Turkish Journal of Computer and Mathematics Education Vol.12 No. 11 (2021), 115- 121
Research Article

Adaptive And Fault Tolerant Random Forest-Based Method For Heart-Rate Observing Utilizing Photoplethysmography In Healthcare

G. Jasmine Christabel[1], A.C. Subhajini[2*]
[1,2]Department of Computer Application, Noorul Islam Centre for Higher Education, Thuckalay, Kanyakumari, India, 629180.
E.mail: acsubhajini@gmail.com[*], jasminejcs@gmail.com

Article History: Received: 11 January 2021; Revised: 12 February 2021; Accepted: 27 March 2021; Published online: 10 May 2021

ABSTRACT: In recent times, healthcare IoT has proven to be highly effective in reducing medical and hospital resource stress produced by an ageing population. The health-care system's quick reply is critical because it is a safety-critical process. Fog computing, which deploys healthcare IoT devices on the edges of clouds, is a strategic solution for meeting the low bandwidth criterion. Such fog devices, on the other hand, consume a huge amount of data sources. Developing a specific system for fog devices in order to make sure secure data transmission and fast data analysis has become a critical issue. In this research, Random Forest Feature Selection (RFFS) algorithm is used to improve dependability of data transmission and processing prompt. Functionalities of RFFS comprise fault-tolerant data transmission, self-adaptive cleaning and data-load-deduction processing. The assessment of heart rate (HR) depends on wear devices is of attention in suitability. Photoplethysmography (PPG) is an auspicious method to evaluate HR owing to its lower cost; nevertheless, it is simply tainted by motion artifacts (MA). In our study, a robust method depend on random forest is projected for precisely assessing HR from the photoplethysmography sign soiled by intense movement artifacts exactly, a dependable transmission mechanism, achieved by a self-adaptive sieve, will recall lost or imprecise data inevitably. Over extensive simulations, we display that our anticipated scheme expands network dependability, and delivers a quicker processing rapidity. The suggested method is more accurate and resistant to extreme motion artefacts, suggesting that it could be used in smart technologies for fitness and health monitoring.

Keywords: Fog device; fault tolerance, heart rate (HR); photoplethysmography (PPG); random forest

INTRODUCTION

Senior citizens have seen changes in healthcare IoT programmed in recent times. These advanced technologies control food quality, nutrition, physical exercise, physiological status, and other variables to provide integrated remote healthcare services [8]. With the advancements, the need for collecting and transferring large amounts of data with lower latency has increased. Since sensor data takes too long to reach core storage and processing nodes, conventional cloud computing cannot fulfil the criteria as a time-sensitive platform [5]. In recent times, the principle of fog computing has been introduced to solve the issue. Fog systems utilize sensor data as well as provide rigorous process of data for e-Health clients, such as local hospitals and other healthcare services, rather than transmitting data to core nodes, reducing latency substantially [3]. Many constraints, however, can limit the development and deployment of fog computing-enabled healthcare IoT systems. When sensors are embedded in everyday items, for example, the network size grows drastically, resulting in vast quantities of data [4]. On the other hand, strong processors in fog systems are difficult to come by. One of the most pressing problems to be tackled is how to process large amounts of complex data quickly and effectively with minimal computing resources. Breaking data into small parts is a popular approach to dealing with the big data analytics issue [6].
According to its important functions in managing the training load or tracking physiologic factors during daily tasks, heart rate (HR) calculation using wearable devices is critical. Photoplethysmography (PPG) is a common technique because it needs less hardware and is less costly than traditional electrocardiography (ECG) [1]. PPG, but at the other hand, is vulnerable to motion artefacts (MA), which could become very high during subjects' vigorous physical activity, making HR calculation with PPG difficult. Different methods, like independent component analysis (ICA), adaptive filtering, and empirical mode decomposition (EMD), were studied to efficiently eradicate MA in the existence of heavy MA [7]. Furthermore, one kind of hybrid process, that combined two MA removal methods, was suggested to enhance denoising efficiency. This mixture may result in high computational effort, as the use of second method may result in excessive computation when the very first method has lessened considerably noise.

