%,

TABLE 7. (a) The results of *ks*-test for the 2-lognormal models' parameters. (b) The results of *ks*-test for the 3-Lognormal models' parameters. (c) The results of *ks*-test for the 4-Lognormal models' parameters.

(a)

P	Y vs O		F vs M		Ht vs Bc		Ht vs Tc		Tc vs Bc	
	h	*p*	*h*	*p*	*h*	*p*	*h*	*p*	*h*	*p*
A_1	1	0	1	0	1	1.1×10^{-122}	1	0	1	0
B_1	1	0	1	0	1	0	1	0	1	0
C_1	1	0	1	3.5×10^{-282}	1	2.8×10^{-231}	1	0	1	0
A_2	1	0	1	0	1	0	1	0	1	0
B_2	1	0	1	0	1	0	1	0	1	0
C_2	1	0	1	0	1	0	1	0	1	0

(b)

P	Y vs O		F vs M		Ht vs Bc		Ht vs Tc		Tc vs Bc	
	h	*p*	*h*	*p*	*h*	*p*	*h*	*p*	*h*	*p*
A_1	1	0	1	0	1	8.4×10^{-119}	1	0	1	0
B_1	1	0	1	0	1	3.9×10^{-248}	1	0	1	0
C_1	1	0	1	0	1	1.8×10^{-316}	1	0	1	0
A_2	1	1.9×10^{-141}	1	0	1	6.0×10^{-109}	1	0	1	1.7×10^{-156}
B_2	1	0	1	0	1	2.6×10^{-242}	1	0	1	0
C_2	1	2.8×10^{-200}	1	0	1	4.3×10^{-244}	1	0	1	1.6×10^{-21}
A_3	1	0	1	0	1	2.1×10^{-126}	1	0	1	2.5×10^{-178}
B_3	1	0	1	0	1	4.3×10^{-181}	1	0	1	0
C_3	1	0	1	0	1	1.3×10^{-318}	1	0	1	4.0×10^{-312}

(c)

P	Y vs O		F vs M		Ht vs Bc		Ht vs Tc		Tc vs Bc	
	h	*p*	*h*	*p*	*h*	*p*	*h*	*p*	*h*	*p*
A_1	1	0	1	0	1	7.3×10^{-138}	1	0	1	2.6×10^{-137}
B_1	1	0	1	0	1	0	1	0	1	0
C_1	1	0	1	0	1	8.3×10^{-141}	1	0	1	3.5×10^{-220}
A_2	1	0	1	0	1	1.7×10^{-99}	1	0	1	0
B_2	1	2.1×10^{-198}	1	0	1	2.3×10^{-209}	1	0	1	0
C_2	1	0	1	0	1	0	1	1.9×10^{-179}	1	1.8×10^{-321}
A_3	1	0	1	0	1	1.5×10^{-205}	1	0	1	1.5×10^{-267}
B_3	1	0	1	0	1	5.3×10^{-69}	1	9.1×10^{-220}	1	1.3×10^{-51}
C_3	1	0	1	0	1	2.4×10^{-113}	1	0	1	4.1×10^{-318}
A_4	1	3.4×10^{-220}	1	0	1	7.3×10^{-100}	1	0	1	0
B_4	1	0	1	0	1	0	1	0	1	0
C_4	1	0	1	0	1	1.7×10^{-271}	1	3.3×10^{-124}	1	1.7×10^{-134}

respectively, while that of PNN-based classifiers are just 96.975 ± 0.204%, 98.453 ± 0.140%, 99.197 ± 0.088%, 97.882 ± 0.132%, 99.424 ± 0.073% and 99.285 ± 0.080%. Meanwhile, the time RF-based classifiers taken is less than that of PNN-based classifiers taken. That is, the RF-based classifiers take less training time and achieve higher accuracy.

For the classifying results of the models with different kernel functions mixtures, the *KC* and the *Time* in Table 8(a) increase as the increase of kernel functions. For instance, the *KC* increases from 98.335 ± 0.119% and 99.425 ± 0.090% to 99.761 ± 0.054% for the 2G, 3G, and 4G models, respectively. The *KC* increases from 98.534 ± 0.093% and 99.602 ± 0.066% to 99.618 ± 0.067% for the 2L, 3L and 4L models, respectively. For models with the same order but different kernel functions, the *KC* of 2G- and 3G-models are less than that of 2L- and 3L-models, respectively. But the *KC* of the 4G-model is higher than that of the 4L-model, and it is the highest value among the classification results, as is bold in Table 8(a).

TABLE 8. (a) The results of classification for the young and old subjects with different models. '2G', '3G', '4G', '2L', '3L' and '4L' mean the models with two gaussian functions mixtures, three gaussian functions mixtures, four gaussian functions mixtures, two Lognormal functions mixtures, three Lognormal functions mixtures, and four Lognormal functions mixtures, respectively. (b) The results of classification for the male and female subjects with different models. (c) The results of classification for the healthy and EB subjects with different models. (d) The results of classification for the Healthy and ET subjects with different models.

(a)

