IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received June 17, 2020, accepted June 27, 2020, date of publication July 6, 2020, date of current version July 20, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3007524

Multi-Phases and Various Feature Extraction and Selection Methodology for Ensemble Gradient Boosting in Estimating Respiratory Rate

SOOJEONG LEE[1], (Member, IEEE), CHANG-HWAN SON[2], (Member, IEEE), MARCELO K. ALBERTINI[3], AND HENRIQUE C. FERNANDES[3,4]

[1]Department of Computer Engineering, Sejong University, Seoul 05006, South Korea
[2]Department of Software Convergence Engineering, Kunsan National University, Gunsan 54150, South Korea
[3]School of Computer Science, Federal University of Uberlandia, Uberlandia 38408-100, Brazil
[4]Fraunhofer IZFP Institute for Nondestructive Testing Saarbrucken, 66119 Saarbrucken, Germany

Corresponding author: Soojeong Lee (leesoo86@sejong.ac.kr)

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. 2020R1A2C1010405).

ABSTRACT Estimating the correct respiratory rate (RR) is an essential technique for intensive care units, hospitals, geriatric hospital facilities, and home care services. Capnography is a standard methodology used to monitor carbon dioxide concentrations or partial pressures of respiratory gases to provide the most accurate RR measurements. However, it is inconvenient to use and has been primarily used while administering anesthesia and during intensive care. Many researchers now use electrocardiogram signals to estimate RR. Despite the recent developments, the current hospital environments suffer from inaccurate respiratory monitoring. While various machine learning techniques, including deep learning, have recently been applied to the medical processing sector, only a few studies have been conducted in the field of RR estimation. Therefore, using photoplethysmography, machine-learning techniques such as the ensemble gradient boosting algorithm are being employed in RR estimation. Multi-phases are used based on various feature extraction and selection methodology to improve the performance for RR estimation. In this study, the number of ensembles is increased, and the proposed ensemble methodology is effectively learned to estimate the RR. The proposed ensemble-based gradient boosting algorithm are compared with those of ensemble-based long-short memory network, and ensemble-based supported vector regression techniques, 3.30 breaths per min (bpm), 4.82 bpm and 5.83 bpm based on mean absolute errors. The proposed method shows a more accurate estimate of the respiration rate.

INDEX TERMS Respiration rate estimation, gradient boosting algorithm, ensemble methodology, photolethysmography signals.

I. INTRODUCTION

Respiratory rate (RR) is one of the four most significant physiological signs that represent the critical functional state of a human body and can often be used to predict symptoms of serious illnesses; the other three being blood pressure, body temperature, and uncontrolled heart rate [1-3]. According to a report in [4], a rapid increase in respiratory activity is a crucial indicator of cardiac arrest in hospital wards. Capnography has been known to provide the most precise RR measurements by monitoring the concentration or partial pressure of carbon dioxide in respiratory gases [3]. However, owing to the inconvenient procedure, it is primarily used while administering anesthesia and during intensive care. Currently, the electrocardiogram signal has been used in various ways to quantify effects of breathing. The well-known respiration-related effect is its influence on the variability of heart rate such as respiratory sinus arrhythmia [5]. However, for this work, we are interested only in the respiration rate estimation. Several notable methods have been devised to measure respiratory rate from the pulse oximeter signal [6]–[10]. Thus, pulse oximetry has become an affordable solution for RR estimation. Based on the principle of photoplethysmography (PPG) [3], it is feasible to estimate the level of oxygen saturation (SpO2) in a person by measuring the attenuation of light passing through the tissues in fingers or earlobes [11].

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

The PPG signal is divided into a series of peaks and troughs, which is well represented in the time domain. It uses techniques such as baseline modulation, amplitude modulation, and frequency modulation methods [12]. These algorithms for estimating RR from PPG signals have been developed based on signal processing techniques such as sparse signal reconstruction, continuous wavelet transform [13], and employing an autoregressive model [14]. In addition, sparse signal reconstruction and spectral fusion have been proposed to obtain sparse representations of PPG signals in the spectral domain [15]. P. S. Addison *et al.* developed a continuous wavelet transform [13]. This method estimates the RR in a PPG signal window. However, for this spectral domain method, they hold windows as a whole and neglect the irregularities in the region [3]. Notwithstanding, the spectrum domain technology can degrade the performance if respiratory modulation changes within the window, or only a part of the window is disturbed.

Recently, various technologies based on PPG signal have been applied to RR estimates [13], [16], [17]. However, a few machine-learning algorithms have been used to estimate RR [18], [19]. Vital signs estimation, including RR using generative boosting based on the long short-term memory network (LSTM), was introduced by Liu *et al.* [19]. LSTM is basically a nonlinear time series model that learns nonlinearity from input features. However, these will not work well with a small number of input features because they need to learn nonlinearity, and the training time will be much longer than looking for the linear regression trees. Also, the regression trees can create over-fitting that does not generalize the data well. Greedy algorithms cannot guarantee the return of a globally optimal regression tree, and regression trees can create biased trees if some classes dominate [20]. Meanwhile, the support vector machine (SVM) was used as one of the most successful machine learning algorithms. The nonlinear function (SVM) is mapped to the derived feature space of the high-dimensional kernel using the linear learner [21]. SVM is more effective when the number of dimensions is higher than the number of samples. The RR measurement system was developed using an SVM based on the accelerometer sensor [22]. However, SVM was not applied to the RR estimation using the PPG signal. Therefore, LSTM and SVM regression (SVR) models are used to compare with the proposed method.

In this regard, it aims to devise a solution to the above case, such as the small number of input features and the over-fitting problems. The ensemble-based gradient boosting algorithm (EGBA) based on regression trees with convolutional kernels are applied for RR estimation. The proposed regression trees based on insufficient PPG data have fewer computational requirements, fewer parameters, and are user-friendly. Besides, multi-phases are used based on various feature extraction and selection methodology to improve the performance for RR estimation. In particular, features are extracted from various methods such as the autoregressive (AR) modeling [23], wavelet packet entropy [24], multifractal wavelet leaders (MWL) [25], and maximal overlap discrete wavelet transform (MODWT) [26], [27]. These techniques can be used to extract important features from insufficient PPG signals. Then, feature selection techniques are also used to select high-weighted feature vectors from among various features. As a result, the number of ensembles was increased, and the proposed EGBA efficiently learned to estimate the RR. As per the authors' knowledge, the proposed technique is one of the first studies conducted based on EGBA with the wavelet transformed domain and the autoregressive (AR) modeling for RR estimation. The proposed technique shows the estimation of human RR signals based on EGBA. This study can contribute to field of RR estimation in the following ways:

- It offers a novel method for RR estimation using the EGBA from insufficient PPG measurement data.
- Multi-phases are used based on feature extraction and selection methodology. In detail, features are extracted from various methods such as the autoregressive (AR) modeling [23], wavelet packet entropy [24], multifractal wavelet leaders (MWL) [25], and maximal overlap discrete wavelet transform (MODWT) [26], [27] to solve insufficient data. Then, feature selection techniques are used to choose high-weighted feature vectors from among various features.
- From the computational aspect, it provides a higher approximation of the target RR than any of the previously used single estimators, because it drives the local search from parameters through a base-learner.
- From a statistical point of view, the proposed methodology finds a superior estimator and mitigates the risk of selecting the wrong estimator.
- The LSTM [19] and SVR [22] algorithms are used as benchmark models to show a comparative evaluation of the proposed method.

The study is organized as follows. In Section II, data collection, feature extraction, and feature selection from PPG signals are represented. The proposed ensemble based GBA (EGBA) is presented in Section III. Section IV describes the experimental scenario, a statistical test, and results. Section V discusses the results. The conclusion is presented in Section VI.