To calculate HR from either a PPG signal polluted by extreme motion artefacts, a vigorous HR prediction model based on random forest is suggested in this study [2]. Its goal is to get rid of MA and then get a clean PPG signal. The suggested MA removal method, which uses a random forest-based feature selection technique, not only improves denoising efficiency but also has a lower computational complexity. Unlike the correlation coefficient-based binary decision method, which can only identify linear relationships, the suggested binary decision algorithm is used to detect not only linear relationships that use the correlation coefficient as one aspect, but also nonlinear relationships that used a variety of other useful procedures, providing a reliable decision result and therefore enhancing the denoisiability.

PROPOSED METHODOLOGY

The utilized method MA removal approach, which combines different MA elimination methods, can not only boost denoising efficiency but also retain a lower computational complexity by using a random forest-based procedure. Unlike the correlation coefficient-based binary decision algorithms, which can only identify linear relationships, the theoretical binary decision method can detect both linear and nonlinear relationships and use a variety of other useful procedures, providing a reliable decision result and therefore enhancing denoising [9].

The suggested framework's computing environment is at the heart of the design process. As previously stated, the proposed solution is made up of three components:

1. The object data is collected at pre-determined intervals by wearable body sensors and a body sensor network. Wearable sensors, on the other hand, have limited knowledge, power, processing, and communication capabilities, and thus cannot handle real-time data. As a consequence, data is collected either when an occurrence is observed or at pre-determined intervals.
2. The data obtained by the sensors is maintained on a remote server, which also acts as a backup for information storage and communication amid the data and the fog computing system.
3. The third aspect is the fog computing platform, which uses fog computing interventions to enhance contact between sensing devices and the device in order to identify operations.

The fog processing element, in turn, conducts online processing. That being said, due to the difficulties in collecting live data from commercial sensors, we are considering offline data analysis for this analysis. The concept behind network virtualization nodes is to take advantage of the ensembles' inherent parallelism by hosting separate outfits on separate fog nodes. As opposed to cloud processing elements, fog computing units have limited processing power capacity. Furthermore, the computing that is predominantly conducted on the fog node, namely algorithms and statistical techniques, is believed to be of medium computation time. Data is removed from the fog node until it has been interpreted and then sent to the cloud services environment, as it acts as an intermediate.

