



Research article

An ECG data sampling method for home-use IoT ECG monitor system optimization based on brick-up metaheuristic algorithm

Qun Song^{1,*}, Tengyue Li², Simon Fong² and Feng Wu³

¹ College of Artificial Intelligence, Chongqing Technology and Business University, Chongqing, China

² Department of Computer and Information Science, University of Macao, Macao SAR, China

³ Zhuhai Institute of Advanced Technology (ZIAT), Chinese Academy of Science, Zhuhai, China

* **Correspondence:** Email: yb57459@umac.mo; Tel: +(86)18645618608.

Abstract: With the rise in the popularity of Internet of Things (IoT) in-home health monitoring, the demand of data processing and analysis increases at the server. This is especially true for ECG data which has to be collected and analyzed continuously in real time. The data transmission and storage capacity of a simple home-use IoT system is often limited. In order to provide a responsive and reasonably high-resolution analysis over the data, the ECG recorder sampling rate must be tuned to an acceptable level such as 50Hz (compared to between 100Hz and 500Hz in lab), a huge amount of time series are to be gathered and dealt with. Therefore, a suitable sampling method that helps shorten the ECG data transformation time and uploading time is very important for cost saving.. In this paper, how to down sample the ECG data is investigated; instead of traditional data sampling methods, the use of a novel Brick-up Metaheuristic Optimization Algorithm (BMOA) that automatically optimizes the sampling of ECG data is proposed. By its adaptive design in choosing the most appropriate components, BMOA can build in real-time a best metaheuristic optimization algorithm for each device user assuming no two ECG data series are exactly identical. This dynamic pre-processing approach ensures each time the most optimal part of the ECG data series is harvested for health analysis from the raw data, in different scenarios from different users. In this study various application scenarios using real ECG datasets are simulated. The experimentation is tested with one of the most commonly used ECG classification methods, Long Short-Term Memory Network. The result shows the ECG data sampling by BMOA is indeed adaptive, the classification efficiency is improved, and the data storage requirement is reduced.

Keywords: ECG data sampling; in-home IoT monitoring system; brick-up metaheuristic optimization algorithm; long short-term memory networks

1. Introduction

According to the World Health Organization, cardiovascular diseases (CVDs) are the number one cause of death globally. Of these deaths 85% are due to heart attack and stroke. The prevalence of CVD in China has been increasing continuously since 2006. Of the approximately 290 million patients with CVD, 13 million, 11 million, 5 million, 4.5 million, 2.5 million, 2 million, and 245 million have stroke, coronary heart disease (CHD), pulmonary heart disease, heart failure (HF), rheumatic heart disease, congenital heart disease, and hypertension, respectively [1]. However, these heart disease are very difficult to be noticed in the early stages. To diagnose heart diseases, abnormal signals are needed to be distinguished from all the recorded signals. This requires some long term monitoring which implies a large volume of signals would be collected and analyzed in a detection system [2].

By the growing popularity of in-home IoT health monitoring system, patients who suffer from heart diseases could have their heart conditions monitored 24 x 7 at home [3]. In-home electrocardiogram (ECG) monitoring system with performance on par with that laboratory grade equipment becomes more affordable than ever. Being a home-use health monitoring system, the processing power in hardware is still limited. Firstly, a home-use ECG monitoring system would have relatively smaller data storage space, compared to a data center at hospital grade. Secondly, most of the in-home ECG monitoring system provide simple detection functions which will sound off an alert when the ECG pattern is detected abnormal. Should the user require a referential consultation, they still have to go an expert. Some prior research seem to have solved these problems, contributing to the upgrades of the in-home systems. By adapting IoT cloud network technology, some new developed in-home ECG monitoring system can upload the data streams to a remote cloud server [4]. Experts then can conduct real time diagnosis during a telemedicine session. Nowadays, as machine learning developed quickly, automated analysis is made possible over ECG data using machines without human intervention. By introducing AI into ECG monitoring, workloads of basic screening are delegated to AI. So human experts only handle cases that are too sophisticated that demand for in-depth human expertise. Research results show that the machine learning methods that classify ECG data can be done at a very high accuracy rate [5], and some well developed algorithm has already been use]. Some well-developed algorithms have already been implemented in the industry [6].