		TP	*FP*	*FN*	*TN*	*Sp* (%)	*Se* (%)	*Ac* (%)	*KC* (%)	KC_1 (%)	KC_2 (%)	*Time* (s)
2G	PNN	7095±63	109±11	118±9	7679±61	98.605±0.145	98.366±0.129	98.490±0.102	96.975±0.204	96.857±0.244	97.094±0.300	22.613±0.387
	RF	7157±62	69±7	55±8	7718±63	99.111±0.094	99.231±0.106	99.169±0.060	98.335±0.119	98.516±0.204	98.154±0.192	4.794±0.146
3G	PNN	7147±50	57±8	58±7	7737±48	99.264±0.101	99.189±0.099	99.228±0.070	98.453±0.140	98.439±0.190	98.467±0.213	22.673±0.373
	RF	7191±51	28±6	15±4	7766±50	99.643±0.072	99.789±0.050	99.713±0.045	99.425±0.090	99.593±0.096	99.258±0.150	5.465±0.119
4G	PNN	7176±55	20±5	40±6	7764±56	99.738±0.060	99.450±0.088	99.599±0.044	99.197±0.088	98.943±0.169	99.453±0.123	25.089±0.239
	RF	7212±55	14±4	4±2	7771±55	99.824±0.048	99.942±0.024	99.881±0.027	**99.761±0.054**	99.889±0.047	99.634±0.100	5.976±0.153
2L	PNN	7155±53	86±8	72±9	7687±54	98.891±0.104	98.997±0.121	98.942±0.066	97.882±0.132	98.061±0.231	97.703±0.209	22.971±0.521
	RF	7186±51	68±7	42±5	7705±51	99.123±0.088	99.424±0.069	99.268±0.046	**98.534±0.093**	98.885±0.133	98.186±0.181	5.047±0.170
3L	PNN	7197±52	18±4	25±5	7760±52	99.773±0.051	99.647±0.070	99.713±0.037	99.424±0.073	99.320±0.135	99.529±0.105	25.087±0.484
	RF	7210±52	17±4	13±4	7760±52	99.780±0.046	99.824±0.055	99.801±0.033	99.602±0.066	99.660±0.107	99.544±0.094	5.710±0.178
4L	PNN	7178±61	23±4	31±4	7769±62	99.708±0.053	99.573±0.062	99.643±0.040	99.285±0.080	99.179±0.119	99.392±0.109	26.643±0.185
	RF	7196±61	15±4	13±3	7776±62	99.802±0.046	99.818±0.045	99.809±0.034	99.618±0.067	99.649±0.088	99.588±0.095	6.504±0.134

(b)

		TP	*FP*	*FN*	*TN*	*Sp* (%)	*Se* (%)	*Ac* (%)	*KC* (%)	KC_1 (%)	KC_2 (%)	*Time* (s)
2G	PNN	6551±58	310±17	318±17	7821±54	96.192±0.210	95.365±0.261	95.814±0.136	91.566±0.276	91.460±0.447	91.675±0.436	22.04±0.093
	RF	6680±55	191±16	189±12	7940±54	97.657±0.195	97.253±0.181	97.472±0.098	94.908±0.198	94.932±0.318	94.885±0.412	5.141±0.074
3G	PNN	6771±49	76±8	100±10	8053±53	99.066±0.102	98.543±0.140	98.826±0.088	97.635±0.176	97.319±0.259	97.954±0.220	22.696±0.122
	RF	6841±51	28±5	30±4	8101±51	99.651±0.064	99.559±0.060	99.609±0.041	99.213±0.083	99.187±0.111	99.239±0.139	5.848±0.120
4G	PNN	6827±58	34±6	46±6	8092±58	99.576±0.068	99.327±0.088	99.462±0.058	98.916±0.117	98.761±0.161	99.073±0.147	25.762±0.357
	RF	6856±58	13±4	18±4	8114±58	99.844±0.047	99.745±0.060	99.799±0.037	**99.594±0.074**	99.530±0.110	99.659±0.103	6.630±0.180
2L	PNN	6634±47	211±17	243±17	7912±52	97.406±0.205	96.467±0.234	96.975±0.159	93.907±0.319	93.502±0.428	94.317±0.436	23.007±0.234
	RF	6689±47	169±15	188±15	7954±50	97.921±0.179	97.269±0.216	97.622±0.116	**95.210±0.234**	94.969±0.393	95.454±0.383	5.905±0.136
3L	PNN	6836±52	25±5	38±6	8100±52	99.687±0.061	99.448±0.081	99.578±0.055	99.149±0.112	98.982±0.149	99.317±0.132	25.147±0.217
	RF	6848±51	24±5	26±5	8102±52	99.703±0.056	99.615±0.079	99.663±0.052	99.321±0.104	99.290±0.146	99.353±0.121	5.790±0.132
4L	PNN	6840±52	20±4	31±6	8109±53	99.752±0.054	99.545±0.088	99.657±0.052	99.310±0.104	99.162±0.164	99.458±0.119	26.985±0.413
	RF	6849±53	13±4	22±4	8116±53	99.834±0.045	99.677±0.059	99.762±0.035	99.521±0.071	99.404±0.107	99.638±0.098	6.964±0.260

(c)

		TP	*FP*	*FN*	*TN*	*Sp* (%)	*Se* (%)	*Ac* (%)	*KC* (%)	KC_1 (%)	KC_2 (%)	*Time* (s)
2G	PNN	779±23	38±6	46±7	14137±24	99.729±0.041	94.481±0.852	99.440±0.066	94.589±0.644	94.163±0.898	95.024±0.726	24.244±0.151
	RF	785±24	12±3	39±6	14164±24	99.915±0.023	95.265±0.794	99.660±0.044	96.672±0.451	95.000±0.834	98.410±0.419	5.424±0.135
3G	PNN	794±27	16±4	31±5	14159±28	99.890±0.031	96.260±0.645	99.690±0.044	96.993±0.414	96.047±0.680	97.963±0.554	24.666±0.513
	RF	802±27	4±2	22±5	14171±28	99.969±0.015	97.274±0.577	99.821±0.030	98.256±0.291	97.120±0.609	99.422±0.271	6.809±0.297
4G	PNN	807±25	7±2	20±4	14166±25	99.951±0.016	97.641±0.519	99.824±0.032	98.293±0.310	97.505±0.547	99.096±0.301	28.772±0.742
	RF	805±25	4±2	22±5	14169±25	99.969±0.017	97.337±0.601	99.824±0.033	98.293±0.328	97.185±0.634	99.430±0.322	10.549±0.475
2L	PNN	798±23	17±4	27±5	14157±24	99.879±0.028	96.681±0.579	99.703±0.043	97.124±0.417	96.490±0.612	97.769±0.513	24.338±0.109
	RF	791±24	10±3	35±6	14165±23	99.932±0.019	95.774±0.677	99.703±0.042	**97.102±0.421**	95.536±0.713	98.724±0.353	6.950±0.193
3L	PNN	792±23	14±3	30±5	14164±22	99.902±0.024	96.371±0.632	99.709±0.042	97.158±0.423	96.165±0.664	98.175±0.447	27.035±0.194
	RF	780±23	7±3	43±7	14171±21	99.952±0.021	94.812±0.868	99.670±0.051	96.745±0.529	94.526±0.911	99.077±0.409	7.850±0.235
4L	PNN	797±26	8±3	24±6	14171±28	99.943±0.019	97.102±0.639	99.787±0.038	97.926±0.356	96.938±0.675	98.938±0.344	29.420±0.474
	RF	800±26	4±2	21±5	14175±28	99.969±0.014	97.505±0.611	99.834±0.036	**98.379±0.339**	97.364±0.646	99.418±0.259	9.072±0.314