II. METHODS

A. DATA COLLECTION AND PREPROCESSING

In this paper, the PPG data were collected by the synthetic dataset derived from peterhcharlton.github.io.[1] The PPG signals were routinely utilized in medical practice to estimate RR, blood pressure (BP), heart rate (HR), and SpO2. The pre-processing stage derived a respiratory signal from the raw PPG signal. In the first step, varying high frequencies were removed using Kaiser window low-pass filters with – 3dB cutoffs of 35 Hz. In the second step, the PPG signal was divided into pulses using an incremental-merge

[1] (http://peterhcharlton.github.io/RRest/synthetic dataset.html) [30]

segmentation (IMS) method [28]. In particular, IMS generated pulse amplitude, the maximum and minimum intensity of each pulse, and pulse period. The signals of irregularly were resampled to 5 Hz using linear interpolation [28]. Finally, very low frequencies were removed using Kaiser window high-pass filters with – 3dB cutoff frequency of 4 breaths per minute (bpm). After pre-processing, the PPG waveforms were obtained from the raw PPG signals, as shown in Fig. 1, where top panel (a) denotes an example of low RR (4 bpm), and bottom panel (b) indicates an example of high RR (42 bpm), respectively. Here, the higher the bpm signal, the more frequent the signal peak change.

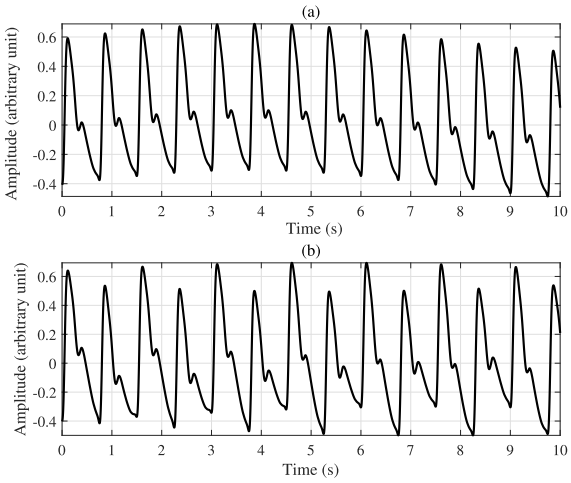

FIGURE 1. Top panel (a) denotes an example (4 bpm) of RR for the first subject and bottom panel (b) denotes the 42 bpm of RR for the 20*th* subject from PPG signals.

Afterward, the PPG signal was dimensionally reduced by transforming the PPG signals into a much smaller set of features such as autoregressive (AR) coefficients, Shannon entropy, wavelet leaders, and maximal overlap discrete wavelet transform (MODWT) as shown in Fig 2. Each feature set was parallel concatenated, and then the feature set was selected using a neighborhood component analysis (NCA) [29]. A set of synthetic datasets consisting of breathing signals obtained was based on PPG and RR reference data received from 192 records (192 adult subjects). The PPG database was composed of 192 records, where each record was 210 (s), and 500Hz was used for sample frequency (Fs). Therefore, 105,000 samples were prepared for each subject. The RR values were determined based on the average time between consecutive breaths in each window. Repetitive experiments were conducted to estimate RR by varying the window size between 16, 32, and 64 s. 32 s window size with the best performance was then selected for the proposed algorithm [30] and [31], which indicated that the RR estimate was based on a 32 s window and was updated every 3 s. This was because the minimum detectable RR for [28] for 32 s was 3.75 (bpm) and the window need to contain signals of at least two cycles so that at least two peaks could be observed [15].

B. FEATURES EXTRACTION BASED ON WAVELET TRANSFORM

Features can be extracted from the wavelet transformed domain and the autoregressive (AR) modeling [23] using the segment PPG signals. Wavelet transformation has been widely used to compensate for the shortcomings of Fourier transform, as it allows the analysis of information on time and frequency axes. The characteristics of this method are used to extract important features from various time-series data. This study uses a parallel combination of autoregressive (AR) modeling [23], wavelet packet entropy [24], multifractal wavelet leaders (MWL) [25], and maximal overlap discrete wavelet transform (MODWT) [26], [27] as shown in Fig 2.

1) AR MODELLING [23]

The signal $x(t)$ at time index t is described as a linear combination of p preceding values of the same signal. The AR process is defined as

$$x(t) = \sum_{n=1}^{p} a(n)x(t-n) + e(t) \tag{1}$$

where $a(n)$ denotes the nth coefficient of the AR model, $e(t)$ denotes a white noise with zero mean, p denotes the AR order, respectively. Note that we use the 4-AR model order, which provides the best fit for PPG signals in a similar RR estimation. The AR model order is selected based on the minimum order to obtain a fitness for each subject [32]. The AR parametric coefficients based on Burg's algorithm [23] are used as features to estimate RR.

2) WAVELET PACKET ENTROPY [24] AND LEADERS [25]

Although WT coefficients represent the local characteristics of PPG signals, the number of such coefficients is generally too large to be used as features for direct estimation. Therefore, for a better estimate, it is necessary to generate some high-level characteristics from these coefficients. Shannon entropy is a method to measure the uncertainty of the information in each system, which is typically used in signal processing. In this work, the Shannon entropy is obtained using the maximal overlap discrete wavelet packet transform (MODWPT) [24]. The wavelet packet decomposition for the signal $x(t)$ is recursively defined as

$$\begin{cases} w_{0,0} = x(t) \\ w_{i,2j-1}(t) = \sqrt{2}\sum_k hi(k) wa_{i-1,j}(2t-k) \\ w_{i,2j}(t) = \sqrt{2}\sum_k lo(k) wa_{i-1,j}(2t-k) \end{cases} \tag{2}$$

where $hi(k)$ and $lo(k)$ denote high-pass and low-pass filters, respectively, and $w_{i,j}$ denotes the reconstruction coefficients of wavelet packet decomposition at the ith level for the jth node. Therefore, the Shannon entropy is calculated based on energy follow equations:

$$\mathrm{En}_{i,j,k} = \|w_{i,j,k}\|^2 \tag{3}$$

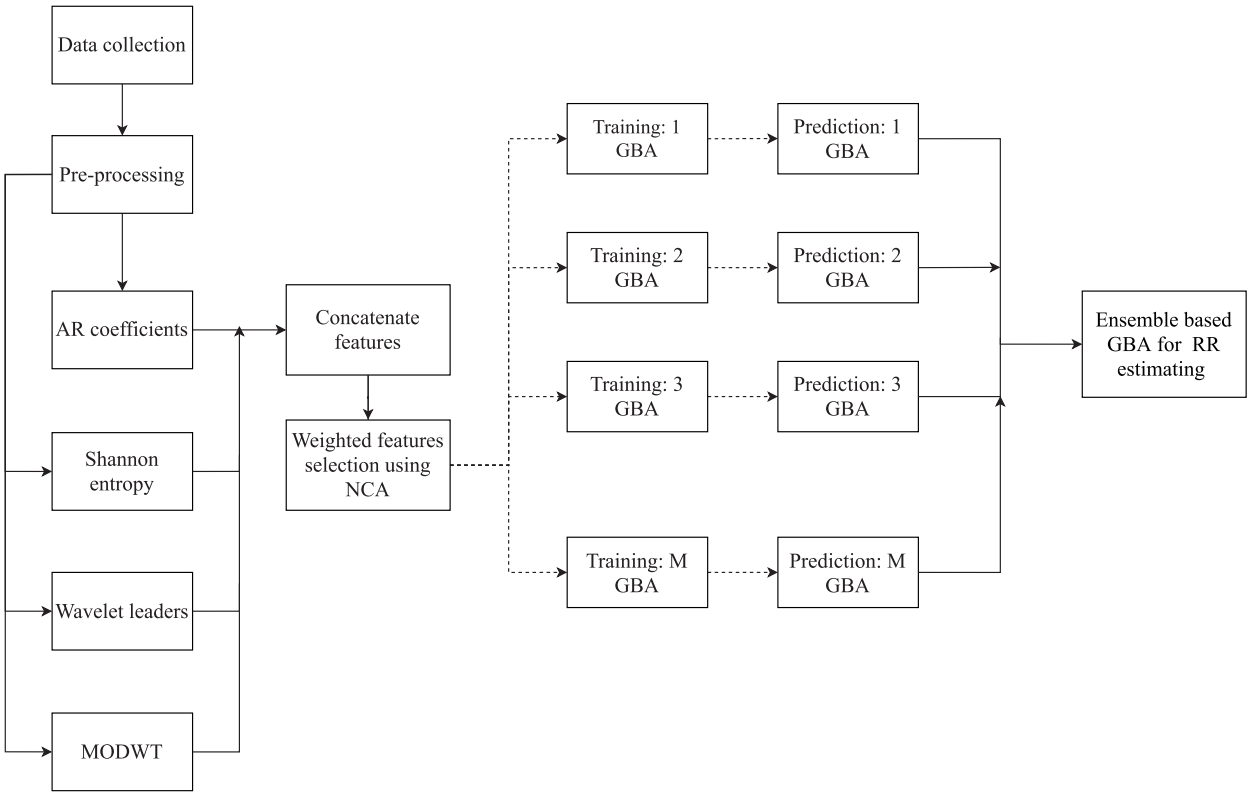

FIGURE 2. Block diagram for RR estimation using the proposed ensemble-based GBA (EGBA).