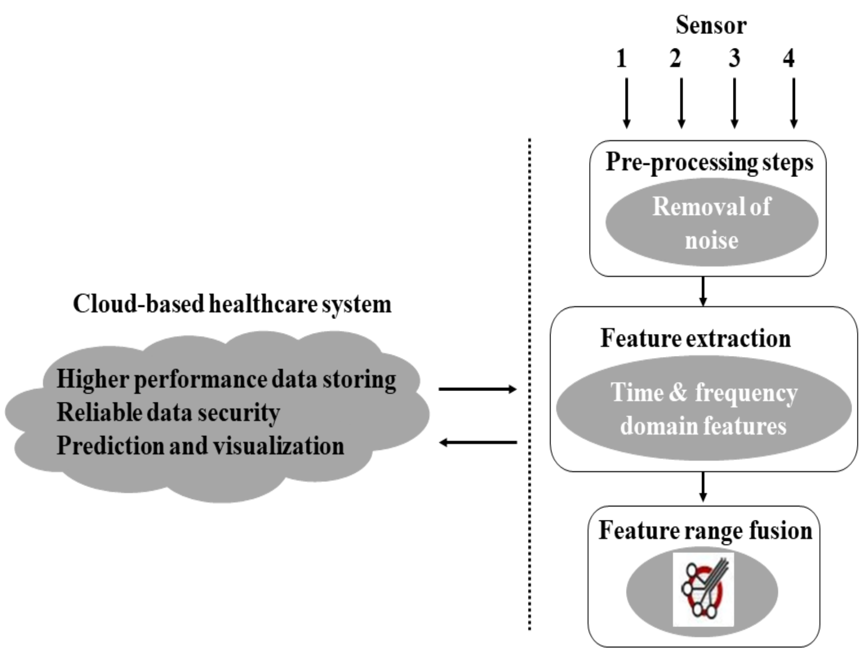

Figure 1: The projected real-time data mixture outline in body sensor channels

Data fusion was done by Random Forest. Let X = $\{X_1,\ldots\ldots,X_n\}$ be the group of random variables, and Y = $\{Y_1,\ldots,Y_n\}$ be the type of rejoinders. A function f (x) = E[Y|X = x] forecasts the response Y for the random variable X. The archives in the dataset D = $\{(X_1, Y_1),\ldots,(X_n,Y_n)\}$ of $[0, 1]^d$ x R-values are independent sets of the arrangement (X, Y), where $E[Y^2] <\alpha$. We utilize immeasurable random forest to calculate f α, n : $[0, 1]^d \rightarrow R$ of f , for the dataset D. For a assortment of P random-based trees, the forecast rate for the k-th tree in the assemblage at point x is $f_n(x; \Delta j)$, where $\Delta_1,\ldots, \Delta_P$ are independent random variable quantity of the dataset D. A combined finite forestry is attained by accumulating the productivities from the distinct trees.

Algorithm 1: The summary of the Random Forest Method

Require: Dataset D in the form of (X, Y) pairs, number of trees P, f € {1,…,P}, τ_n € {1,…..,n}, x € $[0, 1]^P$
nsure: Prediction of the random forest at x
for each j € M **do**
  Choose τ_n points from D
  For all l € τ_n, set $\rho_1 = \Phi$, $\rho_0 = [0, 1]^p$ root partition
  Set $\eta_v = 1$, $\psi = 0$ η number of variables; ψ: level
  while $\eta_{v} < \tau_n$ **do**
    if ψ # ϕ **then**
      $\rho \leftarrow$ point x
      if Σ ρ = 1 **then**
        $\rho_\psi \leftarrow \rho_\psi \cup \rho$
      else
        Generate and split the set ρ into $\rho_{A,}$ ρ_B
        $\rho_{\psi+1} \leftarrow \rho_{\psi+1} \cup \rho_A \cup \rho_B$ ρ updated as a result of split into ρ_A and ρ_B
        $\eta_v \leftarrow \eta_v + 1$
      end if
    else
      $\psi \leftarrow \psi + 1$
    end if
  end while
  Compute f(x, Δ_j, D) for x local prediction for x
Compute $f_{P,n}(x, \Delta_{1,\ldots.}\ \Delta_P, D)$ global prediction for x
return

Extracted features are based on the idea of transforming a signal into some kind of feature vector that depicts operation information. The study's characteristics can be divided into two categories: frequency and time domain functions.
(1) Time-domain Contains: These are statistical metrics which are used to characterize frequency spectrum. We look at time-domain features including central tendency and measures of variability in this analysis.
(2) Frequency-domain Models: are critical for evaluating repeated behaviors, which have a major effect on the heart disease diagnosis. Raw impulses, on the other hand, must be converted into frequency distributions using the Fast Fourier Transform.
Properties like energy, entropy, binned propagation, and time among peaks were obtained from frequency-domain data [10]. Total energy is used to forecast heart disease since it is regarded a global activity. The following are the descriptions of function computation:
(1) Energy: The energy function is computed using the sum of squared FFT magnitudes.
(2) Entropy: The entropy distinguishes between activities with the similar energy function values. The normalized entropy of FFT elements is used to measure the entropy.
(3) Binned Spread: The binned variation is determined by calculating the FFT histogram. This is accomplished by determining the spectrum of values and measuring the proportion of those values that fall within that range.
In order to find the best set of attributes from the available form factor, feature selection is critical in data preparation. Feature extraction reduces the variance of the training phase, eliminating the issue of overfitting. Additionally, the computation time of the training phase can be reduced. We use Correlation-based Feature Selection in this analysis (CFS).

RESULTS AND DISCUSSION

The suggested solution's output in order to improve accuracy, training sample test, and scalability study is the subject of the experiments conducted. One among the extensively utilized machine learning techniques is random forests. They are so common because they have strong predictive efficiency, low overfitting, and are simple to interpret [11]. The reason that it is easy to measure the calculated value on the tree decision leads to its generalizability. In other words, calculating what each variable contributes to the determination is easy.

Step 1: It randomly selects data specimens from a dataset.
Step 2: It then creates a decision tree for each specimen and evaluates all of the decision tree's expected outputs.
Step 3: Using voting, it selects the prediction model with the highest number of votes.

