Research Article

An Efficient Fingertip Photoplethysmographic Signal Artifact Detection Method: A Machine Learning Approach

Tasbiraha Athaya and Sunwoong Choi

School of Electrical Engineering, Kookmin University, Seoul 02707, Republic of Korea

Correspondence should be addressed to Sunwoong Choi; schoi@kookmin.ac.kr

Received 11 March 2021; Revised 14 September 2021; Accepted 16 September 2021; Published 4 October 2021

Academic Editor: Antonio Martinez-Olmos

Copyright © 2021 Tasbiraha Athaya and Sunwoong Choi. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

A photoplethysmography method has recently been widely used to noninvasively measure blood volume changes during a cardiac cycle. Photoplethysmogram (PPG) signals are sensitive to artifacts that negatively impact the accuracy of many important measurements. In this paper, we propose an efficient system for detecting PPG signal artifacts acquired from a fingertip in the public healthcare database named Multiparameter Intelligent Monitoring in Intensive Care (MIMIC) by using 11 features as the input of the random forest algorithm and classified the signals into two classes: acceptable and anomalous. A real-time algorithm is proposed to identify artifacts by using the method. The efficient Fisher score feature selection algorithm was used to order and select 11 relevant features from 19 available features that represented the PPG signal very effectively. Six machine learning algorithms (random forest, decision tree, Gaussian naïve Bayes, linear support vector machine, artificial neural network, and probabilistic neural network) were applied with the extracted features, and their classification accuracy was measured. Among them, the random forest had the best performance using only 11 out of 19 features with an accuracy of 85.68%. Our proposed method also achieved good sensitivity and specificity value of 86.57% and 85.09%, respectively. The proposed real-time algorithm can be an easy and convenient way for real-time PPG signal artifact detection using smartphones and wearable devices.

1. Introduction

Photoplethysmogram (PPG) provides a myriad of information related to the cardiovascular system. Technological advancement has drawn a notable effect on the global healthcare system. A reliable process for measuring different types of physiological signals for daily use is becoming increasingly necessary to minimize hospitalization costs and to save time. Moreover, the use of wearable devices is increasing daily and these devices incorporate a variety of fitness and health strategies into everyday life. Many important physiological parameters are derived from easily obtained biosignals like PPG [1]. Thus, PPG signals are researched widely around the world. In recent studies, PPG signals have been used to measure various clinical parameters, such as blood oxygen saturation, respiratory rate, pulse rate, blood glucose, blood pressure, and many more [2, 3].

A major obstacle to PPG signal acquisition is different artifacts inherent in the signal. For this reason, the use of PPG as the input of various algorithms may lead to incorrect results and lower the reliability of the obtained results. Although various signal processing techniques have been developed to solve the problems of noise in PPG waveforms, it is necessary to remove or at least detect artifacts in available PPG data [4]. There is a high demand for extracting artifact-free PPG signals for the proper assessment of clinical parameters.

Many online free-access databases, such as the Multiparameter Intelligent Monitoring in Intensive Care (MIMIC) [5] dataset and the University of Queensland Vital Signs dataset [6], are available with PPG data of ICU patients. However, several obtained PPG signals are anomalous and are thus unsuitable for deriving other parameters. In many types of research, PPG signals from these databases are used to measure PPG-derived parameters [7].

Several studies have been conducted to detect artifacts in PPG signals obtained from smartphones [8, 9]. In [8], invalid PPG signals were detected due to hand movement. They used six features and fed the extracted features into a probabilistic neural network for motion and noise artifact detection. The authors used two algorithms, personalized and generalized, to detect noise in the PPG signal of individuals and in general, respectively. In [9], variable frequency complex demodulation spectrogram analysis was used to detect noise artifacts in a smartphone signal.

In [10], the support vector machine method was applied to detect motion artifacts. The method performed discrete wavelet transform to denoise and separate the AC and DC components of the PPG signal before feature extraction. Then, four features of the PPG signal obtained from a wearable PPG sensor were calculated. In [11], a linear classifier with seven features of the PPG signal was used to obtain the acceptable segments to measure blood glucose and blood pressure.

A 1D convolution neural network with PPG as input was used in [12] to detect only PPG motion artifact. They used 3 different datasets (customized and 2 independent) for this purpose. They identified the motion artifact-prone signals and good signals in the dataset to train and validate their network.

Except [12], all the mentioned studies used customized datasets with healthy subjects and defined motions for artifacts. In contrast, in our study, the dataset consisted of intensive care unit (ICU) patients, resulting in abnormal PPG shape. Though [12] used the MIMIC-II ICU patient dataset with their custom dataset, they used computationally complex 1D-CNN which is not computationally efficient for portable and wearable devices. The authors in [13–16] proposed for hierarchical decision rules for signal quality assessment to identify and remove signals with low quality. The performance of their method is also very good. But the rule-based methods might have lacking of adaptability for the waveforms under other criteria. So, in those studies, the general scenarios of signal quality assessment might have not been investigated for all signals with artifacts, whereas the machine learning algorithms are different from the hierarchical rule-based methods for signal quality check. Machine learning algorithms are able to generate a multivariate threshold for signal quality checking based on given data and relevant features [17]. So their hierarchical decision rules were not adopted; rather, a simple machine learning approach was proposed.

Our main goal was to detect all kinds of artifacts. We focused on the distorted shaped PPG signals (in both time and frequency domains) caused by involuntary movement, sensor anomalies, sensor displacement, any type of involuntary action, bodily parameters, and severe heart diseases.

Our contributions stated in this manuscript include:

- (i) Detection of artifacts and classification of commonly used PPG signals of the MIMIC database into two groups: acceptable and anomalous for diagnosis

FIGURE 1: An ideal single photoplethysmogram (PPG) signal.

- (ii) Detection of artifacts on both time and frequency domains of the signal avoiding the beat detection process
- (iii) Analyze the extended feature space and construct a reduced efficient feature set containing 11 effective features
- (iv) Develop a standalone real-time artifact detection process using only a smartphone application to detect PPG signal artifacts in real time

In comparison to our preliminary version of the work [18], this extended version includes a significant analysis of the obtained new results. We have extended our work by including a new and efficient feature set and by providing a feature selection technique to select 11 out of 19 features. A standalone real-time implementation with our method to identify PPG artifacts has been proposed, and the complexities related to real-time implementation have been discussed. Moreover, a comparison of our work with the existing works is also shown with our chosen dataset. The comparison is shown on the basis of not only accuracy but also sensitivity, specificity, computational time, and statistical significance test.

The remainder of this paper is organized as follows. Section 2 provides a brief overview of the background of PPG signals and artifacts. Section 3 explains the proposed approach by describing data acquisition, signal preprocessing, feature extraction, feature selection, and selection of a suitable artifact detection machine learning approach. Section 4 provides the results of the machine learning model. The importance of our work and future perspectives have been discussed in Section 5. Finally, Section 6 concludes the paper.

2. Background: Photoplethysmogram and Artifacts

A PPG is a waveform used to noninvasively calculate the variation in blood volume in a cardiac cycle [19]. PPG measurement devices are small, portable, and easy to use. The devices are mainly used in hospitals and medical centers to

FIGURE 2: Acceptable photoplethysmogram (PPG) signals.

monitor patients. The signal has a wave-like motion and represents variation of blood volume in the vessel or capillaries at the measurement site [20]. PPG utilizes an optical approach by using variation in light absorption by tissues to determine the difference in oxygenation levels [21].

A standard cycle of a PPG signal is presented in Figure 1. Usually, a single PPG signal of a healthy individual has a systolic and diastolic peak with a dicotic notch [20]. The period from the start of a signal to the systolic peak is called the systolic phase, while the period from the systolic peak to the end of the single signal is called the diastolic phase.