where $\mathrm{En}_{i,j,k}$ denotes the wavelet energy and k denotes the index of the coefficient. Therefore, the entire energy of the ith level at the jth node is computed as

$$\mathrm{En}_{i,j} = \sum_{k=1}^{K} \mathrm{En}_{i,j,k} \tag{4}$$

where K denotes the number of corresponding coefficients in the node. The probability of the kth corresponding coefficients is calculated as

$$\mathrm{Pr}_{i,j,k} = \frac{\mathrm{En}_{i,j,k}}{\mathrm{En}_{i,j}} \tag{5}$$

where the sum of the probabilities $\mathrm{Pr}_{i,j,k}$ denotes one. Thus, the Shannon entropy is calculated using the probability distribution of energy [33] as

$$\mathrm{Se}_{i,j} = -\sum_{k=1}^{K} \log(\mathrm{Pr}_{i,j,k}) \cdot \mathrm{Pr}_{i,j,k} \tag{6}$$

The MWL is calculated based on the wavelet leaders obtained using the biorthogonal spline wavelet filter [25] that can accurately measure the multifractal spectrum over its full range.

$$[\mathrm{cum}, \mathrm{ho}] = \mathrm{MWL}(x) \tag{7}$$

where cum denotes the first three cumulants characterizing the local maximum of the singularity spectrum, ho is holder exponents.

3) MODWT [27]

In this study, MODWT is also used to replace discrete wavelet transformations (DWT) with shortcoming length limits. Specially, the DWT that requires $N = 2^j$ where j denotes a positive integer. In contrast, the MODWT method has no limitations on the length of the data and can also be useful in decomposition time series data with a constant number of counts on each scale. The MODWT wavelet filter $\widetilde{w}_{j,l}$ and scaling filter $\widetilde{s}_{j,l}$ are computed as $\widetilde{w}_{j,l} = w_{j,l}/2^{j/2}$ and $\widetilde{s}_{j,l} = s_{j,l}/2^{j/2}$, where $w_{j,l}$ and $s_{j,l}$ denote the DTW wavelet filter and scaling filter, respectively. In this study, MODWT allows us to discriminate patterns of multivariate PPG signals from relationships between them based on the variability of individual components of each PPG signal. Totally, we obtain 279 features: 48 AR features (6 coefficients per segment), 192 Shannon entropy (24 features per segment), 24 fractal estimates (3 per segment) acquired using the MWL [25], and 15 wavelet variance estimates.

C. FEATURE SELECTION BASED ON NEIGHBORHOOD COMPONENTS

In this study, neighbor component analysis (NCA) was used to select the high weighted feature vectors in 279 features [29]. Feature weights were learned by a diagonal adaptation of NCA. It learns a feature weighting vector by minimizing an objective function that measures the average leave-one-out regression loss over the training data. The NCA performs

modified feature selection for regression. A set of training samples can be written as $S = \{(x_1, y_1), \ldots, (x_N, y_N), i = 1, \ldots, N\}$. In this case, the goal was to estimate the target vector y given the training set S, where $x_i \in \mathbb{R}^P$ and N denote the number of samples. A randomized regression model was used that randomly selected a point (Rf(x)) in S as the reference point for x. The reference point was chosen to be the nearest neighbor of the new point x. Here, the target value was set at x equal to the target value of the reference point (Rf(x)) [29]. If x_j was closer to x measured by the distance $d_w = \sum_{l=1}^{P} w_l^2 |x_{il} - x_{jl}|$, the probability $\mathbb{P}(\text{Rf}(x) = x_j|S)$, that point x_j was selected in S as the reference point for x was higher. Here, w_l denotes a feature weight with lth feature. It was assumed that $\mathbb{P}(\text{Rf}(x) = x_j|S) \propto \vartheta(d_w(x, x_j))$, where $\vartheta(z) = \exp(-z/\sigma)$ denotes a kernel function, and the kernel width σ denotes an input parameter that influences the probability of each point being chosen as the reference point [29]. Hence, the probability $\mathbb{P}(\text{Rf}(x) = x_j|S)$ can be represented as

$$\mathbb{P}(\text{Rf}(x) = x_j|S) = \frac{\vartheta(d_w(x, x_j))}{\sum_{j=1}^{N} \vartheta(d_w(x, x_j))} \tag{8}$$

We consider the leave-one-out regression, that is, estimation the response for x_i based on the data in S^{-i}, the training set S excluding the point (x_i, y_i). The probability of the point x_j is selected as the reference point for x_i is written as

$$p_{i,j} = \mathbb{P}(\text{Rf}(x_i) = x_j|S^{-i}) = \frac{\vartheta(d_w(x_i, x_j))}{\sum_{j=1, j\neq i}^{N} \vartheta(d_w(x_i, x_j))} \tag{9}$$

$$\Omega_i = \mathbb{E}(\Omega(y_i, \hat{y}_i)|S^{-i}) = \sum_{j=1, j\neq i}^{N} p_{i,j}\Omega(y_i, y_j) \tag{10}$$

where Ω denotes the object function that measures the disagreement between $\hat{y}_i$ and y_i. Therefore, the object function with adding the regularization parameter φ for minimization is follow:

$$f(w) = \frac{1}{N}\sum_{i=1}^{N} \Omega_i + \varphi \sum_{l=1}^{P} w_l^2 \tag{11}$$

After the NCA, the dimension of feature vectors can be significantly reduced from 279 to 13 as shown in Fig. 3. Then, we use a general normalization approach to restrict the interval of the allowed features to lie in between the minimum and maximum.

D. BENCHMARK MODELS

Four benchmark models using machine learning and deep learning to demonstrate the effectiveness of EGBA are represented as follow.

1) SVR [21] AND ENSEMBLE-BASED SVR (ESVR)

Recently, SVR is considered as one of the most successful machine-learning algorithm [21]. The essence of SVR is to map data into a high-dimensional feature space using a nonlinear relationship and then perform linear regression within the space. ESVR is designed to make a balanced comparison with the proposed EGBA.

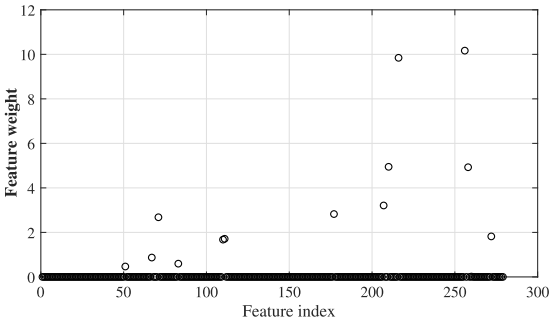

FIGURE 3. Features selection performed using a neighbor component analysis [29].

2) LSTM [19] AND ENSEMBLE-BASED LSTM (ELSTM)

LSTM is capable of learning and remembering over a long sequence of inputs [19]. Therefore, it can overcome the issue of vanishing gradient through memory blocks, which makes LSTM extremely suitable for time-domain data estimations. ELSTM is also constructed to make a balanced comparison to the proposed EGBA.