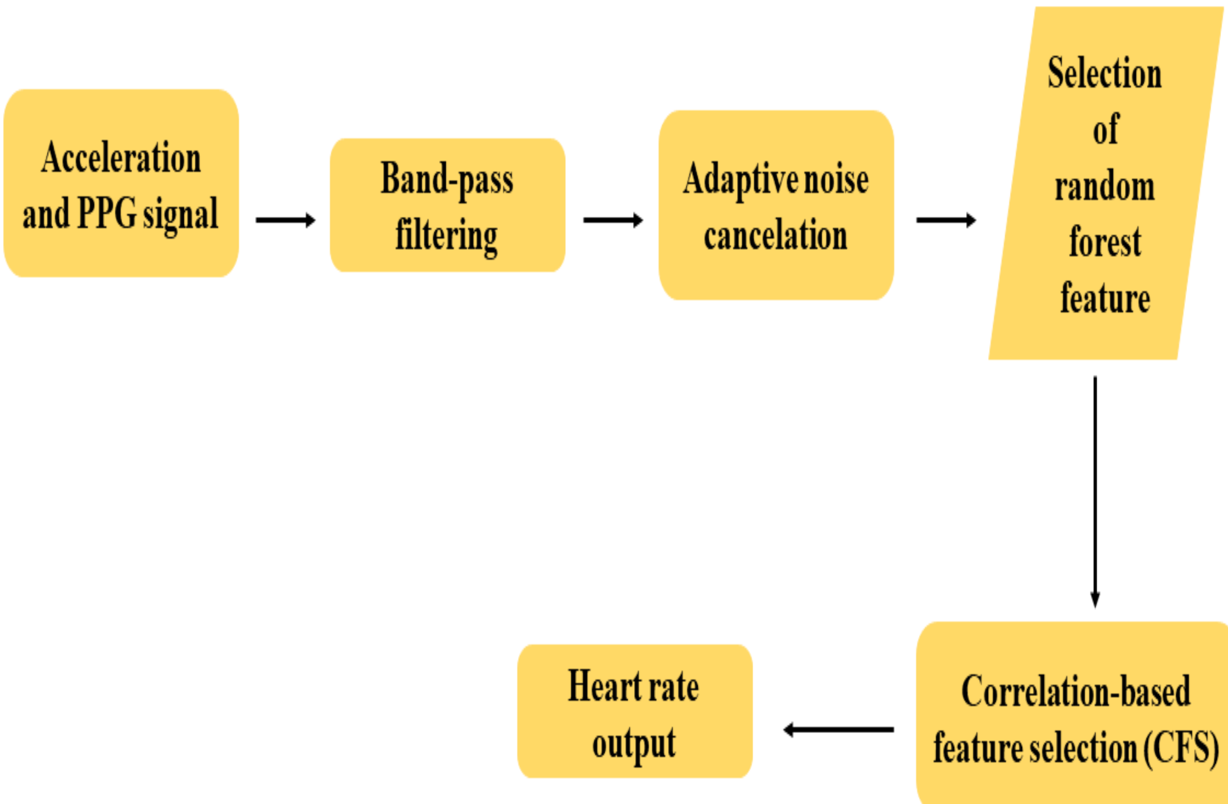

Figure 2. The flow-diagram of our proposed HR assessment method

Figure 2 depicts the flow-diagram of the proposed HR assessment method. The signals in one time window are rated as High, indicating that the spectrum peaks correlated by HR will only become dominant when ANC and CFS are used together; in other terms, a more reliable HR estimate can be achieved by using CFS further. In another way, MA in the PPG sign that have been denoised through Adaptive noise cancellation. All pure signals must be band-pass processed from 0.4 Hz to 5 Hz, which would be the operating frequency, until our suggested HR estimation approach could begin. The goal of adaptive noise cancellation (ANC) is to acquire the PPG signal that has been denoised. By measuring the standard of the triaxial vector at each sample site, we condensed the tri-acceleration impulses into one stream of sign a for simplicity. Random forest, a classification composed of a set of tree-structured classifiers, is a widely used standard to determine whether MA in the denoised PPG sign is Stronger or Not Strong. Random forest was chosen as the classification model in our studies because it can outperform many tree-based methods.
('id' , 0.057856905760074503)
('fault' , 0.10670347383529738)
('priority' , 0.411311584721126)