(d)

		TP	*FP*	*FN*	*TN*	*Sp* (%)	*Se* (%)	*Ac* (%)	*KC* (%)	KC_1 (%)	KC_2 (%)	*Time* (s)
2G	PNN	2036±37	31±5	44±6	12888±36	99.760±0.041	97.864±0.289	99.497±0.059	97.889±0.256	97.523±0.332	98.258±0.298	27.847±0.404
	RF	2045±36	20±4	35±6	12899±36	99.846±0.031	98.296±0.295	99.631±0.050	98.449±0.209	98.024±0.342	98.879±0.222	7.800±0.350
3G	PNN	2045±37	20±5	41±5	12895±37	99.849±0.036	98.030±0.223	99.596±0.049	98.305±0.207	97.715±0.258	98.902±0.265	28.558±0.659
	RF	2067±37	9±3	19±5	12905±37	99.929±0.021	99.083±0.220	99.811±0.037	99.210±0.156	98.936±0.256	99.485±0.151	8.734±0.413
4G	PNN	2047±31	11±3	46±6	12896±33	99.918±0.021	97.817±0.280	99.625±0.043	98.426±0.174	97.470±0.326	99.401±0.155	32.257±0.601
	RF	2084±33	13±4	9±3	12894±32	99.901±0.028	99.590±0.153	99.857±0.026	**99.407±0.109**	99.524±0.178	99.290±0.200	10.245±0.411
2L	PNN	2065±43	24±5	32±5	12878±44	99.812±0.036	98.468±0.243	99.624±0.045	98.433±0.187	98.220±0.282	98.648±0.255	26.997±0.131
	RF	2064±43	19±5	34±6	12883±44	99.851±0.038	98.386±0.276	99.646±0.050	**98.523±0.205**	98.126±0.320	98.924±0.274	6.867±0.193
3L	PNN	2061±42	23±5	23±4	12893±42	99.823±0.040	98.915±0.196	99.697±0.044	98.735±0.185	98.740±0.226	98.730±0.287	29.638±0.188
	RF	2062±40	18±4	22±4	12898±41	99.857±0.031	98.952±0.207	99.731±0.039	98.876±0.163	98.783±0.241	98.971±0.226	7.547±0.286
4L	PNN	2077±39	20±4	19±4	12884±40	99.846±0.034	99.080±0.198	99.739±0.036	98.913±0.148	98.931±0.230	98.895±0.244	32.399±0.361
	RF	2071±40	17±4	25±5	12887±40	99.868±0.029	98.814±0.239	99.721±0.038	98.838±0.160	98.622±0.277	99.054±0.209	10.480±0.362

TABLE 8. *(Continued.)* **(e) The results of classification for the EB and ET subjects with different models. (f) The results of classification for the Healthy, EB and ET subjects with different models.**

(e)

		TP	*FP*	*FN*	*TN*	*Sp* (%)	*Se* (%)	*Ac* (%)	*KC* (%)	KC_1 (%)	KC_2 (%)	*Time* (s)
2G	PNN	518±18	7±3	8±3	1467±19	99.515±0.177	98.446±0.527	99.234±0.171	98.023±0.435	97.893±0.709	98.158±0.654	0.699±0.009
	RF	520±19	5±2	6±2	1469±19	99.682±0.133	98.951±0.459	99.490±0.142	**98.684±0.360**	98.578±0.623	98.793±0.491	0.693±0.019
3G	PNN	519±20	8±3	9±3	1464±20	99.434±0.23	98.222±0.584	99.115±0.244	97.720±0.623	97.587±0.788	97.859±0.854	0.748±0.011
	RF	522±19	5±2	6±2	1467±19	99.678±0.165	98.820±0.459	99.451±0.173	98.585±0.449	98.399±0.623	98.776±0.639	0.882±0.022
4G	PNN	519±19	5±2	10±3	1467±19	99.694±0.148	98.061±0.606	99.264±0.180	98.098±0.474	97.379±0.805	98.832±0.560	0.844±0.011
	RF	523±19	3±2	6±2	1468±19	99.785±0.133	98.918±0.443	99.556±0.135	**98.857±0.346**	98.532±0.597	99.186±0.494	1.072±0.036
2L	PNN	518±19	10±3	11±3	1461±19	99.329±0.198	97.895±0.523	98.948±0.219	97.294±0.557	97.138±0.712	97.454±0.748	0.725±0.015
	RF	522±20	7±2	8±2	1464±20	99.551±0.17	98.525±0.449	99.280±0.179	98.148±0.460	97.997±0.607	98.302±0.629	0.756±0.019
3L	PNN	515±15	8±3	10±3	1467±16	99.451±0.218	98.128±0.573	99.104±0.225	97.681±0.580	97.467±0.772	97.903±0.819	0.797±0.012
	RF	516±16	5±3	8±3	1470±15	99.642±0.17	98.462±0.568	99.333±0.202	98.271±0.528	97.921±0.762	98.629±0.651	0.973±0.026
4L	PNN	511±20	3±2	14±3	1472±20	99.802±0.116	97.357±0.491	99.161±0.146	97.817±0.385	96.444±0.646	99.232±0.445	0.899±0.017
	RF	518±19	3±2	7±2	1471±19	99.77±0.111	98.685±0.390	99.485±0.137	98.666±0.356	98.222±0.524	99.116±0.426	1.152±0.035

(f)