III. ENSEMBLE-BASED GRADIENT BOOSTING ALGORITHM (EGBA)

A. ESTIMATION FUNCTION

Given machine-learning, supervised learning is a powerful constraint for researchers because appropriate target values have to be provided for the data [34]. Therefore, input data and target data need to be prepared as $(\mathbf{x}_i, y_i)_{i=1}^{N}$. The goal was to acquire an estimate $F^*(\mathbf{x}) \approx y$ and minimize the expected value of the specified cost function $\Omega(y, F(\mathbf{x}))$, given as

$$F^*(\mathbf{x}) = \arg\min_{F(\mathbf{x})} \Omega(y, F(\mathbf{x})) \tag{12}$$

where $F^*(\mathbf{x})$ denotes a boosting approximate which is calculated by $F(\mathbf{x}) = \sum_{m=0}^{M} \alpha_m h(\mathbf{x}, \mathbf{a}_m)$, here $h(\mathbf{x}, \mathbf{a}_m)$ denotes a base-learner, α_m are the coefficients, and $\mathbf{a}_m$ denote the parameters. Target data is a continuous variable, i.e., $y \in R$, we use classical cost function.

$$(\alpha_m, \mathbf{a}_m) = \arg\min_{\alpha,\mathbf{a}} \sum_{i=1}^{N} \Omega(y_i, F_{m-1}(\mathbf{x}_i)) + \alpha h(\mathbf{x}_i, \mathbf{a}) \tag{13}$$

$$F_m(\mathbf{x}) = F_{m-1}(\mathbf{x}) + \alpha_m h(\mathbf{x}, \mathbf{a}_m) \tag{14}$$

where a step-size α is specified in each iteration and $\mathbf{a}_m$ is the mth incremental step of the parameter. In general, given cost function $\Omega(y, f(\mathbf{x}))$ and a base-learner $h(\mathbf{x}, \mathbf{a}_m)$, it is difficult to obtain the parameter estimates. Therefore, we can use the negative gradient $\{g_m(\mathbf{x}_i)\}_{i=1}^{N}$ as

$$-g_m(\mathbf{x}_i) = -\left[\frac{\partial \Omega(y_i, F(\mathbf{x}_i))}{\partial F(\mathbf{x}_i)}\right]_{F(\mathbf{x})=F_{m-1}(\mathbf{x})} \tag{15}$$

TABLE 1. Ensemble-based GBA (EGBA).

Algorithm 1: EGBA

Procedure $(\mathbf{X}, Y)_{i=1}^{N}$
1: $\Omega(y, F(\mathbf{x}))$ cost function
2: $h(\mathbf{x}, \mathbf{a})$ learner model
3: $F_0(\mathbf{x}) = \arg\min_{\beta} \sum_{i=1}^{N} \Omega(y_i, \beta)$
4: **for** $m = 1, M$ do,
5: $\quad -g_m(\mathbf{x}_i) = -[\frac{\partial \Omega(y_i, F(\mathbf{x}_i))}{\partial F(\mathbf{x}_i)}]_{F(\mathbf{x})=F_{m-1}(\mathbf{x})}, \quad i = 1, N$
6: $\quad \mathbf{a}_m = \arg\min_{\alpha, \mathbf{a}} \sum_{i=1}^{N} [-g_m(\mathbf{x}_i) - \alpha h(\mathbf{x}_i, \mathbf{a})]^2$
7: $\quad \beta_m = \arg\min_{\beta} \sum_{i=1}^{N} \Omega(y_i, F_{m-1}(\mathbf{x}_i) + \beta h(\mathbf{x}_i, \mathbf{a}_m))$
8: $\quad F_m(\mathbf{x}) = F_{m-1}(\mathbf{x}) + \beta_m h(\mathbf{x}, \mathbf{a}_m)$
9: **endfor**
10: $F_{\varphi}(\mathbf{x}) = M^{-1} \sum_{m=1}^{M} F_m(\mathbf{x})$
11: **endfor**
end procedure

TABLE 2. Split training.

Algorithm 2: Split training [20]

Procedure Input $(\mathbf{x}_{i=1}^{N})$
1: **for** $l = 1, L$ do, step 1: kernel search
2: $\quad \phi_l = \arg\min_{\phi} \sum_{i \in \Psi_1} w_i \left(\phi^T W_{c_l, a_l}(\mathbf{x}_i) - \theta_i\right) + \rho_l \sum_{(m,n) \in \mathcal{N}} \left(\phi^m - \phi^n\right)^2$ detect kernel
3: **endfor**

4: **for** $l = 1, L$ do, step 2: split search
5: $\quad W_l = W_{c_l, a_l}$
6: $\quad (\tau_l, \zeta_{1,l}, \zeta_{2,l}) = \arg\min_{\tau, \zeta_1, \zeta_1} \sum_{i | \phi_l^T W_l \mathbf{x}_i < \tau} w_i \left(\theta_i - \zeta_1\right)^2 + \sum_{i | \phi_l^T W_l \mathbf{x}_i > \tau} w_i \left(\theta_i - \zeta_2\right)^2$
7: **endfor**
8: return $(\phi_l, \tau_l, \zeta_{1,l}, \zeta_{2,l})$ that produces the smallest ν_l
end procedure

Here, $h(\mathbf{x}, \mathbf{a})$ is the most correlated with $-g_m$ in terms of data distribution. Hence, we can solve as

$$\mathbf{a}_m = \arg\min_{\alpha, \mathbf{a}} \sum_{i=1}^{N} [-g_m(\mathbf{x}_i) - \alpha h(\mathbf{x}_i, \mathbf{a})]^2 \tag{16}$$

Here, the constrained negative gradient $h(\mathbf{x}_i, \mathbf{a})$ is utilized in place of the $-g_m(\mathbf{x}_i)$ in the steepest-descent strategy. Thus, (13) can be defined as (17).

$$\beta_m = \arg\min_{\beta} \sum_{i=1}^{N} \Omega(y_i, F_{m-1}(\mathbf{x}_i) + \beta h(\mathbf{x}_i, \mathbf{a}_m)) \tag{17}$$

The approximation can be updated as

$$F_m(\mathbf{x}) = F_{m-1}(\mathbf{x}) + \beta_m h(\mathbf{x}, \mathbf{a}_m) \tag{18}$$

A regression tree is one of the most used machine-learning methods for generating tree-like structures, where each internal node denotes a feature test, each branch is one of the possible test results, and each node denotes a regression. Regression trees are for dependent variables that take continuous values, with prediction error typically measured using the squared difference between the observed and the predicted values [20], [35]. Here, the EGBA that is used with regression trees [35] to improve the performance of the GBA model [34] has been introduced. The input features are given by $(\mathbf{X}, Y)$, where $\mathbf{X}$ and Y denote the input matrix and the output vector, respectively. The aim of the proposed technique is to obtain multiple estimates as $\{F_1(\mathbf{x}), \ldots, F_M(\mathbf{x})\}$ to overcome the unstable estimations such as high variance and bias of error. The ensemble estimators can be provided in Line 10 of Algorithm 1. First, the building of regression trees to optimize the GBA has been shown [20], [34]. Then, the process to learn the relevant features while growing the trees has been described.

B. CONVOLUTION KERNELS FOR TRAINING

The parameter space of the GBA was not as practical as the grid search [20]. Therefore, we use a kernel in a close form that can handle parameters of a large size. Typically, the GBA searches through the sets of basic learners that rely on a fixed set of features. It chooses h_j that minimizes at each iteration j

$$h_j = \arg\min_{h} \sum_{i=1}^{N} w_i^j [h(\mathbf{x}_i) - \theta_i^j]^2 \tag{19}$$

where w_i^j denotes a weight, and θ_i^j denotes a response value calculated by differentiating the cost function Ω. Basic learners, that were regression trees, using convolutions of $\mathbf{x}$ with a set of learned convolution kernels Φ_j, were used. Thus, $h_j(\mathbf{x})$ was rewritten as $\Psi(\mathbf{x}, \Phi_j, \delta_j)$, where δ_j was the tree parameter [20]. The tree training phase was performed one split at a time, as in [36]. The split consisted of a validation function $\nu(\cdot) \in \mathbb{R}$, a threshold τ, and return parameters ζ_1 and ζ_2. Therefore, the estimation function can be defined as

$$p(\cdot) = \begin{cases} \zeta_1 & \text{if} \nu(\cdot) < \tau \\ \zeta_2 & \text{otherwise} \end{cases} \tag{20}$$

The optimal split was detected by minimizing at iteration j, given $\nu(\cdot)$

$$\sum_{i | \nu(\mathbf{x}_i) < \tau} w_i^j \left(\theta_i^j - \zeta_1\right)^2 + \sum_{i | \nu(\mathbf{x}_i) > \tau} w_i^j \left(\theta_i^j - \zeta_2\right)^2 \tag{21}$$

where τ, ζ_1, and ζ_2 were detected using an exhaustive search [36]. Here, the validation function $\nu(\mathbf{x}_i) = \phi^T \mathbf{x}_i$ that performs based on the results of $\mathbf{x}_i$ and a kernel ϕ_l, was used. The split training involved searching for a kernel ϕ, leaf parameters ζ_1 and ζ_2, and split threshold τ that minimized (21) based on $\nu(\mathbf{x}_i)$. First, a set of kernel candidates was built; then, for each candidate, the optimal threshold τ was detected using an exhaustive search [36]. For a specified kernel ϕ and threshold τ, the optimal values for ζ_1 and ζ_2 were then found as the weighted average of θ_i^j values of $\mathbf{x}_i$ samples on the corresponding side of the split [20].