('frequency' , 0.2965677649581345)
('status' , 0.0)
('time-second' , 0.12756027072536755)

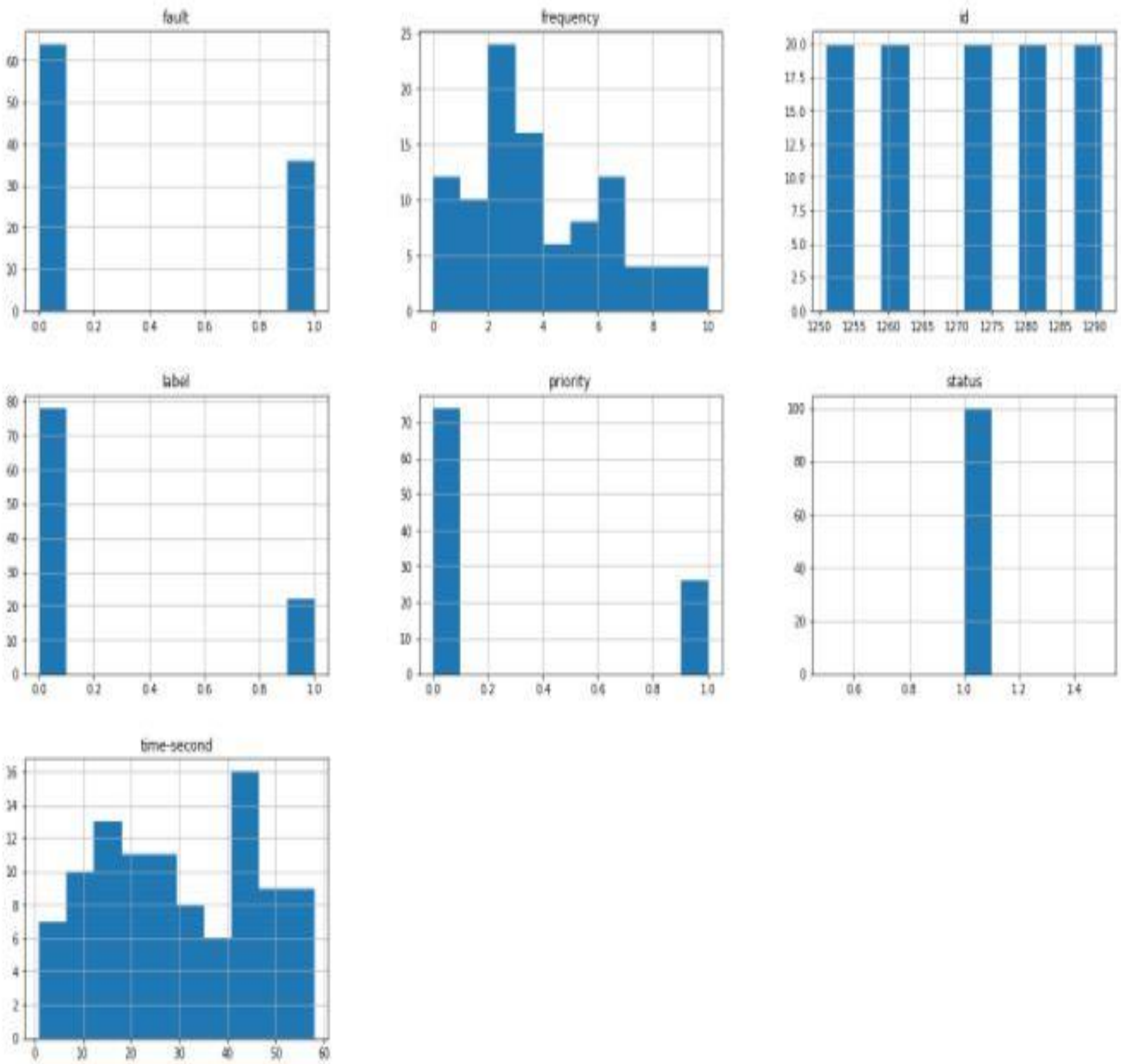

Figure 3: Histogram displays

A histogram is a graphical representation of data that uses bars of varying heights. Each bar in a histogram divides numbers into ranges. More information falls within the range as the bars get taller. The structure and distribution of continuous dataset are represented by a histogram. A predictor variable and several independent variables are used in most multivariate analyses. The majority of univariate techniques focus on explanation, while statistical models focus on testing hypotheses and interpretation. The histogram displays are shown in Figure 3. Here we performed the Histogram and Multivariate analysis (Figure 4) on following parameters, Fault, Priority, Time, Frequency and Status.