In this study, we classify PPG signals into two categories: acceptable and anomalous. Acceptable PPG signals are illustrated in Figure 2, where Figures 2(a) and 2(b) display almost clean PPG signals. In Figures 2(c) and 2(d), a slight dicotic notch is present, and the notch becomes horizontal, which can be considered acceptable. In Figure 2(e), no dicotic notch develops; however, a significant change in

the angle of the notch is clearly observed. Thus, this signal is considered acceptable.

PPG signal artifacts are mostly caused by different kinds of motions caused by the subjects in measurement sites. But artifacts not only are limited to motion, different kinds of environmental noises, equipment errors, and finger pressure but also can be the cause of PPG signal artifacts. Moreover, for variation in dermis thickness, fat layer, and nail of subjects, sometimes proper signals are not obtained, which results in artifacts. People suffering from cardiac arrhythmia give irregular signal patterns or long absence in the heartbeat [22]. So, irregular or long absence in heartbeat is also considered artifacts. Segments with different artifacts of PPG signals of the MIMIC dataset are considered anomalous signals, and some examples are shown in Figure 3. Figures 3(a) and 3(b) display uneven systolic and diastolic peaks, which indicate a motion artifact. Figure 3(c) indicates high amount of noise levels in the diastolic phase.

FIGURE 3: Anomalous photoplethysmogram (PPG) signals.

Figure 3(d) exhibits a long absence in the signal. Figure 3(e) contains multiple artifacts, and the signal is considered anomalous.

3. Materials and Methods

A block diagram of our overall workflow is presented in Figure 4. PPG signals are acquired from the publicly available MIMIC database. Then, the signals are preprocessed for feature extraction and for ordering based on importance. The ordered features are fed to different machine learning models. For an accurate evaluation of the models, the accuracy of the models is obtained by performing 10-fold cross-validation. Finally, the best model with a minimum number of features is obtained based on the performance matrices.

3.1. Data Collection. The MIMIC database [23] contains a collection of recordings from 121 ICU patients. In each case, the data contain periodic measurements and signals collected from a patient's bedside monitor or medical record

[24]. Different waveforms for 72 subjects are publicly available for practical experiments. Of these signals, a subset of PPG signals from 55 subjects with a 125 Hz sampling rate was used in this study. The signals were downloaded using the native Python Waveform Database (WFDB) package, which contains a library for writing, reading, and processing WFDB signals and annotations [25].

3.2. Data Preprocessing. The steps of data preprocessing are presented in Figure 4. As a PPG signal is a low-frequency signal, each signal file was filtered with cut-off frequencies of 0.5 Hz to 8 Hz using the equiripple FIR filter [26]. Values below 0.5 Hz were identified as baseline wandering, while values above 8 Hz were considered high-frequency noise. The FIR filter was selected as it is always stable with a linear phase. As we required the exact shape of the PPG signal, maintaining a constant phase was necessary. The filtering coefficients were calculated using the Python Filter Design Analysis Tool. In the magnitude (dB) and phase response curve of the filter, the lower corner of the band-pass filter stop frequency was set to

FIGURE 4: Block diagram of the proposed workflow.

0.1 Hz, the pass frequency was set to 0.5 Hz, and the stop attenuation was set to -60 dB. For the higher corner, the pass frequency was set to 8 Hz, the stop frequency was set to 9 Hz, and the stop attenuation was set to -80 dB. The PPG signals were filtered using 525 filtering coefficients.

The equiripple filter is designed based on the weighted error term of

$$E_{\{b[n]\}}(\omega) = W(\omega) \left[H_{dr}(\omega) - H_{r,\{b[n]\}}(\omega) \right], \quad (1)$$

where $E_{\{b[n]\}}$, W , H_{dr} , and $H_{r,\{b[n]\}}$ indicate the weighted error value, weights, desired frequency response, and designed frequency response, respectively. The weight values can emphasize a certain frequency band. Then, the filter is designed to optimize the expression in

$$\min_{\text{over } \{b[n]\}} \left[\max_{\omega \in \text{PBUSB}} |E_{\{b[n]\}}(\omega)| \right]. \quad (2)$$

The value of the above expression is optimized to have as lower value as possible. The PB and SB indicate the passbands and stopbands, respectively.

For ease of calculation, the filtered signals were divided into windows of 350 samples with a 28.5% overlap. Usually for a normal person, the heartbeat is between 60 and 100 beats per minute. That means 1 heartbeat corresponds to 1 s for the minimum heartbeat. The sampling rate of the MIMIC database is 125 Hz. So, for 2 heartbeats, there are 250 samples, if the heartbeat is minimum (i.e., 60 bpm). 350 samples were taken

TABLE 1: Categorization of PPG windows by two experts.

		Expert 2		Total
		Acceptable	Anomalous	
Expert 1	Acceptable	954	450	1404
	Anomalous	52	1044	1096
Total		1006	1494	2500

because we considered 100 samples for overlapping so that no information of signal would miss while doing segmentation. Furthermore, two or more heartbeats can be found in a window of 350 samples which results in accurate feature calculation. The divided PPG windows are referred to as segments.

Next, 2,500 windows were selected randomly. The labeling of acceptable and anomalous windows was performed manually and carefully by two experts independently. The categorization of the selected 2,500 segments by the experts is presented in Table 1 as acceptable and anomalous segments, and 1,998 common PPG segments were used for the final dataset. Among common PPG segments, 954 segments were found to be acceptable, while 1044 segments were found to be anomalous.

3.3. Feature Extraction. To detect artifacts in the PPG segments based on quality, 19 features or parameters were extracted from each segment. The features were selected after reviewing related studies. The extracted features are listed in Table 2 with their notation and references. The features are described below.

TABLE 2: List of features.

No.	Feature description	Notation	Reference work
F1	Signal mean	$S\mu$	[8]
F2	Signal variance	$S\sigma^2$	[27]
F3	Signal skewness	S_{skew}	[28, 29]
F4	Signal interquartile range	S_{IQR}	[30]
F5	Kurtosis	S_K	[10, 31]
F6	Sample number less than or equal to 10% of the maximum AVPG	N_{AVPG}	New
F7	Standard deviation of consecutive peak amplitude	SD_{amp}	[8, 10]
F8	Standard deviation of consecutive peak interval	SD_{CPI}	[8, 10, 20]
F9	Mean absolute deviation of consecutive peak amplitude	MAD_{amp}	[10]
F10	Slope ratio	SR	[8]
F11	Mean of moving standard deviation	$MSD\mu$	[8]
F12	Optimal AR model order	AR_O	[8]
F13	Spectral entropy	H_S	[11, 29, 32]
F14	Qi-Zheng energy	QZE	[32]
F15	Zero crossing rate	ZCR	[11, 29, 33]
F16	KTE mean	$KTE\mu$	[11]
F17	KTE variance	$KTE\sigma^2$	[11]
F18	KTE interquartile range	KTE_{IQR}	[11]
F19	KTE skewness	KTE_{skew}	[11]

3.3.1. *Signal Mean ($S\mu$)*. The mean, denoted as $S\mu$ and defined below, is the average value of a signal. As it is the most common parameter of a signal, we observed that it plays an important role in artifact detection [8].

$$S\mu = \frac{1}{N} \sum_{i=1}^N x_i, \quad (3)$$

where x_i is the i th sample of the filtered PPG segment x . Here, $i = 1, 2, \dots, N$, where N is the number of samples in a PPG segment. In this study, the value of N was 350.

3.3.2. *Signal Variance ($S\sigma^2$)*. The signal variance, which is defined below, expresses the variation in the instantaneous values of the standardized squared sum and the mean. It represents the deviation of a signal from the signal mean and is an effective indicator of signal classification [27].

$$S\sigma^2 = \frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^2, \quad (4)$$

where $\bar{x}$ denotes the mean of a PPG segment and is the same as $S\mu$.