To facilitate the operation, the kernel ϕ was limited to square windows within $\mathbf{x}$. This is more common than most previous methods. The dimensionality of the problem was

TABLE 3. The parameters were summarized for methodologies.

	Algorithm List					
Parameters	SVR	ESVR	LSTM	ELSTM	GBA	EGBA
Dimension of Feature	13	13	13	13	13	13
Dimension of Output	1	1	1	1	1	1
Number of Ensemble	N/A	5 to 20	N/A	5 to 20	N/A	5 to 20
Epsilon	3	3	$1.00\,e-08$	$1.00\,e-08$	N/A	N/A
KernelFunction	gaussian	gaussian	N/A	N/A	N/A	N/A
KernelScale	auto	auto	N/A	N/A	N/A	N/A
Number of Hidden Unit	N/A	N/A	50 to 200	50 to 200	N/A	N/A
FullyConnectdLayer	N/A	N/A	50	50	N/A	N/A
Dropout	N/A	N/A	0.5	0.5	N/A	N/A
MaxEpoch	N/A	N/A	50 to 100	50 to 100	N/A	N/A
Solver	N/A	N/A	adam	adam	N/A	N/A
GrandientThreshold	N/A	N/A	1	1	N/A	N/A
ShrinkageFactor	N/A	n/a	N/A	N/A	0.5	0.5
SubsamplingFactor	N/A	N/A	N/A	N/A	0.3	0.3
MaxTreeDepth	N/A	N/A	N/A	N/A	3 to 4	3 to 4
Number of Iterations	N/A	N/A	N/A	N/A	500 to 800	500 to 800

reduced, and it allowed the divisions to focus on regional features [20]. An operator $W_{c,a}(\mathbf{x}_i)$ that returned the pixel values of $\mathbf{x}$ in vector form within a square window centered at c on side length a, was introduced. The criteria in (19) were as follows:

$$\sum_{i=1}^{N} w_i^j \left(\phi^T W_{c,a}(\mathbf{x}_i) - \theta_i^j \right)^2 \tag{22}$$

where ϕ was limited to a square window parametrized by c and a. Considering c and a, the optimal ϕ can be calculated in closed form by solving the least squares problem in (22). Two improvements, regularization and splitting the training set, were introduced to prevent over-fitting. In the criteria given in (22), the term regularization was used to prefer a smooth kernel as

$$\sum_{i} w_i^j \left(\phi^T W_{c,a}(\mathbf{x}_i) - \theta_i^j \right)^2 + \rho \sum_{(m,n)\in\mathcal{N}} \left(\phi^m - \phi^n \right)^2 \tag{23}$$

where, $(m, n) \in \mathcal{N}$ denotes an index pair corresponding to the neighboring pixels and ϕ^m denotes the mth pixel of kernel ϕ. The second term of (23) imposes a smooth kernel that is controlled by $\rho \geq$ zero. Note that (23) can be minimized in a closed form using least squares. The recursive splitting step of Algorithm 2 then produces regression trees that are utilized as the base learner in GBA [20].

IV. EXPERIMENTAL RESULTS

A. STATISTICAL TEST

The first step was to extract the breathing signal from the raw PPG signal. This step removed very low frequencies from the PPG signal. A wide range of existing wearable sensors and smart devices use PPG signals. Three types of modulations, as the baseline wander (BW), the amplitude modulation (AM), and the frequency modulation (FM), were observed [30]. Only AM type data was used because of its higher performance and simple structure [30]. Consequently, RR was estimated from the AM signals based on the PPG data set. 192 AM signals were obtained; however, 64 AM signals were excluded because these data had unmodulated characteristics even after AM processing. Hence, AM signals were used to develop the proposed algorithm. These signals were separated sequentially into training data and test data; 90 signals were used as training data, and 38 signals were used as validation data. Based on a custom breath detection algorithm, reference RRs were computed from the oral-nasal pressure signals [30].

TABLE 4. The complexity was compared in terms of running time (training and testing at the number of ensemble (20)) [37], where the specifications of system are Intel® Core(TM) i7-8700 CPU 3.20 GHz, RAM 32.0 GB, OS 64 bit, and Matlab® 2019 (The MathWorks Inc., Natick, Ma, USA).

Method	LSTM [19]	GBA	ELSTM (5)	EGBA (5)
Total running time (s)	92.99	74.05	413.80	361.87

The configuration and parameter settings of the proposed and the conventional methods are summarized in Table 3, where the parameters were tuned to the best performance for each algorithm. The total running time was calculated using MATLAB ®2019 [37]. The EGBA presented a lower total running time than the ensemble-based LSTM (ELSTM) algorithm, as shown in Table 4. The results to differentiate LSTM and ELSTM techniques, are shown in the left columns of Table 5; the results to differentiate LSTM with NCA and ELSTM with NCA are summarized in the right columns of Table 5; the results to differentiate SVR and ESVR are represented in the left columns of Table 6; the results to differentiate SVR with NCA and ESVR with NCA are shown in the right columns of Table 6; the results to differentiate GBA and EGBA are shown in the left columns of Table 7, and the results to differentiate GBA with NCA and EGBA with NCA are shown in the right columns of Table 7. Based on the results of RR estimation, all algorithms were evaluated to verify the mean absolute error (MAE), $n^{-1}\sum_{i=1}^{n}|e_i|$) and the standard deviation (SD) of the MAE, as shown in Tables 5–7. These results represent the average results of 10 experiments for Ensemble (5)–Ensemble (20), respectively, where 10 denoted the number of experiments. In Tables 5-7, the MAE and SD results represent the average values of the ten tests. The lower costs of SD and MAE showed higher performance results. In addition, the Bland-Altman plots were used to

TABLE 5. The RR estimation results obtained using LSTM [19] and ELSTM were calculated with the reference RR method to represent MAE and SD, where the number in parentheses denotes the number of ensemble.

	Algorithm List									
Errors	LSTM	ELSTM (05)	ELSTM (10)	ELSTM (15)	ELSTM (20)	LSTMNCA	ELSTMNCA (5)	ELSTMNCA (10)	ELSTMNCA (15)	ELSTMNCA (20)
MAE	5.29	4.96	4.95	4.84	4.82	9.01	6.46	5.94	5.97	6.65
SD	0.75	0.51	0.19	0.12	0.01	4.62	1.65	0.96	0.79	0.73

TABLE 6. The RR estimation results obtained using SVR [21] and ESVR were calculated with the reference RR method to represent MAE and SD, where the number in parentheses denotes the number of ensemble.

	Algorithm List									
Errors	SVR	ESVR (05)	ESVR (10)	ESVR (15)	ESVR (20)	SVRNCA	ESVRNCA (5)	ESVRNCA (10)	ESVRNCA (15)	ESVRNCA (20)
MAE	5.85	5.83	5.83	5.82	5.83	6.11	6.03	6.01	6.02	6.02
SD	0.03	0.01	0.02	0.01	0.01	0.16	0.04	0.03	0.03	0.03

TABLE 7. The RR estimation results obtained using GBA [34] and EGBA were calculated with the reference RR method to represent MAE and SD, where the number in parentheses denotes the number of ensemble.

	Algorithm List									
Errors	GBA	EGBA (05)	EGBA (10)	EGBA (15)	EGBA (20)	GBANCA	EGBANCA (5)	EGBANCA (10)	EGBANCA (15)	EGBANCA (20)
MAE	5.49	4.73	3.83	3.65	3.30	4.93	3.05	3.22	2.82	2.88
SD	1.99	1.54	0.18	0.78	0.12	0.13	1.01	0.30	0.09	0.08

compare the performance of proposed EGBA methodology to the reference RR (bpm), as shown in Figs. 7-8.