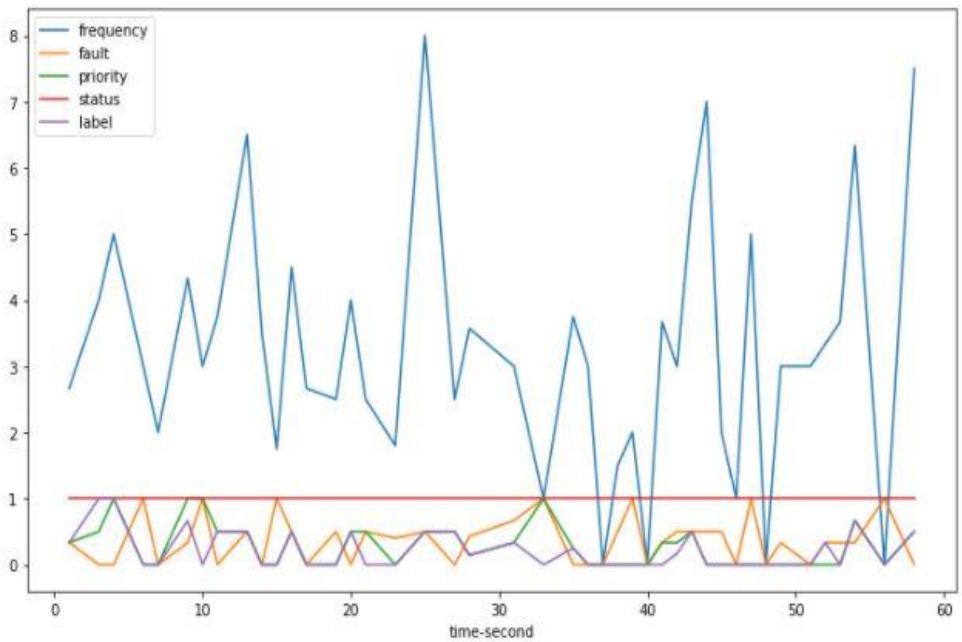

Figure 4: Multivariate analysis

Our suggested approach's evaluation results on measurements of HR traces were mapped, and this has been comparison to the ground reality, which was reported concurrently from ECG. The approximate HR ranges of our suggested HR assessment method was very near to the ground reality, showing a higher accuracy of our efficiency once again, as shown in the figure 5.