3.3.3. *Signal Skewness (S_{skew})*. In [28, 29], the authors reported that signal skewness is necessary to identify anomalous PPG signals. Skewness expresses the uniformity or

absence of it for a probability distribution and is defined as follows:

$$S_{skew} = \frac{(1/N) \sum_{i=1}^N (x_i - \bar{x})^3}{\sigma^3}, \quad (5)$$

where σ is the standard deviation of the PPG segment.

3.3.4. *Signal Interquartile Range (S_{IQR})*. The interquartile range (IQR) can be used to identify outliers in data. It is also known as H -spread, which represents the measurement of statistical scattering. The IQR is the measurement of variations based on dividing a dataset into quartiles [30] and can be expressed as follows:

$$S_{IQR} = Q_3 - Q_1, \quad (6)$$

where Q_1 and Q_3 denote the first and third quartiles of a PPG segment, respectively.

3.3.5. *Kurtosis (S_K)*. Kurtosis expresses whether a distribution is flat or peaked relative to the normal distribution. It has been demonstrated in [10, 31] as an effective measure of signal quality and is expressed as follows:

$$S_K = \frac{(1/N) \sum_{i=1}^N (x_i - \bar{x})^4}{\sigma^4}. \quad (7)$$

FIGURE 5: (a) Velocity photoplethysmogram (VPG). (b) Absolute velocity photoplethysmogram (AVPG) of an acceptable photoplethysmogram segment (Figure 2(a)).

FIGURE 6: (a) Velocity photoplethysmogram (VPG). (b) Absolute velocity photoplethysmogram (AVPG) of an anomalous photoplethysmogram segment (Figure 3(d)).

3.3.6. Sample Number Less Than or Equal to 10% of the Maximum Absolute Velocity Photoplethysmogram (N_{AVPG}). The velocity photoplethysmogram (VPG) is the first derivative of a PPG signal that contains a myriad of important information [34]. If the number of sample values in the absolute VPG (AVPG) is less than or equal to 10% of the maximum value, then the window may be noisy. A value of 10% was selected after observing the segments. The AVPGs are illustrated in Figures 5 and 6. Figures 5(a) and 5(b) display the VPG and AVPG, respectively, of the acceptable PPG segment in Figure 2(a), while Figures 6(a) and 6(b) display the VPG and AVPG, respectively, of the anomalous PPG segment in Figure 3(d). The number of samples less than or equal to 10% of the maximum magnitude of the AVPG is higher in Figure 6(b) than in Figure 5(b). A larger number indicates anomalous segments; such as Figure 3(d).

3.3.7. Standard Deviation of Consecutive Peak Amplitude (SD_{amp}). Because the standard deviation represents the range of variation from the mean value, in [8, 10], the standard deviation of the consecutive peak amplitude was selected as a feature and is expressed as follows:

$$SD_{amp} = \sqrt{\frac{1}{N_{peak}} \sum_{i=1}^{N_{peak}} (A_{x_i} - \bar{A})^2}, \quad (8)$$

where A_{x_i} is the consecutive peak amplitude of the i^{th} single signal in a PPG segment and N_{peak} is the number of peaks in a PPG segment. The mean of the consecutive peak amplitude of the segment is denoted as $\bar{A}$.

3.3.8. Standard Deviation of Consecutive Peak Interval (SD_{CPI}). According to [8, 10], the standard deviation of the consecutive peak interval (i.e., the timing of consecutive peaks) can be used as a feature:

$$SD_{CPI} = \sqrt{\frac{1}{N_{peak} - 1} \sum_{i=1}^{N_{peak}-1} (T_{x_i} - \bar{T})^2}, \quad (9)$$

where T_{x_i} is the consecutive peak time of the i^{th} single pulse in a PPG segment. The mean of the consecutive peak interval of the segment is denoted as $\bar{T}$.

3.3.9. Mean Absolute Deviation of Consecutive Peak Amplitude (MAD_{amp}). The mean absolute deviation (MAD) is the average distance between each data point and the mean. In [1, 10], the MAD of the consecutive peak amplitude was used to classify the PPG signal. The MAD is expressed in

$$MAD_{amp} = \sum_{i=1}^{N_{peak}} \frac{|A_{x_i} - \bar{A}|}{N_{peak}}. \quad (10)$$

3.3.10. Slope Ratio (SR). In [8], the slope ratio of each segment was measured by

$$SR = \left| \frac{SP}{SN} \right|, \quad (11)$$

where SP denotes the highest value of positive slopes and SN denotes the lowest value of negative slopes in a segment. These can be expressed as follows:

$$SP = \max ((x_i - x_{i-1}) > 0, i = 2, 3, 4, \dots, N), \quad (12)$$

$$SN = \min ((x_i - x_{i-1}) < 0, i = 2, 3, 4, \dots, N). \quad (13)$$

Note that $x_i - x_{i-1}$ becomes the slope between i^{th} and $(i - 1)^{\text{th}}$ sample point of that PPG segment. It is known that the slope ratio of anomalous segments is higher than that of acceptable segments [8].

3.3.11. Mean of Moving Standard Deviation ($MSD\mu$). $MSD\mu$ in each segment is calculated using the following equation used in [8]:

$$MSD\mu = \frac{\sum_{m=W}^N MSD_{m,W}}{N - W + 1}, \quad (14)$$

where $MSD_{m,W}$ is the moving standard deviation of the m th sample point in a sliding window of length W of a PPG segment. $W = 3$ samples.

$$MSD_{m,W} = \sqrt{\frac{1}{W-1} \sum_{i=m-W+1}^m (x_i - \bar{x}_W)^2}, \quad (15)$$

where $\bar{x}_W$ is the mean of the samples in the sliding window. In anomalous PPG segments, the samples are scattered; thus, $MSD\mu$ is higher for anomalous segments.

3.3.12. Optimal Autoregressive Model Order (AR_O). The optimal autoregressive model order was used in [8] to detect artifacts by minimizing the Akaike information criterion value. It is calculated from the first detailed component of the discrete wavelet transform of a segment and has demonstrated effectiveness in quantifying motion artifacts. The Daubechies 4 wavelet function is used in discrete wavelet transform.

3.3.13. Spectral Entropy (H_S). Spectral entropy can be an efficient indicator for PPG signal classification [11, 31]. For an absent or noisy signal, it has a high value. In contrast, the value for a tuned or harmonic signal is low. Spectral entropy is defined in [35] as

$$H_S = - \sum_{i=1}^N x_i^2 * \log_e(x_i^2), \quad (16)$$

where $*$ denotes the convolution of two terms.

3.3.14. Qi-Zheng Energy (QZE). This feature was mentioned in [32] and denotes the outcome of filtering the selected energy sequence for every segment. It generally uses a filter that is designed to be appropriate for identifying the start or endpoints of a given segment. As this feature detects high values at the endpoints, it is complementary to the other features.

3.3.15. Zero Crossing Rate (ZCR). According to [11, 29, 33], this feature denotes the rate of change of the sign of the filtered signal and is defined as

$$ZCR = \frac{Z_n^c}{N}, \quad (17)$$

where Z_n^c is the number of zero crossings in a PPG segment.