		KC (%)	KC_1 (%)	KC_2 (%)	*Time* (s)	*KC* (%)
2G	PNN	96.802±0.208	96.570±0.317	94.509±0.872	98.003±0.281	30.070±0.493
	RF	97.931±0.229	97.220±0.375	98.344±0.632	98.695±0.271	9.255±0.441
3G	PNN	97.804±0.236	97.299±0.382	97.240±0.635	98.689±0.321	31.294±0.346
	RF	98.852±0.158	98.567±0.255	99.041±0.355	99.146±0.223	10.654±0.294
4G	PNN	98.253±0.193	97.534±0.349	98.466±0.518	99.109±0.234	34.791±0.292
	RF	**99.045±0.123**	99.023±0.194	99.242±0.337	99.000±0.214	13.599±0.330
2L	PNN	97.743±0.199	97.825±0.308	96.648±0.709	98.070±0.322	29.860±0.935
	RF	**97.993±0.172**	97.435±0.343	98.325±0.520	98.591±0.257	9.643±0.640
3L	PNN	98.177±0.211	98.287±0.284	97.250±0.556	98.397±0.301	32.386±0.127
	RF	98.257±0.192	97.745±0.337	99.094±0.357	98.602±0.258	10.298±0.194
4L	PNN	98.576±0.181	98.448±0.259	98.975±0.448	98.588±0.298	35.713±1.035
	RF	98.647±0.173	98.451±0.294	99.426±0.300	98.601±0.247	13.105±0.689

Therefore, if we only consider the classification accuracy, the 4G-model is the best choice. However, time-consumption is also a very important factor in practical applications. Thus, the 2L-model may be a good choice because it has the *KC* of 98.534 ± 0.093% and the least modeling time.

Table 8(b), Table 8(c), Table 8(d), Table 8(e), and Table 8(f) display the classification results of males vs females, healthy subjects vs patients with EB, healthy subjects vs patients with ET, patients with EB vs patients with ET, and healthy subjects vs patients with EB vs patients with ET. The performance of most PNN-based classifiers is worse than that of RF-based classifiers, and only three of the results are opposite, i.e., the classification of 2L- and 3L-models in Table 8(c), and the classification of 4L-model in Table 8(d). There is no much difference in the *KC* between two classification methods for the 4L-model, and their average difference of *KC* is just 0.075%. The time PPN-based classifiers taken is still more than that of RF-based classifiers.

Meanwhile, the accuracies and time consumption of the most classifiers also increase as more kernel functions used, and only the 3L-model in Table 8(c), 4L-model in Table 8(d), and 3G-model in Table 8(e) are unusual. But it does not matter, and the *KC* of them is still over 96%, which does not have a bad influence on our optimal modeling method selecting. The highest *KC* from Table 8(b) to Table 8(f) are 99.594 ± 0.074%, 98.379 ± 0.339%, 99.407 ± 0.109%, 98.857 ± 0.346%, and 99.045 ± 0.123%, which correspond to the 4G-, 4L-, 4G-, 4G-, and 4G-models in each classification, respectively. The 4L-model is in the classification between healthy subjects and patients with EB. While when the modeling time is taken consider for each kind of classification, the most promising models in practical application from Table 8(b) to Table 8(f) are 2L-, 2L-, 2L-, 2G-, and 2L-models, whose *KC* are 95.210 ± 0.234%, 97.102 ± 0.421%, 98.523 ± 0.205%, 98.684 ± 0.360%, and 97.993 ± 0.172%, respectively. The 2G-model is in the classification between patients with EB and ET.

Table 9 presents the statistical results of the classification for different models. The '*mKC*', '*mCTime*' and '*mRMSE*'

TABLE 9. **The statistical results of classification and ABP modeling.**

		mKC (%)	*mCTime* (s)	*mRMSE*	*MTime* (s)
2G	PNN	95.974±2.303	21.252±9.640	0.192±0.060	4216.859
	RF	97.496±1.360	5.518±2.697		
3G	PNN	97.818±0.587	21.772±9.931	0.101±0.040	5620.965
	RF	98.923±0.471	6.399±3.052		
4G	PNN	98.530±0.471	24.586±11.177	0.053±0.025	6322.197
	RF	**99.160±0.537**	8.012±4.036		
2L	PNN	97.064±1.513	21.316±9.548	0.116±0.042	3805.274
	RF	**97.585±1.203**	5.861±2.705		
3L	PNN	98.388±0.861	23.348±10.426	0.077±0.033	5284.429
	RF	98.512±0.988	6.361±2.866		
4L	PNN	98.638±0.643	25.343±11.402	0.051±0.030	6810.224
	RF	98.945±0.515	7.880±3.754		

are the mean of *KC*, the mean time consumption of the different classification, and the mean of *RMSE*, respectively. The '*MTime*' is the sum of the ABP wave modeling time of healthy subjects and patients with EB and ET. We can obtain some conclusions as follows:

(1) The modeling accuracy increases with the increase of kernel function, and the Lognormal-function-based models have less *mRMSE* than the Gaussian-function-based models under the same number of kernel functions.
(2) The RF-based classifiers have better performance than the PNN-based classifiers. The *mKC* of PNN-based classifiers is less than that of the RF-based classifiers, while they take more time in training and testing classifiers.
(3) The classifying accuracy increases with the increase of kernel function, while the time consumption in modeling, classifiers' training, and testing also increases. Especially for the 4G-model and 4L-model, the modeling time is even over 6000 s.
(4) Classifying accuracy is not directly proportional to the modeling accuracy. Though we can see that the *mKC* increases with the increase of the same kernel functions, and the *mRMSE* decreases. However, we found that this conclusion is wrong when we compare the *mRMSE* and *mKC* of the models between two different kernel functions. For example, the *mKC* of RF-based classifiers for 3G- and 4G-models are higher than that 3L- and 4L-model, i.e., $98.923 \pm 0.471\%$ vs $98.512 \pm 0.988\%$ and $99.160 \pm 0.537\%$ vs $98.945 \pm 0.515\%$, while their *mRMSE* are higher than 3L- and 4L-model, i.e., 0.101 ± 0.040 vs 0.077 ± 0.033 and 0.053 ± 0.025 vs 0.051 ± 0.030.
(5) The 2L-model seems to be cost-effective in the ABP-related classification, though the 4G-model based classification has the best performance. The *mKC* of 4G-model is $99.160 \pm 0.537\%$, while that of 2L-model is $97.585 \pm 1.203\%$. The average accuracy only increases by 1.575%, but we have to pay nearly twice as much as time in modeling, training and testing the classifiers and store 6 extra variables. The results indicate the information we extracted from the models with three and four kernel function is redundant. Therefore, the 2L-model has greater potential in practical applications.