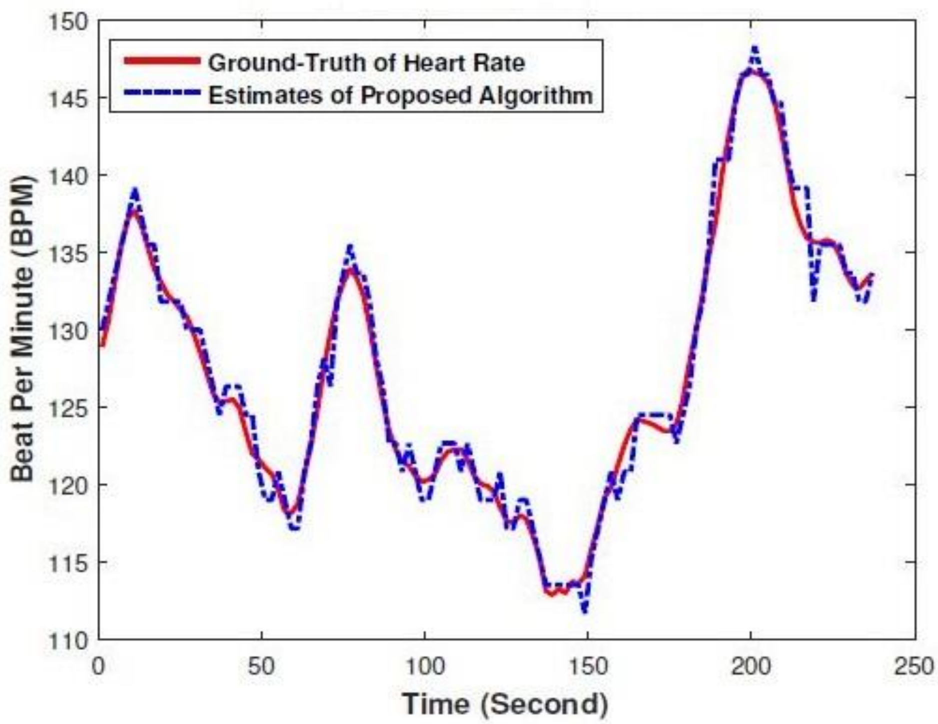

Figure 5. Estimation results on recordings

CONCLUSIONS

In our paper, a vigorous HR prediction model utilizing random forest is suggested for fitness monitoring through wearable devices like smarter watch and smart wristbands, focused on the PPG signal polluted by extreme motion artefacts. The projected method contains the phase of motion artifacts elimination such as adaptive noise termination, random forest-based feature selection, singular spectral study, Correlation-based Feature Selection. It can probably eliminate MA with lower computational complexity and pinpoint the spectral peak contributing to HR with improved sturdiness and generalization, resulting in higher HR approximation correctness and efficiency, suggesting its possible usage in wearable devices for healthcare applications and fitness tracking.

REFERENCES

1. Arunkumar, K. R., & Bhaskar, M. (2019). Heart rate estimation from photoplethysmography signal for wearable health monitoring devices. Biomedical Signal Processing and Control, 50, 1-9.
2. Chaudhuri, T., Zhai, D., Soh, Y. C., Li, H., & Xie, L. (2018). Random forest based thermal comfort prediction from gender-specific physiological parameters using wearable sensing technology. Energy and Buildings, 166, 391-406.
3. Ismail, S., Akram, U., & Siddiqi, I. (2021). Heart rate tracking in photoplethysmography signals affected by motion artifacts: a review. EURASIP Journal on Advances in Signal Processing, 2021(1), 1-27.
4. Li, Y., Yan, C., Liu, W., & Li, M. (2018). A principle component analysis-based random forest with the potential nearest neighbor method for automobile insurance fraud identification. Applied Soft Computing, 70, 1000-1009.
5. Roy, M. S., Gupta, R., Chandra, J. K., Sharma, K. D., & Talukdar, A. (2018). Improving photoplethysmographic measurements under motion artifacts using artificial neural network for personal healthcare. IEEE Transactions on Instrumentation and Measurement, 67(12), 2820-2829.
6. Singh, K., Guntuku, S. C., Thakur, A., & Hota, C. (2014). Big data analytics framework for peer-to-peer botnet detection using random forests. Information Sciences, 278, 488-497.
7. Tamura, T. (2019). Current progress of photoplethysmography and SPO 2 for health monitoring. Biomedical engineering letters, 9(1), 21-36.
8. Vergutz, A., Noubir, G., & Nogueira, M. (2020). Reliability for Smart Healthcare: A Network Slicing Perspective. IEEE Network, 34(4), 91-97.
9. Wager, S., & Athey, S. (2018). Estimation and inference of heterogeneous treatment effects using random forests. Journal of the American Statistical Association, 113(523), 1228-1242.
10. Xu, K., Jiang, X., & Chen, W. (2019). Photoplethysmography motion artifacts removal based on signal-noise interaction modeling utilizing envelope filtering and time-delay neural network. IEEE Sensors Journal, 20(7), 3732-3744.
11. Zailan, K. A. B. M., Hasan, M. H. B., & Witjaksono, G. (2020). Comparative Analysis Of Machine Learning Algorithms For Optimizing Variable Step-Size Least Mean Square In Motion Artifact Reduction. Compusoft: An International Journal Of Advanced Computer Technology, 9(3).