B. ANOVA TEST

The analysis of variance (ANOVA) test [38] was used to efficiently evaluate and compare the performance of the proposed EGBA with GBA, LSTM, ELSTM, SVR, and ESVR algorithms. ANOVA is a statistical approach generally used in all situations that require comparison between two or more population means. That is, the hypothesis of interest in ANOVA are as follow:

$$H_0 : \mu_1 = \mu_2 \ldots = \mu_j, \quad H_1 : \mu_1 \neq \mu_2 \ldots \neq \mu_j \quad (24)$$

The null hypothesis in ANOVA is that there is no difference in means. The alternative hypothesis is that the means are not all equal. Hence, multi-comparison was used to determine the results of the group average that were different from others. One-way ANOVA is a simple and illustrative example of a linear model, given $e_{ij} = \alpha_j + \epsilon_{ij}$. Here, it was assumed that e_{ij} was the experimental result (MAEs) of the GBA, EGBA, LSTM, ELSTM, SVR, and ESVR, where $i = 10$ was the number of measurements, and j denoted the number of groups.

V. DISCUSSION

This study presents a novel approach for estimating RR from PPG signals. Based on the statistical tests, ELSTM was compared to LSTM. MAEs and SDs were computed for five categories, as shown in the left columns of Table 5. The MAE results of ELSTMs exhibited slightly higher performance than LSTM, as shown in Table 5. This was assumed to be the result of an over-fitting due to the reduction of NCA-based input capabilities in complex ELSTM methods. In detail, in ELSTMNCA, the input feature dimensions of ELSTM were reduced from 279 to 13, which did not work well with a small number of input feature dimensions. As shown in the left part of Table 6, the MAE results of ESVRs did not exhibit higher performance than those of SVR. Hence, these results imply that, compared to the conventional SVR method, ESVRs do not affect the performance. Even ELSTMs with NCA (ELSTMNCA) were found to have lower performance than ELSTMs. Even the ESVRs with the NCA (ESVRNCA) were confirmed to have lower performance than the ESVRs. This was due to overfitting as the input feature dimensions decrease from 279 to 13, as shown in Table 6 and Fig. 4. The MAEs in the left columns of Table 6 were directly consistent with the boxplots in Figure 4 (a). Also, the MAEs in the right columns in Table 6, were directly connected with the boxplots in Figure 4 (b). In Fig. 4, plot (a) denoted MAEs and SDs relative to the reference RR method obtained with the SVR and ESVR, where the leftmost box denoted the result of SVR, and the rest were the results of ESVR. Plot (b) indicated MAEs and SDs relative to the reference RR method obtained with the SVRNCA and ESVRNCA, where the leftmost box denoted the result of SVRNCA, and the rest were the results of ESVRNCA.

The proposed EGBA was compared with conventional GBA. The MAE and SD were evaluated for five categories, as shown in Table 7, where the numbers in parentheses denote the number of ensembles. EGBA could be used as an effective method to increase the number of ensembles, such as GBA, EGBA (5), EGBA (10), EGBA (15), and EGBA (20). This was because the MAEs of GBA, EGBA (05), EGBA (10), EGBA (15), and EGBA (20) were 5.49, 4.73, 3.83, 3.65, and 3.30, respectively, as shown in Table 7. Here, GBA denotes a single estimate. These results demonstrate that it is vital to use the ensemble methodology for robust estimation of RR. The MAEs in the left columns in Table 7, were directly matched with the box plots in Figure 5 (a). The MAEs in the right columns in Table 7, were directly connected with the

TABLE 8. The result of ANOVA test was obtained from the GBA and EGBA.

Source	SS	df	MS	F	Prob.>F(p-value)					
Group	31.815	4	7.954	11.43	$1.742 e-06$	31.476	4	7.869	28.49	$7.999 e-12$
Error	31.320	45	0.696			12.431	45	0.276		
Total	63.135	49				43.907	49			

TABLE 9. The result of ANOVA test was obtained from the LSTM and ELSTM.

Source	SS	df	MS	F	Prob.> F(p-value)					
Group	1.447	4	0.362	1.49	0.2208	64.579	4	16.145	3.09	0.0249
Error	10.917	45	0.243			235.141	45	5.225		
Total	12.364	49				299.725	49			

TABLE 10. The result of ANOVA test was obtained from the SVR and ESVR.

Source	SS	df	MS	F	Prob.> F(p-value)					
Group	0.0033	4	0.0008	2.62	0.0459	0.0641	4	0.016	2.32	0.0716
Error	0.0142	45	0.0003			0.3113	45	0.0069		
Total	0.0175	49				0.3754	49			

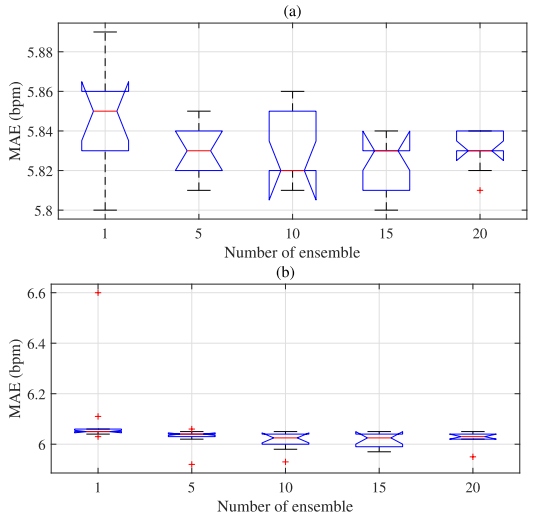

FIGURE 4. Plots (a) and (b) denotes MAEs and SDs relative to the reference RR method obtained with the SVR, ESVR, SVRNCA, and ESVRNCA.

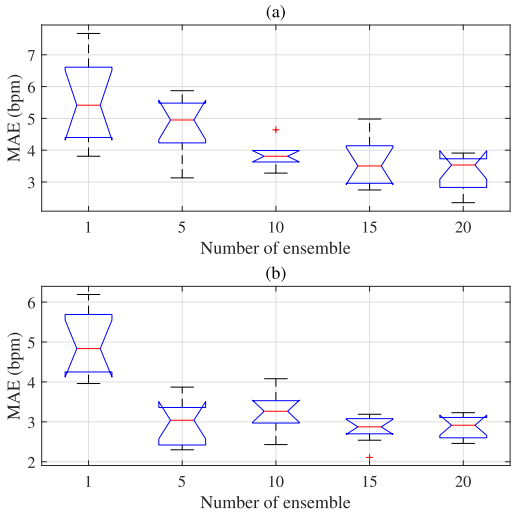

FIGURE 5. Plots (a) denotes MAEs and SDs relative to the reference RR method obtained with the GBA, EGBA, GBANCA and EGBANCA.

box plots in Figure 5 (b). In Fig. 5, plot (a) denoted MAEs and SDs relative to the reference RR method obtained with the GBA and EGBA, where the leftmost box denoted the result of GBA, and the rest were the results of EGBA. Plot (b) indicated MAEs and SDs relative to the reference RR method obtained with the GBANCA and EGBANCA, where the leftmost box denoted the result of GBANCA, and the rest were the results of EGBANCA. In addition, the results of EGBAs for the SD were provided, and it was confirmed that there was a little uncertainty as the SDs obtained from EGBAs were smaller than the SDs obtained from GBA. It is evident that MAEs and SDs generally decreased as the number of ensembles increased from EGBA (05) to EGBA (20). Therefore, EGBA estimates were found to represent higher accuracy as the number of ensembles increased, as shown in Table 7. However, the accuracies of the EGBA estimates were almost similar in the number of ensembles (25) and above. Furthermore, as indicated in Table 7, the SD of EGBA served 1.54 in Ensemble (5); however, decreased continuously to 0.12 in Ensemble (20), as the number of ensembles increased. Also, the variability SD of EGBANCA was found to be 1.01 in Ensemble (5); however, it decreased continuously to 0.08 in Ensemble (20), as shown in the last column of Table 7. This means that the proposed EGBA provides more accurate RR estimates than GBA, LSTM, ELSTM, SVR, and ESVR methods, as shown in Tables 5–7.