3.3.16. Kaiser-Teager Energy (KTE). It was reported in [11, 36] that the Kaiser-Teager energy (KTE) of a signal is a standard tool that can classify anomalous and acceptable signals by separating noise, transients, and artifacts from the signal. The mean ($KTE\mu$), variance ($KTE\sigma^2$), interquartile range (KTE_{IQR}), and skewness (KTE_{skew}) of the KTE of a PPG segment are computed as four features.

$$KTE_i = x_i^2 - x_{i+1}x_{i-1}, \quad (18)$$

where $i = 2, 3, 4, \dots, N$.

$$KTE\mu = \frac{1}{N-2} \sum_{i=2}^{N-1} KTE_i, \quad (19)$$

$$KTE\sigma^2 = \frac{1}{N-2} \sum_{i=2}^{N-1} (KTE_i - KTE\mu)^2, \quad (20)$$

$$KTE_{IQR} = KTE_{Q_3} - KTE_{Q_1}, \quad (21)$$

where KTE_{Q_1} and KTE_{Q_3} denote the first and third quartiles of the KTE, respectively.

$$KTE_{skew} = \frac{(1/(N-2)) \sum_{i=2}^{N-1} (KTE_i - KTE\mu)^3}{(KTE\sigma)^3}, \quad (22)$$

where $KTE\sigma$ is the standard deviation of the KTE of a PPG segment.

3.4. Feature Selection Algorithm. In addition to the preprocessing step, feature selection is also important for machine learning and pattern recognition problem [37–39]. A large number of features is not always beneficial for classification tasks; instead, it may increase complexity and be irrelevant to the problem. Therefore, the feature selection step increases learning accuracy by reducing the computational complexity, removing redundant features, and improving the comprehensibility of the obtained results [35]. This step reduces the dimensionality problem and can thus facilitate model simplicity. This can help researchers interpret model results more easily. In this study, we utilized the Fisher score feature selection algorithm to order and select important features for a machine learning model to detect PPG segments containing artifacts.

The Fisher score is a supervised feature selection method that is typically used for binary classification problems. The algorithm identifies features that can separate samples into different classes. The algorithm selects features independently by following the scores. A threshold value is

calculated by taking the average of the Fisher scores to select the effective feature set among all features. If the Fisher score of a feature is greater than the threshold value, the feature is selected as an important feature. The variation in sample points of different classes is as large as possible, whereas the variation in sample points of the same class is as small as possible [40]. The Fisher score is obtained using

$$FS = \frac{1}{(\sigma_f^j)^2} \sum_{k=1}^2 n_k (f_{j,k}^- - \bar{f}_j)^2, \quad (23)$$

where $(\sigma_f^j)^2$ is the variance of the j^{th} feature, n_k is the number of segments in a class, $f_{j,k}^-$ is the mean of the j^{th} feature in the k^{th} class, and $\bar{f}_j$ is the mean of the j^{th} feature.

3.5. Machine Learning Model. Based on the 19 extracted feature vectors, we used six machine learning approaches to detect artifacts and classify PPG segments. In this study, we selected the model parameters after comparing them with different values.

3.5.1. Random Forest Model (RF). RF is an ensemble learning technique for classification, regression, and many other tasks that are performed by building a myriad of decision trees at the time of training. It produces a class that represents different classes, known as classification [41]. In our experiment, we used the default RF implementation of scikit-learn with the extracted features as input. The model performed best for 650 estimators. scikit-learn calculates a node's importance by calculating the Gini importance for each binary decision tree. The combined prediction from all the trees makes the final decision. The Gini importance is given in

$$i_j = w_j c_j - w_{lj} c_{lj} - w_{rj} c_{rj}, \quad (24)$$

where i_j denotes the importance value and c_j is the impurity of node j . w_j is the number of weighted samples that reach node j . lj and rj are the child node from the left and right split of the node, respectively.

3.5.2. Decision Tree (DT). The DT algorithm is widely applied when model comprehensibility is required. It is one of the simplest and most popular algorithms for a task. This algorithm is a decision-making method that has a tree-like structure for modeling decisions and probable consequences. The features, decision rules, and results are specified by the internal nodes, branches, and leaves of each node, respectively [42]. Similar to RF, it also calculates the Gini importance in scikit-learn using (24). But the difference is that it relies on singular decision.

3.5.3. Artificial Neural Network (ANN). An ANN generally learns to perform operations by studying examples and is usually not programmed with any task-specific rules. An ANN is constructed on a collection of connected units or nodes called neurons. In this study, the ANN consisted of six hidden layers along with an input and output layer.

FIGURE 7: Architecture of the ANN model.

The performance of an ANN is highly dependent on the number of hidden layers and hidden neurons. The hidden layers in our model had 200, 150, 80, 40, 20, and 16 neurons, respectively.

The architecture of the ANN model is shown in Figure 7. The input are the selected features that go to the six hidden layers and give two outputs. We selected output 01 for anomalous signal windows and 10 for acceptable signal windows.

The activation function used was ReLu. The ANN was trained with 350 epochs and a batch size of 10. Categorical cross-entropy was selected as the loss function with the Adam optimizer. All parameters were selected experimentally.

For the ANN, the features were standardized and normalized because neural networks require the same range of feature values to provide a uniform impact on the network [43]. Equations (25) and (26) were used for standardization ($\text{Standard}_{f_{sj}}$) and normalization ($\text{Normal}_{f_{sj}}$):

$$\text{Standard}_{f_{sj}} = \frac{f_{sj} - \bar{f}_j}{\sigma_{f_j}}, \quad (25)$$

where f_{sj} denotes the j^{th} feature of the s^{th} PPG segment, $\bar{f}_j$ is the mean, and σ_{f_j} is the standard deviation of the j^{th} feature. Here, $j = 1, 2, 3, \dots, 19$ and $s = 1, 2, 3, \dots, 1998$.

$$\text{Normal}_{f_{sj}} = \frac{\sinh(\text{Standard}_{f_{sj}})}{\cosh(\text{Standard}_{f_{sj}})}, \quad (26)$$

Here, normalization is performed by calculating the hyperbolic tangent function of the standardized features. This normalization technique scales the feature vector to the range of -1 to $+1$.

3.5.4. Probabilistic Neural Network (PNN). A PNN is a feed-forward neural network that is specifically designed for classification and pattern recognition problems. It has input and output layers coupled with two hidden layers. The first is called the pattern layer while the second is called the

summation layer [44]. We used the PNN algorithm of the NeuPy Python library with a batch size of 3. The distance between a sample and the input is calculated using [8]

$$d_i = \frac{1}{\sigma^s (2\pi)^{s/2}} \times \frac{1}{n} \sum_{k=1}^n e^{-(x-x_i)^T (x-x_i)/2\sigma^2}, \quad (27)$$

where k denotes the input parameter number, n denotes total inputs' number, the input sample is x_i , s denotes the dimension of the measuring sample, and σ denotes the amount of variance. The sum of the distances is calculated for the inputs to obtain the output probabilities.

Standardization and normalization were performed sequentially for the PNN using the same formulas as for the ANN.

3.5.5. Gaussian Naïve Bayes (GNB). Naïve Bayes classifiers are a group of simple probabilistic classifiers on the basis of Bayes' theorem. These algorithms are used to classify binary or multiclass data. They provide a strong or naïve conjecture of liberty to the data to be classified. The GNB classifier is the simplest model of Bayesian networks and provides high scalability [45]. We used the simplest GNB model for our data using python. The probability density p of a given observation v for a given class c can be computed using

$$p(v | c) = \frac{1}{\sqrt{2\pi\sigma^2}} \times e^{-(v-\mu)^2/2\sigma^2}, \quad (28)$$

where μ and σ are the mean and variance of all the observations, respectively. Standardization and normalization were performed sequentially with the data for better results using the same equations as for the ANN.

3.5.6. Linear Support Vector Machine (LSVM). The support vector machine (SVM) is a learning method that can classify a dataset into multiple categories. The goal of the SVM is to determine the best separating hyperplane. The SVM is a powerful classifier and is mainly used for classification, pattern recognition, and regression for its high accuracy and robustness [46]. In this study, we applied SVM, which performed well for a C value of 2^5 . C is a tunable parameter that can determine the number of decision boundaries as well as the number of misclassified features. We used a linear kernel. The kernel can be expressed using

$$K(x_1, x_2) = (C + x_1^T x_2), \quad (29)$$

where C is the constant term and x_1 and x_2 are two points in the hyperplane. The dot product of the two points calculates the distance between them.