IV. DISCUSSION

In this work, we explored the optimal morphological model for ABP-related classification by the principle as shown in Fig. 1. For two popular ABP wave models with the kernel functions mixtures of Gaussian or Lognormal, as shown in (3) and (4), we compare the modeling and classification performance of the models with different kernel functions mixture and different order. For the results of ABP wave modeling, we found the accuracy grows with the increase of kernel functions, whereas more time is consumed. The models of Lognormal function have higher precision than the models of Gaussian function under the same model order, and the lognormal function is more suitable for ABP wave modeling. This conclusion is consistent with [21] and [22].

The two-sample *ks*-test results show that the parameters of all models are markedly different at a highly significant level ($h = 1, p < 0.001$, as shown in Table 6(a) to Table 7(b)). Thus, all the parameters are employed as the feature vectors to train the classifiers using PNN and RF. The results of different classification are illustrated in Table 8(a) to 8(f), and the average results are shown in Table 9. We concluded that the classifying accuracy is not proportional to the modeling accuracy, and the 4G-models has the highest classification accuracy. However, the accuracy is most important but not the only index we considered, and the time-consumption and the number of variables are the other indexes that should be addressed. Then, the 2L-model is acceptable in practical applications. Table 9 presents that the accuracy of the 4G-model related classification is only 1.575% high than that of 2L-model, while the modeling time increase by 2516.923 s since 6 model parameters is added with the increase of model order. The *mKC* of 2L-model related classifications is still $97.585 \pm 1.203\%$. Therefore, the increase of the model order will inevitably lead to information redundancy in classification.

Therefore, we explored the relationship between the features' number and the classifiers' performance. For 2L-model, there are 45 combinations of the features, as shown in Table 10, where '$\circ$' means the feature in this row is selected as an element of a feature vector. The combination is numbered as *Nu*, *Nu* is the integer ranging from 1 to 45 for the 2L-model. Combination 1 means a feature vector contains one parameter $\boldsymbol{A}_1, \ldots$, combination 9 means a feature vector contains all the parameters, $\ldots$, and combination 45 means a feature vector contains ***PPI***. Similarly, there are 120 combinations constructed by model parameters of 4G-model. The numbering process is similar to Table 10 and we were not about to repeat.

Table 10(a) shows the 2L-models related classification results of Ht vs EB vs ET. There are 45 results corresponding to different parameter combinations, and 11 combinations of them whose *KC* is over 97% (combinations 5 to 9, 14 to 17, 24, and 30). The combination 9 has the best performance with the *KC* of $98.016 \pm 0.223\%$ and contains all features. The combinations 5 and 14 contain the fewest features, and their *KC* are $97.382 \pm 0.201\%$ and $97.582 \pm 0.237\%$, respectively. The combination 14 only consists of $\boldsymbol{B}_1, \boldsymbol{C}_1, \boldsymbol{A}_2, \boldsymbol{B}_2$, and $\boldsymbol{C}_2$. The mean of *KC* increases by only 0.207% from combination 6 to 9, which indicates the information redundancy exists. Combination 6 consists of $\boldsymbol{A}_1, \boldsymbol{B}_1, \boldsymbol{C}_1, \boldsymbol{A}_2, \boldsymbol{B}_2$, and $\boldsymbol{C}_2$. When the $\boldsymbol{K}$ is added into combination 6, we obtain the combination 7 with the *KC* of $97.844 \pm 0.220\%$, and the mean of *KC* increases by only 0.035%. However, the mean of *KC* for combination 8 decreases after adding $\boldsymbol{B}$ into combination 7, which means the $\boldsymbol{B}$ has a negative impact on the classification. After adding ***PPI*** into combination 8,

TABLE 10. Different combinations of features. '×' means null, '○' means the features in this row is selected. (a) The results of 2L-models related classification for Ht vs EB vs ET under different feature vectors. (b) The results of 4G-model related classification for Ht vs EB vs ET under different feature vectors.

P	*Nu*															
	1	2	3	⋯	8	9	10	11	⋯	17	18	19	⋯	24	⋯	45
A_1	○	○	○	⋯	○	○	×	×	⋯	×	×	×	⋯	×	⋯	×
B_1	×	○	○	⋯	○	○	○	○	⋯	○	×	×	⋯	×	⋯	×
C_1	×	×	○	⋯	○	○	×	○	⋯	○	○	○	⋯	○	⋯	×
A_2	×	×	×	⋯	○	○	×	×	⋯	○	×	○	⋯	○	⋯	×
B_2	×	×	×	⋯	○	○	×	×	⋯	○	×	×	⋯	○	⋯	×
C_2	×	×	×	⋯	○	○	×	×	⋯	○	×	×	⋯	○	⋯	×
K	×	×	×	⋯	○	○	×	×	⋯	○	×	×	⋯	○	⋯	×
B	×	×	×	⋯	○	○	×	×	⋯	○	×	×	⋯	○	⋯	×
PPI	×	×	×	⋯		○	×	×	⋯	○	×	×	⋯	○	⋯	○

(a)