Based on the results of the ANOVA test, the performance of EGBA was analyzed. As shown in Fig. 5 (a), as the number of ensembles increased, the errors in EGBA decreased. Table 8 shows the between-groups variation (Group) and within-groups variation (Errors), where SS denotes the sum of squares, and df is the degrees of freedom. The total degree of freedom is the total number of measurements (MAEs) minus one, which denotes 49 (= 50-1). The between-groups degrees of freedom are the number of groups minus one, which denotes 4 (= 5-1). MS denotes the mean squared error, which denotes SS/df. The F-statistic denotes the ratio of the mean squared errors (7.954/0.696). The p-value, $1.742 e-06$, denotes the probability that the test statistic can obtain a value greater than the value of the calculated test statistic, i.e., P(F > 11.43). The small p-value,

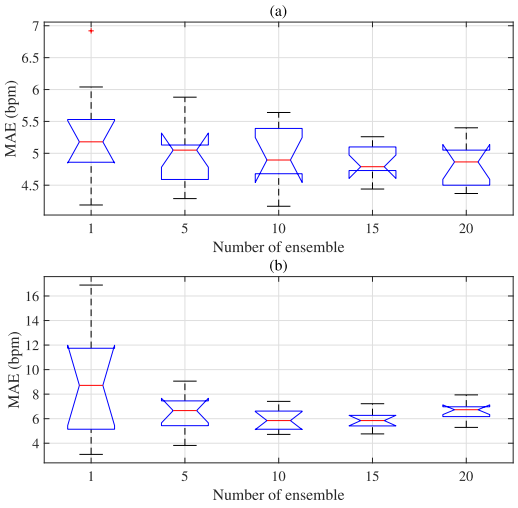

FIGURE 6. Plot (a) denotes MAEs and SDs relative to the reference RR method obtained with the LSTM and ELSTM, LSTMNCA and ELSTMNCA.

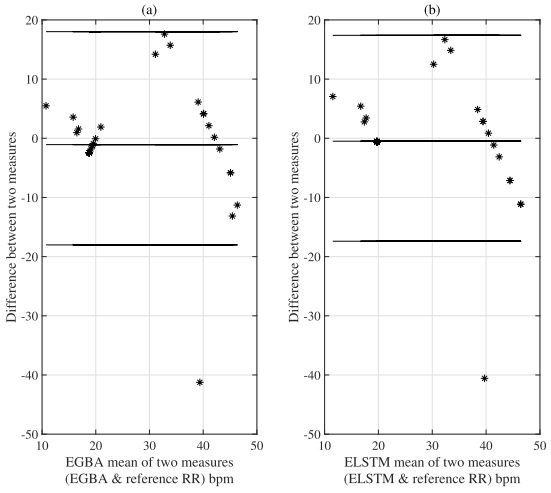

FIGURE 7. The Bland-Altman denotes comparing the performance between the EGBA (a) and ELSTM (b).

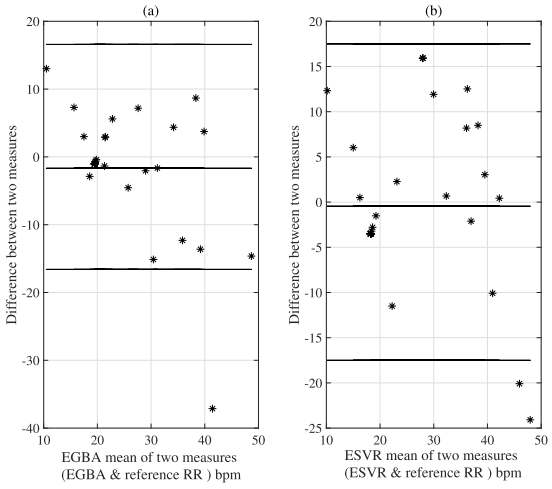

FIGURE 8. The Bland-Altman denotes comparing the performance between the EGBA (a) and ESVR (b).

1.742 $e-06 < (\alpha = 0.05)$, shows that differences between group means are significant. Namely, the MAE of the first column GBA, MAE of the second column EGBA (5), MAE of the third column, MAE of the fourth column, and MAE of the last column EGBA (20) are different, as shown in Tables 7 and 8. As shown in the right part of Table 7, MAEs of EGBANCAs [29] was 3.05, 3.22, 2.82, and 2.88 for Ensembles (5), (10), (15), and (20), respectively. These results are more accurate than the result of MAEs (4.93) of GBA with NCA (GBANCA). Here, the p-value, 7.999 $e-12$ was less than the significant value (0.05) in the last column of Table 8. The MAE of GBANCA, MAE of the EGBANCA (5), MAE of the EGBANCA (10), MAE of the EGBANCA (15), and MAE of the EGBANCA (20) are also different, as denoted in Tables 7 and 8. Furthermore, the performance of the proposed EGBA was higher compared to the ensemble-based LSTM (ELSTM) approach. MAEs of the EGBA (5), (10), (15), and (20) were 4.73, 3.83, 3.65, and 3.30, respectively, as shown in Table. 7. It was confirmed that the MAEs for the proposed EGBAs were lower than those for LSTM, ELSTM, SVR, and ESVR methods, as shown in Tables 5–7. As shown in Fig. 5 (a) and (b), the EGBAs and EGBANCAs resulted in lower MAEs for Ensembles (5), (10), (15), and (20), compared to the ELSTMs and ELSTMNCAs algorithms in Fig. 6 (a) and 6 (b). In addition, the performances of EGBANCA by MAE between the estimated results in Fig. 5 (b) and reference measurements were higher to those obtained with ELSTMNCA as shown in Fig. 6 (b). These results indicate a higher performance than benchmark algorithms. Therefore, it can be concluded that EGBA decreases uncertainty, such as SD and MAE, and increases performance reliability.

Moreover, the performance between the EGBA and ELSTM models using the Bland-Altman plots was expressed, as shown in Fig.7 (a) and (b). The limits of agreement (see horizontal lines in Fig. 7) that we used were MAE $\pm$ 2 $\times$ SD for plot; most black asterisks lie within the limits of agreement. The bias (see horizontal central lines) for the plots were small ($\leq \pm$ 1.0 bpm). The result of Fig.7 (a) was slightly closer to the centerline than the result of Fig. 7 (b). Figure 7 well shows an exemplary result corresponding to the third column in Table 5 (ELSTM) and in Table 7 (EGBA). The performance between the EGBA and ESVR models was also compared using the Bland-Altman plots as shown in Fig. 8 (a) and (b). Hence, from the results of overall evaluation, we show that performance of proposed EGBA model is as accurate as those obtained using the ELSTM and ESVR models.

VI. CONCLUSION

In conclusion, this study proposes a novel method using the EGBA for RR estimation using insufficient PPG signal data. Multi-phases based on feature extraction and selection methodology were used to choose the high weighted feature vectors. The primary contribution of this study is the use of

EGBA to achieve higher accuracy and stability. The proposed EGBA with a small sample acquired lower SDs and MAEs for RR estimation compared to the LSTM, SVR, ELSTM, and ESVR methods. The proposed EGBA scheme lowers the SDs of MAEs for RR estimation as opposed to single GBA estimators. The multi-phases based on feature extraction and selection methodology are expected to be suitable for the proposed EGBA. Therefore, this study provides a method to increase the accuracy in RR estimation and offers a solution that can reduce estimation errors such as MAEs and SDs. Based on the MAEs, the best result of proposed EGBA was 3.30 bpm, those of the ELSTM was 4.82 bpm, and those of the ESVR technique was 5.83 bpm, respectively. Additional experimental tests will be performed on a new subject population in the future. Moreover, since the proposed EGBA is time-consuming when compared with the single GBA, optimization can be further used to simplify the enhance the efficiency of the EGBA estimator.

ACKNOWLEDGMENT

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. 2020R1A2C1010405).