Standardization and normalization were performed sequentially using the same equations as for the ANN to keep the feature value in the same range.

4. Experimental Results

We used six methods for performance evaluation: model accuracy, 10-fold cross-validation, standard deviation of

TABLE 3: Score of feature importance.

No.	Notation	Importance score
F17	$KTE\sigma^2$	0.23078
F16	$KTE\mu$	0.19048
F18	KTE_{IQR}	0.13860
F3	S_{skew}	0.13610
F11	$MSD\mu$	0.08795
F6	N_{AVPG}	0.07854
F15	ZCR	0.04432
F9	MAD_{amp}	0.03455
F5	S_K	0.03328
F7	SD_{amp}	0.03266
F13	H_S	0.02830
F10	SR	0.02608
F4	S_{IQR}	0.00986
F8	SD_{CPI}	0.00915
F14	QZE	0.00830
F19	KTE_{skew}	0.00732
F12	AR_O	0.00265
F2	$S\sigma^2$	0.00217
F4	$S\mu$	0.00008

the accuracy of the k folds, a confusion matrix, computational time, and sensitivity and specificity analysis.

In 10-fold cross-validation, the PPG segments were divided into 10 subclusters. With this method, nine subclusters were used for training, while one subcluster was used to test the trained model. The method was performed until the training and testing of all subclusters were completed. The average result of testing for 10-fold is considered a single accuracy value. This technique is generally used in machine learning to compare and select a model to estimate its performance on new data. Cross-validation is mainly a resampling process that is used to evaluate machine learning models on a limited data sample. It is a common method, as it is easy to understand. This method usually results in a less biased or less optimistic estimate of model ability than other approaches, such as an easy train and test split.

4.1. Result of Feature Selection Algorithm. We aimed to identify efficient features from the set of 19 features. To achieve this, the features were ordered using the Fisher score feature selection algorithm based on the obtained score. The score denotes the importance of a feature for the problem.

The ordering of the features using the Fisher score algorithm is presented in Table 3 with the importance value, where a higher importance value for a feature indicates higher importance in signal classification. There were no features with zero importance values.

All six machine learning models (RF, DT, ANN, PNN, GNB, and LSVM) were evaluated using a 10-fold cross-validation method with the 19 ordered features as input.

FIGURE 8: Change of (a) accuracy and (b) standard deviation with an increasing number of features using the Fisher score algorithm.

Figure 8 displays the model accuracy rates and standard deviation using the ordered feature set of the Fisher score algorithm. Figure 8(a) shows the number of features in the x -axis and accuracy (%) in the y -axis. It demonstrates that even using only one feature, some models obtained comparatively high accuracy rates. This indicates that the obtained feature order was efficient. The graph reveals that the RF model outperformed all models with the highest accuracy using 11 features. The accuracy of the RF model increased notably compared to that of other models after seven features. An investigation of the remaining models indicates that the accuracy of the LSVM and GNB was similar from 9 to 12 features. The GNB exhibited better performance with 10 features; however, the accuracy was lower than that of the RF. In contrast, the LSVM model performed well with 14 features. The accuracy rate of the ANN and PNN was greater than 80% after 11 features. Figure 8(a) also reveals that the DT model had the lowest accuracy per feature among all models. As RF performed best with the first 11 features from the ordered 19 features using the Fisher score algorithm, the first 11 features with higher importance value were selected as the final feature set.

Figure 8(b) shows the number of features (x -axis) versus standard deviation (y -axis). A lower standard deviation indicates a proper distribution of data and model efficiency. The LSVM demonstrated the lowest standard deviation with only one feature. However, the value increased from two features, whereas the accuracy increased, as displayed in Figure 8(a). The GNB exhibited a low standard deviation with 19 features. For the ANN, the accuracy value with 17 features was the highest, while the standard deviation was the lowest. In the RF model, there was less fluctuation in the standard deviation, while the DT and PNN produced comparatively higher fluctuation.

As shown in Figure 8(b), though the standard deviation of the random forest model at 11 features is not the lowest, the value is relatively in the lower region, compared to the other models. So there is a trade-off between accuracy and standard deviation. We are taking the highest accuracy, and at that point, standard deviation is not the lowest, but at the lower side. Among all the classifiers, the random forest

FIGURE 9: ROC curve with 11 high ranked features.

TABLE 4: Comparison of performance for PPG signal artifact detection using various machine learning models.

Model	Feature number	Accuracy $\pm$ STD (%)	Sensitivity (%)	Specificity (%)
RF	11	85.68 $\pm$ 2.43	86.57	85.09
DT	14	77.60 $\pm$ 2.74	69.67	82.35
ANN	17	82.07 $\pm$ 3.00	73.09	86.75
PNN	17	81.27 $\pm$ 2.87	72.09	86.25
GNB	11	79.93 $\pm$ 3.44	64.8	85.79
LSVM	18	80.53 $\pm$ 4.34	66.25	87.90

shows a balance between standard deviation and accuracy with the lowest number of high ranked features. So, it is selected as the best model. We showed the accuracy and standard deviation with all the features from 1 to 19 to make the difference visible between combination of high rank features and low rank features.

FIGURE 10: Computational time of six machine learning algorithms listed in Table 4.

FIGURE 11: Confusion matrix of the best model, random forest (RF).

The receiver operating characteristic (ROC) curve is shown in Figure 9 for all the machine learning models with 11 effective features. Here, AUC means area under the ROC curve. The more AUC, the better is the performance of the model. The curve shows that RF has the highest AUC among all the models. It indicates that RF performs best with 11 high ranked features.

4.2. Comparison of Machine Learning Models. A comparison of the performance matrices of all models with the number of features and the standard deviation is displayed in Table 4. It can be seen that the RF model obtained the highest accuracy rate of 85.68% for 11 features with a standard deviation of 2.43. The RF also demonstrated the highest sensitivity value of 86.57%. The model performed better in identifying anomalous PPG segments than the other models.

Following the RF model, the best accuracy was obtained with the ANN, PNN, LSVM, GNB, and DT, in that order. After testing all six machine learning approaches with 10-fold cross-validation, the RF model exhibited the highest performance based on accuracy and standard deviation

and was selected as the best model for detecting motion artifacts in PPG segments of the MIMIC dataset. The ordering of efficient features obtained using the RF model was $KTE\sigma^2$, $KTE\mu$, KTE_{IQR} , S_{skew} , $MSD\mu$, N_{AVPG} , ZCR , MAD_{amp} , S_K , SD_{amp} , and H_S .

Moreover, the computational time of the models in Table 4 is displayed in Figure 10. The GNB model worked the fastest. The DT, PNN, and LSVM models' computational time was a little more than that of GNB. The best-performed RF model in terms of classification accuracy and standard deviation also did not need a very long time to do the whole calculation. The highest computational time was needed for ANN.

The confusion matrix of the proposed RF model is presented in Figure 11. Out of 1044 anomalous and 954 acceptable segments, RF detected 911 anomalous and 797 acceptable segments correctly.

4.3. Result of Proposed Standalone Real-Time Process. Implementation and the result of our proposed real-time PPG signal artifact detection process using a standalone android

FIGURE 12: Implementation of real-time PPG signal artifact detection using an android application.