Nu	*KC* (%)	*Nu*	*KC* (%)	*Nu*	*KC* (%)
1	26.235±0.788	16	**97.636±0.207**	31	83.627±0.526
2	62.898±0.940	17	**97.951±0.204**	32	91.093±0.428
3	82.096±0.606	18	22.647±0.856	33	96.044±0.286
4	94.284±0.320	19	66.375±0.714	34	23.736±0.744
5	**97.382±0.201**	20	91.472±0.475	35	53.237±0.757
6	**97.809±0.231**	21	94.713±0.332	36	73.463±0.524
7	**97.844±0.220**	22	95.020±0.288	37	92.222±0.352
8	**97.796±0.225**	23	96.217±0.283	38	26.750±0.906
9	**98.016±0.223**	24	**97.318±0.267**	39	57.745±0.724
10	29.772±0.928	25	23.901±0.716	40	85.843±0.501
11	61.931±0.826	26	82.581±0.647	41	22.500±0.758
12	87.540±0.462	27	90.066±0.466	42	83.042±0.547
13	96.823±0.253	28	92.471±0.392	43	75.180±0.692
14	**97.582±0.237**	29	95.487±0.309	44	83.627±0.526
15	**97.629±0.220**	30	**97.047±0.230**	45	91.093±0.428

(b)

Nu	*KC* (%)	*Nu*	*KC* (%)	*Nu*	*KC* (%)	*Nu*	*KC* (%)	*Nu*	*KC* (%)	*Nu*	*KC* (%)
1	20.523±0.656	21	**97.324±0.198**	41	**98.831±0.144**	61	**98.505±0.168**	81	**98.169±0.182**	101	81.138±0.456
2	70.661±0.659	22	**98.168±0.162**	42	**98.960±0.150**	62	**98.652±0.167**	82	**98.197±0.193**	102	91.375±0.397
3	88.295±0.486	23	**98.486±0.136**	43	10.602±0.623	63	**98.644±0.158**	83	**98.207±0.212**	103	92.805±0.400
4	94.378±0.353	24	**98.538±0.170**	44	72.733±0.674	64	**98.653±0.155**	84	**98.470±0.193**	104	94.678±0.367
5	**97.024±0.252**	25	**98.736±0.155**	45	89.713±0.375	65	**98.760±0.174**	85	40.259±0.863	105	96.598±0.310
6	**97.659±0.198**	26	**98.831±0.157**	46	95.013±0.298	66	45.347±0.671	86	75.333±0.435	106	51.918±0.708
7	**97.965±0.190**	27	**98.846±0.143**	47	**97.501±0.205**	67	74.055±0.625	87	86.420±0.391	107	77.713±0.535
8	**98.600±0.145**	28	**98.862±0.146**	48	**98.097±0.183**	68	93.912±0.315	88	96.476±0.301	108	84.041±0.536
9	**98.719±0.137**	29	**98.928±0.132**	49	**98.307±0.186**	69	96.300±0.265	89	**97.384±0.238**	109	91.668±0.402
10	**98.798±0.166**	30	46.639±0.846	50	**98.630±0.147**	70	**97.155±0.244**	90	**97.552±0.232**	110	95.916±0.296
11	**98.901±0.154**	31	72.566±0.839	51	**98.728±0.146**	71	**98.269±0.168**	91	**97.777±0.219**	111	23.252±0.825
12	**98.953±0.140**	32	93.028±0.380	52	**98.760±0.147**	72	**98.493±0.164**	92	**98.135±0.189**	112	54.490±0.870
13	**99.011±0.149**	33	95.814±0.241	53	**98.772±0.149**	73	**98.532±0.159**	93	31.440±0.750	113	72.331±0.753
14	**98.981±0.150**	34	**97.075±0.210**	54	**98.862±0.165**	74	**98.522±0.159**	94	62.714±0.664	114	92.778±0.423
15	**99.054±0.128**	35	**98.084±0.156**	55	37.446±0.747	75	**98.634±0.175**	95	92.642±0.352	115	26.984±0.761
16	36.106±0.698	36	**98.490±0.154**	56	76.904±0.599	76	5.954±0.688	96	96.080±0.270	116	57.781±0.895
17	68.053±0.845	37	**98.567±0.150**	57	87.427±0.454	77	77.155±0.722	97	96.396±0.295	117	85.727±0.556
18	84.369±0.586	38	**98.705±0.152**	58	96.497±0.218	78	90.816±0.432	98	96.917±0.253	118	22.253±0.756
19	95.325±0.295	39	**98.799±0.153**	59	**97.656±0.200**	79	94.168±0.307	99	**97.444±0.256**	119	82.891±0.592
20	96.649±0.228	40	**98.840±0.138**	60	**97.971±0.173**	80	**97.759±0.186**	100	11.660±0.723	120	75.063±0.715

we get the combination 9 whose *KC* increases by 0.22%. Thus, the optimal feature combination should contain $\boldsymbol{A}_1$, $\boldsymbol{B}_1$, $\boldsymbol{C}_1$, $\boldsymbol{A}_2$, $\boldsymbol{B}_2$, $\boldsymbol{C}_2$, and $\boldsymbol{PPI}$.

Table 10(b) is the 4G-model related classification results of Ht vs EB vs ET. Half of the combinations' *KC* are over 97% (combinations 5 to 15, 21 to 29, 34 to 42, 47 to 54, 59 to 65, 70 to 75, 80 to 84, 89 to 92, and 99), which are bold in table. The combinations 5, 34, 47, 59, 70, 80, and 89 contain only 5 features. The combination 80 consists of $\boldsymbol{A}_3$, $\boldsymbol{B}_3$, $\boldsymbol{C}_3$, $\boldsymbol{A}_4$, and $\boldsymbol{B}_4$ with the highest *KC* of 97.759 ± 0.186%. Then, the *KC* of the combinations increases with the increase of features. For example, the mean *KC* increases by only 0.454% from combination 8 to 15, while the number of features increases by 7. This means that we spend more time identifying the parameters, but the accuracy of the classification is not significantly increased. The same cases occur in the combinations 22 to 29, the combinations 35 to 42, the combinations 48 to 54, the combinations 60 to 65, the combinations 71 to 75, and the combinations 81 to 84. There is excessive information redundancy in the classification based on 4G-model. Therefore, the 2L-model has greater potential in practical applications.