REFERENCES

[1] S. Ridley, "The recognition and early management of critical illness," *Ann. Roy. College Surgeons England*, vol. 87, no. 5, pp. 315–322, Sep. 2005.
[2] M. H. Ebell, "Predicting pneumonia in adults with respiratory illness," *Amer. Fam. Phys.*, vol. 76, no. 4, 560–562, 2007.
[3] X. Zhang and Q. Ding, "Respiratory rate estimation from the photoplethysmogram via joint sparse signal reconstruction and spectra fusion," *Biomed. Signal Process. Control*, vol. 35, pp. 1–7, May 2017.
[4] S. C. Gandevia and D. K. McKenzie, "Respiratory rate: The neglected vital sign," *Med. J. Aust.*, vol. 189, no. 9, pp. 531–532, Nov. 2008.
[5] S. Dash, K. H. Shelley, D. G. Silverman, and K. H. Chon, "Estimation of respiratory rate from ECG, photoplethysmogram, and piezoelectric pulse transducer signals: A comparative study of time–frequency methods," *IEEE Trans. Biomed. Eng.*, vol. 57, no. 5, pp. 1099–1107, May 2010.
[6] D. Clifton, J. G. Douglas, P. S. Addison, and J. N. Watson, "Measurement of respiratory rate from the photoplethysmogram in chest clinic patients," *J. Clin. Monitor. Comput.*, vol. 21, no. 1, pp. 55–61, Jan. 2007.
[7] S. G. Fleming and L. Tarassenko, "A comparison of signal processing techniques for the extraction of breathing rate from the photoplethysmogram," *Int. J. Biol. Med. Sci.*, vol. 2, no. 4, pp. 232–236, 2007.
[8] P. Leonard, "Standard pulse oximeters can be used to monitor respiratory rate," *Emergency Med. J.*, vol. 20, no. 6, pp. 524–525, Nov. 2003.
[9] L. Nilsson, A. Johansson, and S. Kalman, "Respiration can be monitored by photoplethysmography with high sensitivity and specificity regardless of anaesthesia and ventilatory mode," *Acta Anaesthesiol. Scandinavica*, vol. 49, no. 8, pp. 1157–1162, 2005.
[10] K. H. Shelley, A. A. Awad, R. G. Stout, and D. G. Silverman, "The use of joint time frequency analysis to quantify the effect of ventilation on the pulse oximeter waveform," *J. Clin. Monitor. Comput.*, vol. 20, no. 2, pp. 81–87, Jun. 2006.
[11] L. Mirmohamadsadeghi and J.-M. Vesin, "Respiratory rate estimation from the ECG using an instantaneous frequency tracking algorithm," *Biomed. Signal Process. Control*, vol. 14, pp. 66–72, Nov. 2014.
[12] M. A. F. Pimentel, A. E. W. Johnson, P. H. Charlton, D. Birrenkott, P. J. Watkinson, L. Tarassenko, D. A. Clifton, and "Toward a robust estimation of respiratory rate from pulse oximeters," *IEEE Trans. Biomed. Eng.*, vol. 64, no. 8, pp. 1914–1923, Aug. 2017.
[13] P. S. Addison, J. N. Watson, M. L. Mestek, and R. S. Mecca, "Developing an algorithm for pulse oximetry derived respiratory rate (RRoxi): A healthy volunteer study," *J. Clin. Monitor. Comput.*, vol. 26, no. 1, pp. 45–51, Feb. 2012.
[14] J. Lee and K. H. Chon, "Respiratory rate extraction via an autoregressive model using the optimal parameter search criterion," *Ann. Biomed. Eng.*, vol. 38, no. 10, pp. 3218–3225, Oct. 2010.
[15] X. Zhang and Q. Ding, "Respiratory rate monitoring from the photoplethysmogram via sparse signal reconstruction," *Physiol. Meas.*, vol. 37, no. 7, pp. 1105–1119, Jun. 2016.
[16] S. Khreis, D. Ge, H. A. Rahman, and G. Carrault, "Breathing rate estimation using Kalman smoother with electrocardiogram and photoplethysmogram," *IEEE Trans. Biomed. Eng.*, vol. 67, no. 3, pp. 893–904, May 2020.
[17] A. P. Prathosh, P. Praveena, L. K. Mestha, and S. Bharadwaj, "Estimation of respiratory pattern from video using selective ensemble aggregation," *IEEE Trans. Signal Process.*, vol. 65, no. 11, pp. 2902–2916, Jun. 2017.
[18] A. Johansson, "Neural network for photoplethysmographic respiratory rate monitoring," *Med. Biol. Eng. Comput.*, vol. 41, no. 3, pp. 242–248, May 2003.
[19] S. Liu, J. Yao, and M. Motani, "Early prediction of vital signs using generative boosting via LSTM networks," in *Proc. IEEE Int. Conf. Bioinf. Biomed. (BIBM)*, Nov. 2019, pp. 437–444.
[20] C. Becker, R. Rigamonti, V. Lepetit, and P. Fua, "Supervised feature learning for curvilinear structure segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, Nagoya, Japan, Sep. 2013, pp. 1–8.
[21] A. A. Rakotomamonjy, "Analysis of SVM regression bound for variable ranking," *Neurocomputing*, vol. 70, pp. 1489–1491, Feb. 2007.
[22] A. R. Fekr, M. Janidarmian, K. Radecka, and Z. Zilic, "A medical cloud-based platform for respiration Rate measurement and hierarchical classification of breath disorders," *Sensors*, vol. 2014, no. 14, pp. 1204–1224, 2014.
[23] S. Orfanidis, *Optimum Signal Processing*, 2nd ed. New York, NY, USA: Macmillan, 1988, ch. 5.
[24] T. Li and M. Zhou, "ECG classification using wavelet packet entropy and random forests," *Entropy*, vol. 18, no. 8, pp. 1–16, 2016.
[25] R. F. Leonarduzzi, G. Schlotthauer, and M. E. Torres, "Wavelet leader based multifractal analysis of heart rate variability during myocardial ischaemia," in *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol.*, Buenos Aires, Argentina, Aug. 2010, pp. 110–113.
[26] Q. Zhao and L. Zhang, "ECG feature extraction and classification using wavelet transform and support vector machines," in *Proc. Int. Conf. Neural Netw. Brain*, Beijing, China, Oct. 2005, pp. 1089–1092.
[27] E. A. Maharaj and A. M. Alonso, "Discriminant analysis of multivariate time series: Application to diagnosis based on ECG signals," *Comput. Statist. Data Anal.*, vol. 70, pp. 67–87, Feb. 2014.
[28] W. Karlen, S. Raman, J. M. Ansermino, and G. A. Dumont, "Multiparameter respiratory rate estimation from the photoplethysmogram," *IEEE Trans. Biomed. Eng.*, vol. 60, no. 7, pp. 1946–1953, Jul. 2013.
[29] W. Yang, K. Wang, and W. Zuo, "Neighborhood component feature selection for high-dimensional data," *J. Comput.*, vol. 7, no. 1, pp. 161–168, Jan. 2012.
[30] P. H. Charlton, T. Bonnici, L. Tarassenko, D. A. Clifton, R. Beale, and P. J. Watkinson, "An assessment of algorithms to estimate respiratory rate from the electrocardiogram and photoplethysmogram," *Physiol. Meas.*, vol. 37, no. 4, pp. 610–626, Apr. 2016.
[31] A. L. Goldberger, A. LAN, L. Glass, J. M. Hausdorff, P. H. Ivanov, P. Mark, J. E. Mietus, G. Moody, C. K. Peng, and H. E. Stanley, "Physio bank, physioToolkit, and physioNet: Components of a new research resource for complex physiologic signals," *Circulation*. vol. 101, no. 23, 13 Jun. 2000, pp. 215–220. [Online]. Available: http://circ.ahajournals.org/content/101/23/e215.full
[32] M. Amiri, E. Zahedi, and F. Behnia, "Autoregressive modeling of the photoplethysmogram AC signal amplitude changes after flow-mediated dilation in healthy and diabetic subjects," in *Proc. 19th Iranian Conf. Biomed. Eng. (ICBME)*, Tehran, Iran, Dec. 2012, pp. 170–173.
[33] C. E. Shannon. *A Mathematical Theory of Communication*. Accessed: Mar. 1, 2020. [Online]. Available: http://people.math.harvard.edu/ctm/home/text/others/shannon/entropy/entropy.pdf
[34] J. H. Friedman, "Greedy function approximation: A gradient boosting machine," *Ann. Statist.*, vol. 29, no. 5, pp. 1189–1232, Oct. 2001.
[35] P. Gupta. *Decision Trees in Maching Learning*. Accessed: Mar. 1, 2020. [Online]. Available: https://towardsdatascience.com/decision-trees-in-machine-learning-641b9c4e8052
[36] T. Hastie, R. Tibshirani, and J. Friedman, "Introduction," in *The Elements of Statis- Tical Learning*. New York, NY, USA: Springer, 2009.
[37] M. Knapp-Cordes and B. McKeeman, "Improvements to tic and TOC functions for measuring absolute elapsed time performance in MATLAB," in *MATLAB Technical Articles and Newsletters*. Natick, MA, USA: The MathWorks, 2011.
[38] R. A. Bailey, *Design of Comparative Experiments*. Cambridge, U.K.: Cambridge Univ. Press, 2008.