FIGURE 13: Runtime of the real-time algorithm for every two samples.

application are shown in Figure 12. The algorithm takes input of the PPG signal and shows the result in real time. The signal is taken using the rear camera of the android smartphone. The camera needs to be fully covered by lightly pressing the index finger on the rear camera to get proper PPG signal. The value from red LED is considered PPG data. We used android studio with Java to develop the whole application. After taking the signal from the mobile camera, the processing is done according to Figure 4. To match the sampling frequency of the obtained signal with the previously trained model, linear interpolation is done before the

segmentation. The previously trained random forest model was compiled in a native binary code using the python m2cgen library. The signal is obtained and processed using a HUAWEI Nova 3i android smartphone with a 4 GB RAM and 8 core (2 clusters: 2.19 GHz and 1.71GHz) 64-bit Kirin 710 processor. The processor instruction set was arm64-v8a.

The main complexity for implementing the algorithm is space and time. The standalone android application takes 15.93 kB space in smartphone memory. The time required to detect artifact using the application is shown in

TABLE 5: Comparison with other works.

Work	Model	#Feature/ rule	Threshold value	Dataset	Accuracy (%)	Sensitivity (%)	Specificity (%)
Our work	RF	11	Not used	55 subjects from the MIMIC database	85.68	86.57	85.09
[8]	PNN	6	Not used		60.42	64.20	57.14
[10]	LSVM	4	Not used		70.93	69.45	72.41
[12]	1D-CNN	No feature	Not used		72.56	71.89	73.19
[13]	Rule-based	2	Used	MIMIC-II, BIDMC, PRRB, and self-made	96.5%	—	—
[14]	Rule-based	4	Used	PCinC and JR	—	91%	95%
[15]	SDAE-MLPNN	—	Used	MIMIC-II	95.75%	99.12	93.50%
[16]	Rule-based	#Rule: 4 #Feature: 11	Used	CSL, MIT-BIH SLP, and MIMIC-II	98.09%	98.62%	97.37%

Figure 13. For the real-time implementation, we changed the equiripple FIR filtering process to real-time filtering. We took a single sample to filter the signal instead of taking the whole signal. The PPG signal acquired from the android mobile has a sampling rate of 30 Hz. So two samples acquiring time is 66.6 ms. The total runtime for our proposed standalone real-time algorithm is 57.5 ms. The total time is divided into three steps: signal filtering and interpolation, feature calculation, and prediction of signal artifact using a random forest classifier. After that, there is idle time of 9.1 ms. From Figure 13, it can be said that standalone real-time signal artifact detection is possible with the proposed system.

4.4. Comparison of the Best Model with Related Works. It was difficult to compare our results with the results of related studies, as related studies used customized datasets with their own subjects. Most of the existing studies used healthy subjects for their experiments. However, we implemented the methodology and features of [8, 10, 12] in our dataset, and the obtained results are illustrated in Table 5. We also compared our result with the rule-based methods [13–16] as shown in Table 5.

In [8], the PNN algorithm was used with six features. The model with the selected features was applied in our dataset. As illustrated in Table 5, the obtained accuracy, sensitivity, and specificity were much lower than those of our method.

In [47], 42 features were extracted from the PPG signals of their customized small dataset and an SVM classifier has been used to classify PPG signal into 2 classes (good and bad) with the RBF kernel. They used a small dataset with only 13 patients, and the classification of the PPG signal is only based on atrial fibrillation. They used a large number of features (42 features) which is not computationally efficient. Use of unnecessary features causes redundancy and take more resources than required. No feature selection method was applied. As applying 42 features is not computationally efficient for any system, this method was not applied in our dataset and not listed in Table 5. Rather, we

TABLE 6: Statistical significance test result using “ $5 \times 2cv$ paired t -test”.

Model 1	Model 2	p value
RF	PNN [8]	0.001
RF	LSVM [10]	0.002
RF	1D-CNN [12]	0.001

considered the work described in [10] selected four features and used the SVM with 10-fold cross-validation with linear kernel to classify PPG segments. As they used the linear kernel, they named the model LSVM. So, we applied this method to our dataset and obtained lower accuracy, sensitivity, and specificity for the LSVM than the RF. However, the four selected features alone could detect anomalous segments comparatively better than acceptable segments. But none of the methods have used a feature selection method to identify effective features.

The work of [12] used 1D-CNN with PPG signals from three datasets. We implemented their network with our dataset and obtained 70.93% accuracy. The network’s specificity and sensitivity rates were also lower than our selected model for our dataset. It might be due to the selection of data. Therefore, the network is not generalizable for all kinds of data. Moreover, the 1D-CNN needs a large computational time which is not efficient, whereas a feature-based method needs very negligible time compared to that and gives better accuracy and balanced performance in our dataset. We focused on the shape of PPG signals and considered artifacts on both frequency and time domain, so feature selection was important for our method.

In [13–16], several constant threshold values have been used to check the PPG signal quality. The threshold values worked well for their chosen processed dataset, and so the accuracy is good. But there is a varied range of PPG data. So, the threshold value might not work for all kinds of PPG signals. Moreover, in [15], a stacked denoising autoencoder (SDAE) is used with each PPG beat to extract multiple

FIGURE 14: Misclassification error with an increasing number of estimators for the RF model.

features from the beat, and then, the extracted features are fed into a multilayer perceptron neural network (MLPNN) to get the acceptable beats. But the number of features used or the process of choosing the features is not mentioned. Also, the classification rule using the threshold value is not explained properly. [14–16] need beat segmentation for their algorithm. However, beat segmentation in intense noisy signals is a very difficult process, and fair success is observed in limited cases. Additionally, [13, 14, 16] used signals from combination of multiple datasets. The PPG signals of different datasets are acquired using different measurement devices. So, the voltage level and wavelength vary due to the difference in dataset and measurement devices. Hence, signals of each dataset need individual preprocessing techniques and need to be taken under similar criteria to determine the common threshold level. Furthermore, the same amount of data need to be taken from each dataset so that the threshold is not biased for a specific set of data. But none of the works considered the data inconsistency problem and used common threshold values.

4.5. Statistical Significance Test Result. We evaluated the RF model using a commonly used resampling method called 10-fold cross-validation. Although it directly calculates and compares the mean accuracy score, sometimes this evaluation process may be misleading due to its difficulty to understand that the obtained result is fair or a statistical fluke. The statistical significance test can solve this issue by determining whether one classification algorithm outperforms another one on a specific classification task. The rejection of the null hypothesis by a statistical significant test indicates that the difference in accuracy is statistically significant and real.

So, to show the fairness of our obtained result, we did a $5 \times 2cv$ paired t -test with the proposed RF model and the two compared papers listed in Table 5. This method was proposed in [48] that claimed to overcome the shortcomings of other statistical significant test methods. In this method, the dataset was divided into 2 parts: 50% training and 50% testing, and the division was repeated for 5 times. Each time

we took two models: RF (model A) and PNN [8] (model B), RF (model A) and LSVM [10] (model B), or RF (model A) and 1D-CNN [12] (model B). A and B models were fitted with the training set, and then, their performance was evaluated using the test set. Then, the training set and test set were swapped and the performance was computed again. Finally, from the two performance differences, the p value was calculated using the equations of [48]. The p values are shown in Table 6.

The result denotes that the p values are less than the chosen statistical significance level of 0.05. It means that the null hypothesis is rejected and a significant difference is presented between our proposed model and the compared models.

5. Discussion

Our result was comparatively better than the compared papers of Table 5 for our dataset. The simple RF method performed well for the ICU dataset of the MIMIC database. The main problem was the selection of acceptable and anomalous signals in the huge dataset. Our target was to propose an easy way to detect the artifacts of the PPG signal that will be limited not only for the healthy subjects but also for the ICU patients. The proposed RF model can successfully detect involuntary artifacts in ICU patients. This is important because recently a large number of researches are being conducted with the publicly available datasets containing PPG signals of ICU patients to automatically detect premature ventricular contractions [49], to estimate breathing rate [50], blood pressure [51], blood oxygen saturation [52], heart rate, respiration [7], arterial fibrillation [53], and so on. An artifact-prone PPG signal will give unreliable and false results for these kinds of important measurements. Hence, it is essential to detect the artifacts of PPG signals, and the machine learning method is an efficient and easy way for this purpose. Our chosen classification algorithm performed quite well in comparison to the other methods. So, our proposed approach can be used in the preprocessing module of different PPG-based applications that use the

signal to get various bodily parameters and can achieve reliable results. A standalone real-time artifact detection algorithm is proposed for this purpose.