We can infer from the experimental results that the two types of modeling mixtures can accurately describe the change of ABP waves. The features extracted from the models are effective in ABP signal related classifications, and the lowest *mKC* is over 95%, as shown in Table 9. Here, the experimental data are from the databases *Fantasia* and *2015 PhysioNet/CinC Challenge*, and the ABP signals are recorded from 74 subjects (20 healthy subjects, 17 subjects with EB, and 37 subjects with ET). One of the critical reasons for the high performance of the classifiers is a limited dataset, which leads to overfitting. In the future, we would collect more clinical data to build a 'big dataset' for classifiers training with deep learning. The ABP wave modeling method can be engaged as a feature extractor in deep neural networks. Moreover, there are three other arrhythmias, ventricular tachycardia, ventricular flutter, and ventricular fibrillation in the database. We would improve the 2L-model related classification method to detect the three arrhythmias in the future.

V. CONCLUSION

In this work, we explored the optimal morphological model for ABP signal related classification. Two kinds of most popular modeling methods with Gaussian or Lognormal kernel function mixtures are engaged to fit the measured ABP waves of different subjects who are from the Fantasia and 2015 PhysioNet/CinC Challenge databases. The 2G-, 3G-, 4G-, 2L-, 3L- and 4L-models of the young, old, male, female, and healthy subjects and the patients with EB and ET are obtained, and the modeling time consumption and accuracy are compared. The results show that the models with Lognormal function mixtures are more suitable for ABP wave modeling than the models with Gaussian function mixtures. Then, the parameters of different models are exploited to train the classifiers by PNN and RF, by which we classify the ABP waves according to the age and gender of healthy subjects and whether the subjects are healthy or the patients with EB or ET. We compare the performance of classifications and the results of ABP modeling. The results indicate that the classification performance is not directly proportional to the modeling accuracy, and the 4G-model based classification has the best performance, while the 2L-model has greater potential in practical applications. In the future, we will apply the ABP modeling method to more data and explore its practical applications in m-health.

REFERENCES

[1] D. Martin-Martinez, P. Casaseca-de-la-Higuera, M. Martin-Fernandez, and C. Alberola-Lopez, "Stochastic modeling of the PPG signal: A synthesis-by-analysis approach with applications," *IEEE Trans. Biomed. Eng.*, vol. 60, no. 9, pp. 2432–2441, Sep. 2013.

[2] N. Paradkar and S. R. Chowdhury, "Coronary artery disease detection using photoplethysmography," in *Proc. 39th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Seogwipo, South Korea, Jul. 2017, pp. 100–103.

[3] C. Liu, D. Zheng, L. Zhao, and C. Liu, "Gaussian fitting for carotid and radial artery pressure waveforms: Comparison between normal subjects and heart failure patients," *Biomed. Mater. Eng.*, vol. 24, no. 1, pp. 271–277, Jan. 2014.

[4] Y. Sun and N. Thakor, "Photoplethysmography revisited: From contact to noncontact, from point to imaging," *IEEE Trans. Biomed. Eng.*, vol. 63, no. 3, pp. 463–477, Mar. 2016.

[5] R. Couceiro, P. Carvalho, R. P. Paiva, J. Henriques, I. Quintal, M. Antunes, J. Muehlsteff, C. Eickholt, C. Brinkmeyer, M. Kelm, and C. Meyer, "Assessment of cardiovascular function from multi-Gaussian fitting of a finger photoplethysmogram," *Physiol. Meas.*, vol. 36, no. 9, pp. 1801–1825, 2015.

[6] B. Sun, C. Wang, X. Chen, Y. Zhang, and H. Shao, "PPG signal motion artifacts correction algorithm based on feature estimation," *Optik*, vol. 176, pp. 337–349, Jan. 2018.

[7] C. Liu, L. Zhao, and C. Liu, "Effects of blood pressure and sex on the change of wave reflection: Evidence from Gaussian fitting method for radial artery pressure waveform," *PLoS ONE*, vol. 9, no. 11, 2014, Art. no. e112895.

[8] D. He, L. Wang, X. Fan, Y. Yao, N. Geng, Y. Sun, L. Xu, and W. Qian, "A new mathematical model of wrist pulse waveforms characterizes patients with cardiovascular disease—A pilot study," *Med. Eng. Phys.*, vol. 48, no. 2017, pp. 142–149, 2017.

[9] S. Banerjee, R. Bailón, J. Lazáro, V. Marozas, P. Laguna, and E. Gil, "A two step Gaussian modelling to assess PPG morphological variability induced by psychological stress," in *Proc. CinC*, Rennes, France, 2017, pp. 1–4.

[10] A. Wang, L. Yang, W. Wen, S. Zhang, G. Gu, and D. Zheng, "Gaussian modelling characteristics changes derived from finger photoplethysmographic pulses during exercise and recovery," *Microvasc. Res.*, vol. 116, pp. 20–25, Mar. 2018.

[11] K. Li, S. Zhang, L. Yang, H. Jiang, D. Hao, L. Zhang, and D. Zheng, "Gaussian modelling characteristics of peripheral arterial pulse: Difference between measurements from the three trimesters of healthy pregnancy," *J. Healthcare Eng.*, vol. 2018, 2018, Art. no. 1308419.

[12] A. M. Alqudah, "An enhanced method for real-time modelling of cardiac related bio-signals using Gaussian mixtures," *J. Med. Eng. Technol.*, vol. 41, no. 8, pp. 600–611, 2017.

[13] M. Sorelli, A. Perrella, and L. Bocchi, "Detecting vascular age using the analysis of peripheral pulse," *IEEE Trans. Biomed. Eng.*, vol. 65, no. 12, pp. 2742–2750, Dec. 2018.

[14] M. Sorelli, A. Perrella, and P. Francia, "Multi-Gaussian decomposition of the microvascular pulse detects alterations in type 1 diabetes," in *Medical Physics and Biomedical Engineering*. Singapore: Springer, 2019, pp. 173–176.

[15] X. He, R. A. Goubran, and X. P. Liu, "Secondary peak detection of PPG signal for continuous cuffless arterial blood pressure measurement," *IEEE Trans. Instrum. Meas.*, vol. 63, no. 6, pp. 1431–1439, Jun. 2014.