Our adopted RF model was evaluated starting from 10 estimators to 1100 estimators increasing 10 estimators each time as shown in Figure 14. And in order to understand the inconsistent behaviour of the misclassification error data, we also performed 2nd-order polynomial fit. This polynomial fit curve shows the trend of the error plot, which is gradually decreasing from the start, and at 650 estimators, the error ceases to go lower. On the other hand, the computational time also increases linearly with the increasing number of estimators. So, taking care of the trade-off between the error and computational time, we selected the number of estimators at 650 as the most optimized value for this study.

However, the result depends on the selection of proper features for the model, and it is a feature-dependent approach which is the main limitation of our method. We focused on features and different machine learning algorithm to express the effectiveness. Our artifact detection was based on the shape of the PPG signal that depends on both the time domain (i.e., noise and motion) and frequency domain (i.e., shape distortion and absence of notch), because many kinds of artifacts affect the frequency domain. The classification on the basis of the time domain cannot identify all the artifacts correctly. So, for this research, we focused on feature selection other than deep learning approach. In the future, we want to use our real-time algorithm for measuring different PPG-based biological parameters.

6. Conclusions

In this paper, various machine learning methods are presented and compared to detect all kinds of artifacts in PPG segments and classify the segments into acceptable and anomalous categories. The acceptable signals will enhance the reliability of the result to measure PPG-based parameters. Our focus was to meet the existing challenge of detecting all kinds of artifacts in this highly studied biosignal. We used the publicly available MIMIC database which is the most popular biosignal database and is mainly used in research. In this study, we examined the performance of six machine learning methods and proposed the best method for detecting artifacts in the PPG signals in this database.

Feature selection is an important step for this purpose, as detection is dependent on the selection of exact features for accurate classification. To select 11 high ranked important features, we used the Fisher score algorithm. Then, a 10-fold cross-validation method was performed for model verification. Among all machine learning methods, the RF classifier exhibited the best performance in terms of accuracy for ICU patients of the MIMIC dataset. We also proposed a standalone real-time artifact detection algorithm using our method that is implemented to detect artifacts of the PPG signal acquired using an android smartphone. In conclusion, it can be stated that the proposed method is a simple and efficient process for detecting artifacts in the PPG signal.

We hope that the proposed method will be able to help the researchers to get reliable PPG-driven parameters.

Data Availability

Publicly available datasets were analyzed in this study. This data can be found at <https://physionet.org/content/mimicdb/1.0.0/> (accessed on 11 March 2021).

Conflicts of Interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

Acknowledgments

This work was supported by the National Research Foundation of Korea (NRF) Grant funded by the Korean Government (MSIT) (Nos. 2016R1A5A1012966 and 2021R1F1A1062285).

References

- [1] D. Pollreis and N. TaheriNejad, "Detection and removal of motion artifacts in PPG signals," *Mobile Networks and Applications*, 2019.
- [2] J. Allen, "Photoplethysmography and its application in clinical physiological measurement," *Physiological Measurement*, vol. 28, no. 3, pp. R1–39, 2007.
- [3] D. Yang, J. Zhu, and P. Zhu, "SpO2 and heart rate measurement with wearable watch based on PPG," in *2015 IET International Conference on Biomedical Image and Signal Processing (ICBISP 2015)*, pp. 1–5, Beijing, China, Nov. 2015.
- [4] K. T. Sweeney, T. E. Ward, and S. F. McLoone, "Artifact removal in physiological signals—practices and possibilities," *IEEE Transactions on Information Technology in Biomedicine*, vol. 16, no. 3, pp. 488–500, 2012.
- [5] G. B. Moody and R. G. Mark, "Integration of real-time and off-line clinical data in the MIMIC database," in *Computers in Cardiology 1997*, pp. 585–588, Lund, Sweden, 1997.
- [6] D. Liu, M. Görges, and S. A. Jenkins, "University of Queensland Vital Signs dataset," *Anesthesia & Analgesia*, vol. 114, no. 3, pp. 584–589, 2012.
- [7] M. A. Motin, C. K. Karmakar, and M. Palaniswami, "An EEMD-PCA approach to extract heart rate, respiratory rate and respiratory activity from PPG signal," in *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 3817–3820, Orlando, FL, USA, 2016.
- [8] F. Tabei, R. Kumar, T. N. Phan, D. D. McManus, and J. W. Chong, "A novel personalized motion and noise artifact (MNA) detection method for smartphone photoplethysmograph (PPG) signals," *IEEE Access*, vol. 6, pp. 60498–60512, 2018.
- [9] S. K. Bashar, D. Han, A. Soni, D. D. McManus, and K. H. Chon, "Developing a novel noise artifact detection algorithm for smartphone PPG signals: preliminary results," in *2018 IEEE EMBS International Conference on Biomedical Health Informatics (BHI)*, pp. 79–82, Las Vegas, NV, USA, Mar. 2018.
- [10] W.-J. Lin and H.-P. Ma, "A physiological information extraction method based on wearable PPG sensors with motion artifact removal," in *2016 IEEE International Conference on*

- Communications (ICC)*, pp. 1–6, Kuala Lumpur, Malaysia, May 2016.
- [11] E. Monte-Moreno, “Non-invasive estimate of blood glucose and blood pressure from a photoplethysmograph by means of machine learning techniques,” *Artificial Intelligence in Medicine*, vol. 53, no. 2, pp. 127–138, 2011.
 - [12] C.-H. Goh, L. K. Tan, N. H. Lovell, S.-C. Ng, M. P. Tan, and E. Lim, “Robust PPG motion artifact detection using a 1-D convolution neural network,” *Computer Methods and Programs in Biomedicine*, vol. 196, article 105596, 2020.
 - [13] S. Alam, R. Gupta, and K. D. Sharma, “On-board signal quality assessment guided compression of photoplethysmogram for personal health monitoring,” *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1–9, 2021.
 - [14] C. Orphanidou, T. Bonnici, P. Charlton, D. Clifton, D. Vallance, and L. Tarassenko, “Signal-quality indices for the electrocardiogram and photoplethysmogram: derivation and applications to wireless monitoring,” *IEEE Journal of Biomedical and Health Informatics*, vol. 19, no. 3, pp. 832–838, 2014.
 - [15] M. Singha Roy, B. Roy, R. Gupta, and K. Das Sharma, “On-device reliability assessment and prediction of missing photoplethysmographic data using deep neural networks,” *IEEE Transactions on Biomedical Circuits and Systems*, vol. 14, no. 6, pp. 1323–1332, 2020.
 - [16] S. Vadrevu and M. S. Manikandan, “Real-time PPG signal quality assessment system for improving battery life and false alarms,” *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 66, no. 11, pp. 1910–1914, 2019.
 - [17] Y.-C. Yang, W.-K. BEH, Y.-C. Lo, A.-Y. A. Wu, and S.-J. Lu, “ECG-aided PPG signal quality assessment (SQA) system for heart rate estimation,” in *2020 IEEE workshop on signal processing systems (SiPS)*, pp. 1–6, Coimbra, Portugal, Oct. 2020.
 - [18] T. Athaya and S. Choi, “Evaluation of different machine learning models for photoplethysmogram signal artifact detection,” in *2020 International conference on information and communication technology convergence (ICTC)*, pp. 1206–1208, Jeju, Korea (South), Oct. 2020.
 - [19] A. A. Alian and K. H. Shelley, “Photoplethysmography,” *Best Practice & Research Clinical Anaesthesiology*, vol. 28, no. 4, pp. 395–406, 2014.
 - [20] M. Elgendi, “On the analysis of fingertip photoplethysmogram signals,” *Current Cardiology Reviews*, vol. 8, no. 1, pp. 14–25, 2012.
 - [21] E. Ghosh and C. M. P. Blandon, “Method to quantify photoplethysmogram (PPG) signal quality,” 2017, <https://patents.google.com/patent/WO2017089921A1/en>.
 - [22] N. Paradkar and S. R. Chowdhury, “Cardiac arrhythmia detection using photoplethysmography,” in *2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 113–116, Jeju, Korea (South), Jul. 2017.
 - [23] G. B. Moody and R. G. Mark, “The MIMIC database,” 1992, <http://physionet.org/>.
 - [24] G. B. Moody and R. G. Mark, “A database to support development and evaluation of intelligent intensive care monitoring,” in *Computers in Cardiology 1996*, pp. 657–660, Indianapolis, IN, USA, Sep. 1996.
 - [25] T L for C. Physiology, *wfdb: The WFDB Python Toolbox* July 2020, <https://github.com/MIT-LCP/wfdb-python>.
 - [26] W. Waugh, J. Allen, J. Wightman, A. J. Sims, and T. A. W. Beale, “Novel signal noise reduction method through cluster analysis, applied to photoplethysmography,” *Computational and Mathematical Methods in Medicine*, vol. 2018, Article ID 6812404, 8 pages, 2018.
 - [27] K. Vandecasteele, J. Lázaro, E. Cleeren et al., “Artifact detection of wrist photoplethysmograph signals,” *BIOSIGNALS*, 2020.
 - [28] R. Krishnan, B. Natarajan, and S. Warren, “Two-stage approach for detection and reduction of motion artifacts in photoplethysmographic data,” *IEEE Transactions on Biomedical Engineering*, vol. 57, no. 8, pp. 1867–1876, 2010.
 - [29] M. Elgendi, “Optimal signal quality index for photoplethysmogram signals,” *Bioengineering*, vol. 3, no. 4, p. 21, 2016.
 - [30] S. M. Thampi, A. Gelbukh, and J. Mukhopadhyay, *Advances in Signal Processing and Intelligent Recognition Systems*, Springer International Publishing, 2014.
 - [31] N. Selvaraj, Y. Mendelson, K. H. Shelley, D. G. Silverman, and K. H. Chon, “Statistical approach for the detection of motion/noise artifacts in photoplethysmogram,” in *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 4972–4975, Boston, MA, USA, Aug. 2011.
 - [32] Qi Li, Jinsong Zheng, A. Tsai, and Qiru Zhou, “Robust endpoint detection and energy normalization for real-time speech and speaker recognition,” *IEEE Transactions on Speech and Audio Processing*, vol. 10, no. 3, pp. 146–157, 2002.
 - [33] M. Elgendi, R. Fletcher, Y. Liang et al., “The use of photoplethysmography for assessing hypertension,” *NPJ Digital Medicine*, vol. 2, no. 1, 2019.
 - [34] M. Elgendi, Y. Liang, and R. Ward, “Toward generating more diagnostic features from photoplethysmogram waveforms,” *Diseases*, vol. 6, no. 1, p. 20, 2018.
 - [35] R. R. Coifman and M. V. Wickerhauser, “Entropy-based algorithms for best basis selection,” *IEEE Transactions on Information Theory*, vol. 38, no. 2, pp. 713–718, 1992.
 - [36] A.-O. Boudraa and F. Salzenstein, “Teager-Kaiser energy methods for signal and image analysis: a review,” *Digital Signal Processing*, vol. 78, pp. 338–375, 2018.
 - [37] A. Onan and S. Korukoğlu, “A feature selection model based on genetic rank aggregation for text sentiment classification,” *Journal of Information Science*, vol. 43, no. 1, pp. 25–38, 2017.
 - [38] A. Onan, “Classifier and feature set ensembles for web page classification,” *Journal of Information Science*, vol. 42, no. 2, pp. 150–165, 2016.
 - [39] B. Cao, D. Shen, J.-T. Sun, Q. Yang, and Z. Chen, “Feature selection in a kernel space,” in *Proceedings of the 24th international conference on Machine learning*, pp. 121–128, Corvallis, Oregon, USA, Jun. 2007.
 - [40] Q. Gu, Z. Li, and J. Han, “Generalized Fisher score for feature selection,” 2020, <http://arxiv.org/abs/1202.3725>.
 - [41] E. Scornet, “Random forests and kernel methods,” 2020, <http://arxiv.org/abs/1502.03836>.
 - [42] H. Sharma and S. Kumar, “A survey on decision tree algorithms of classification in data mining,” *International Journal of Science and Research (IJSR)*, vol. 5, no. 4, pp. 2094–2097, 2016.
 - [43] S.-N. Yu and Y.-H. Chen, “Electrocardiogram beat classification based on wavelet transformation and probabilistic neural network,” *Pattern Recognition Letters*, vol. 28, no. 10, pp. 1142–1150, 2007.

- [44] B. Mohebbali, A. Tahmassebi, A. Meyer-Baese, and A. H. Gandomi, "Chapter 14 - probabilistic neural networks: a brief overview of theory, implementation, and application," in *Handbook of Probabilistic Models*, P. Samui, D. T. Bui, S. Chakraborty, and R. C. Deo, Eds., Butterworth-Heinemann, 2020.
- [45] H. Zhang, "The optimality of naive Bayes," in *Florida Artificial Intelligence Research Society Conference*, Miami Beach, Florida, USA, 2004.
- [46] M. Awad and R. Khanna, "Support vector machines for classification," in *Efficient Learning Machines: Theories, Concepts, and Applications for Engineers and System Designers*, M. Awad and R. Khanna, Eds., pp. 39–66, Apress, Berkeley, CA, 2015.
- [47] T. Pereira, K. Gadhoumi, M. Ma et al., "A supervised approach to robust photoplethysmography quality assessment," *IEEE Journal of Biomedical and Health Informatics*, vol. 24, no. 3, pp. 649–657, 2020.
- [48] T. G. Dietterich, "Approximate statistical tests for comparing supervised classification learning algorithms," *Neural Computation*, vol. 10, no. 7, pp. 1895–1923, 1998.
- [49] A. Sološenko, A. Petrėnas, and V. Marozas, "Photoplethysmography-based method for automatic detection of premature ventricular contractions," *IEEE Transactions on Biomedical Circuits and Systems*, vol. 9, no. 5, pp. 662–669, 2015.
- [50] P. H. Charlton, D. A. Birrenkott, T. Bonnici et al., "Breathing rate estimation from the electrocardiogram and photoplethysmogram: a review," *IEEE Reviews in Biomedical Engineering*, vol. 11, pp. 2–20, 2018.
- [51] L. Wang, W. Zhou, Y. Xing, and X. Zhou, "A novel neural network model for blood pressure estimation using photoplethysmography without electrocardiogram," *Journal of Healthcare Engineering*, vol. 2018, Article ID 7804243, 9 pages, 2018.
- [52] T. Y. Abay and P. A. Kyriacou, "Photoplethysmography for blood volumes and oxygenation changes during intermittent vascular occlusions," *Journal of Clinical Monitoring and Computing*, vol. 32, no. 3, pp. 447–455, 2018.
- [53] A. Sološenko, A. Petrėnas, V. Marozas, and L. Sörnmo, "Modeling of the photoplethysmogram during atrial fibrillation," *Computers in Biology and Medicine*, vol. 81, pp. 130–138, 2017.