[16] T. Tigges, A. Pielmus, M. Klum, A. Feldheiser, O. Hunsicker, and R. Orglmeister, "Model selection for the pulse decomposition analysis of fingertip photoplethysmograms," in *Proc. 39th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Seogwipo, South Korea, Jul. 2017, pp. 4014–4017.

[17] X. Jiang, S. Wei, J. Ji, F. Liu, P. Li, and C. Liu, "Modeling radial artery pressure waveforms using curve fitting: Comparison of four types of fitting functions," *Artery Res.*, vol. 23, pp. 56–62, Sep. 2018.

[18] L. Wang, L. Xu, D. Zhao, Y. Yao, and D. Song, "FPGA-based design and implementation of arterial pulse wave generator using piecewise Gaussian–cosine fitting," *Comput. Biol. Med.*, vol. 59, pp. 142–151, Apr. 2015.

[19] A. Sološenko, A. Petrėnas, V. Marozas, and L. Sörnmo, "Modeling of the photoplethysmogram during atrial fibrillation," *Comput. Biol. Med.*, vol. 81, pp. 130–138, Feb. 2017.

[20] L. Wang, L. Xu, S. Feng, M. Q.-H. Meng, and K. Wang, "Multi-Gaussian fitting for pulse waveform using weighted least squares and multi-criteria decision making method," *Comput. Biol. Med.*, vol. 43, no. 11, pp. 1661–1672, 2013.

[21] T. Tigges, J. Rockstroh, A. Pielmu, M. Klum, A. Feldheiser, O. Hunsicker, and R. Orglmeister, "In-ear photoplethysmography for central pulse waveform analysis in non-invasive hemodynamic monitoring," *Current Directions Biomed. Eng.*, vol. 3, no. 2, pp. 587–590, 2017.

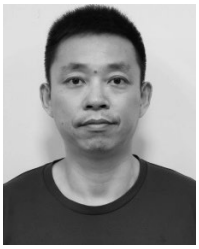
[22] C. Liu, D. Zheng, A. Murray, and C. Liu, "Modeling carotid and radial artery pulse pressure waveforms by curve fitting with Gaussian functions," *Biomed. Signal Process. Control*, vol. 8, no. 5, pp. 449–454, Sep. 2013.

[23] A. L. Goldberger, L. A. N. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C.-K. Peng, and H. E. Stanley, "PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals," *Circulation*, vol. 101, no. 23, pp. e215–e220, 2000.

[24] G. D. Clifford, I. Silva, B. Moody, Q. Li, D. Kella, A. Shahin, T. Kooistra, D. Perry, and R. G. Mark, "The PhysioNet/computing in cardiology challenge 2015: Reducing false arrhythmia alarms in the ICU," in *Proc. CinC*, Nice, France, 2015, pp. 273–276.
[25] S. T. Lawless, "Crying wolf: False alarms in a pediatric intensive care unit," *Crit. Care Med.*, vol. 22, no. 6, pp. 981–985, Jun. 1994.
[26] Y. Chou, B. Xu, Y. Gu, R. Zhang, L. Wang, and Y. Jin, "A fast mathematical morphology filter on one dimensional sampled signal," in *Proc. ICCAIS*, Chiang Mai, Thailand, 2017, pp. 233–238.
[27] Y. D. Teng, L. Ge, and Y. X. Chou, "A novel abnormal segments detection method for photopletysmography signal," in *Proc. CCC*, Wuhan, China, 2018, pp. 4113–4117.
[28] Y. X. Chou, A. H. Zhang, and P. Wang, "Pulse rate variability estimation method based on sliding window iterative DFT and Hilbert transform," *J. Med. Biol. Eng.*, vol. 34, no. 4, pp. 347–355, Aug. 2014.
[29] Z. Chen, Y.-J. Pan, and J. Gu, "Integrated adaptive robust control for multilateral teleoperation systems under arbitrary time delays," *Int. J. Robust Nonlinear Control*, vol. 26, no. 12, pp. 2708–2728, Aug. 2016.
[30] D. F. Specht, "Probabilistic neural networks," *Neural Netw.*, vol. 3, no. 1, pp. 109–118, 1990.
[31] A. Liaw and M. Wiener, "Classification and regression by randomforest," *R News*, vol. 2, no. 3, pp. 18–22, 2002.
[32] U. Farooq, J. Gu, M. El-Hawary, M. U. Asad, and J. Luo, "An extended state convergence architecture for multilateral teleoperation systems," *IEEE Access*, vol. 5, pp. 2063–2079, 2017.
[33] M. S. Islam, W. Khreich, and A. Hamou-Lhadj, "Anomaly detection techniques based on kappa-pruned ensembles," *IEEE Trans. Rel.*, vol. 67, no. 1, pp. 212–229, Mar. 2018.

YONGXIN CHOU was born in Qingyang, China, in 1987. He received the B.S. and Ph.D. degrees from the College of Electrical and Information Engineering, Lanzhou University of Technology, Lanzhou, China, in 2010 and 2015, respectively. In January 2016, he joined the School of Electric and Automatic Engineering, Changshu Institute of Technology, Suzhou, China, where he is currently a Lecturer. His main research interests include physiological signal measurement and analysis, cardiovascular disease detection, and machine learning.

PING WANG received the B.S. and M.S. degrees in control theory and control engineering from the Lanzhou University of Technology, Lanzhou, China, in 2011 and 2014, respectively, and the Ph.D. degree in instrument science and technology from the Nanjing University of Aeronautics and Astronautics, Nanjing, China, in 2019. He is currently a Lecturer with the College of Electrical and Information Engineering, Lanzhou University of Technology. His research interests include computer vision, image processing, machine learning, and signal processing.



YUFENG FENG was born in Changshu, China, in 1973. He received the B.S. degree in clinical medicine from Nanjing Medical University, Nanjing, China, in 1996. In August 1996, he joined Changshu No.1 People's Hospital, Suzhou, China, where he is currently the Chief Physician in ICU. His main research interests include physiological signal measurement and analysis, and cardiovascular disease clinical diagnosis and rescue.