Nowadays the study of ECG and machine learning have advanced to maturity. ECG data sampling, in contrast, has been a standard technique for a long time. By down-sampling raw ECG data, the sever can save storage resource and speed up data transformation because it deals with less data. Normally down-sampling ECG data works by filtering data, and/or transformation them to other formats. For example, down-sampling wavelet transform [7] or other deceptively down-sampling methods [8]. have been widely used. However, after down-sampling with these methods, the transformed data will be very different from the original ECG data format which is a time-series. Having a very different data format means the same machine at which a trained AI model is programmed to detect abnormality will not be compatible with the original ECG data anymore. In this study, a new method for ECG data sampling is proposed. This new method is aimed at not only reduce the data size, improve the system decision accuracy, but also preserving the format of the ECG data, allowing them readable by human experts for easy referencing them over other ECG data.

The proposed study in this paper is based based on a brick-up metaheuristic optimization system (BMOS). BMOS is first published in 2016 [9], and it has been perfected gradually in the following

years. This system can generate a specific metaheuristic optimization algorithm for a certain situation. As in-home ECG monitoring systems are used by many users of different demographics, the use cases can be very different from one another. The data sampling methods are required to have a strong adaptability to all the situations. BMOS is an excellent choice for solving this problem. In this paper, we discussed about customizing an IoT network structure, for pairing with this ECG data sampling method using BMOS. The overall system improves the ECG monitoring efficiency and classification accuracy.

In Section 2, the ECG data is first discussed, then the brick-up metaheuristic optimization system (BMOS) is introduced in Section 2.2. The proposed ECG data sampling method by BMOS is discussed in detail in Section 3. Then this method is simulated and tested by experiment in Section 4. The simulation experiment is based on the long short-term memory network classification, which introduced in detail in Section 4.1. The experiment result is summarized in Section 4.2. At last, the conclusion of this study is drawn in Section 5, and the further works are discussed in Section 6.

2. Research background

2.1. ECG data down-sampling analysis

An ECG signal is generated by electrodes placed on the skin. The electrodes detect the electrical activity of the heart [10]. As shown in Figure 1, the ECG data is a periodic data, for these data are the small electrical changes recorded during each cardiac cycle (heartbeat). The data cycle repeats with the heartbeat. A common ECG signal image is a graph of the measured voltage levels versus time[11].

The P wave represents atrial depolarization. The QRS complex represents ventricular depolarization. The T wave represents ventricular repolarization.

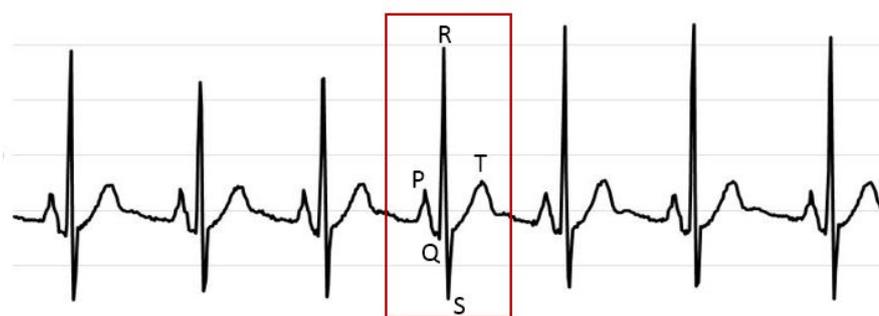


Figure 1. Part of a normal ECG signal.

This is how an in-home IoT ECG monitoring system works: collect ECG data, upload to the server, machine learning model perform classification (diagnose), and feedback the result. The process runs in real-time continuously. Traditional sampling method is conducted only before the classification. Normally, sampling ECG data is carried out by typical time series transformation methods, such as Discrete Fourier Transform (DFT) [12], Discrete Wavelet Transform (DWT) [13] and Singular Value Decomposition (SVD) [14]. However after applying these data sampling methods to reduce ECG data, the data would be transformed into frequency domain, no longer readable to human experts. So the server still need to store all the original ECG data for human interpretation. Different from other time-series

data, the ECG have its own feature that can be considered for down-sampling.

Figure 1 is part of an ECG report from a healthy volunteer. Experts diagnose heart disease by observing the pattern of ECG signals. Figure 2 is a full-length abnormal ECG report. An abnormal ECG signal does not show the abnormality in every cardiac cycle. For example, if the abnormal signal in Figure 2 is down-sampled to the blue part, the patient would be misdiagnosed. However, if the red part in Figure 2 is taken as the sampling result, an expert can easily identify the abnormality. With this feature, the IoT ECG monitoring system can keep only the salient part of the data on the server instead of the whole length of ECG time-series data.

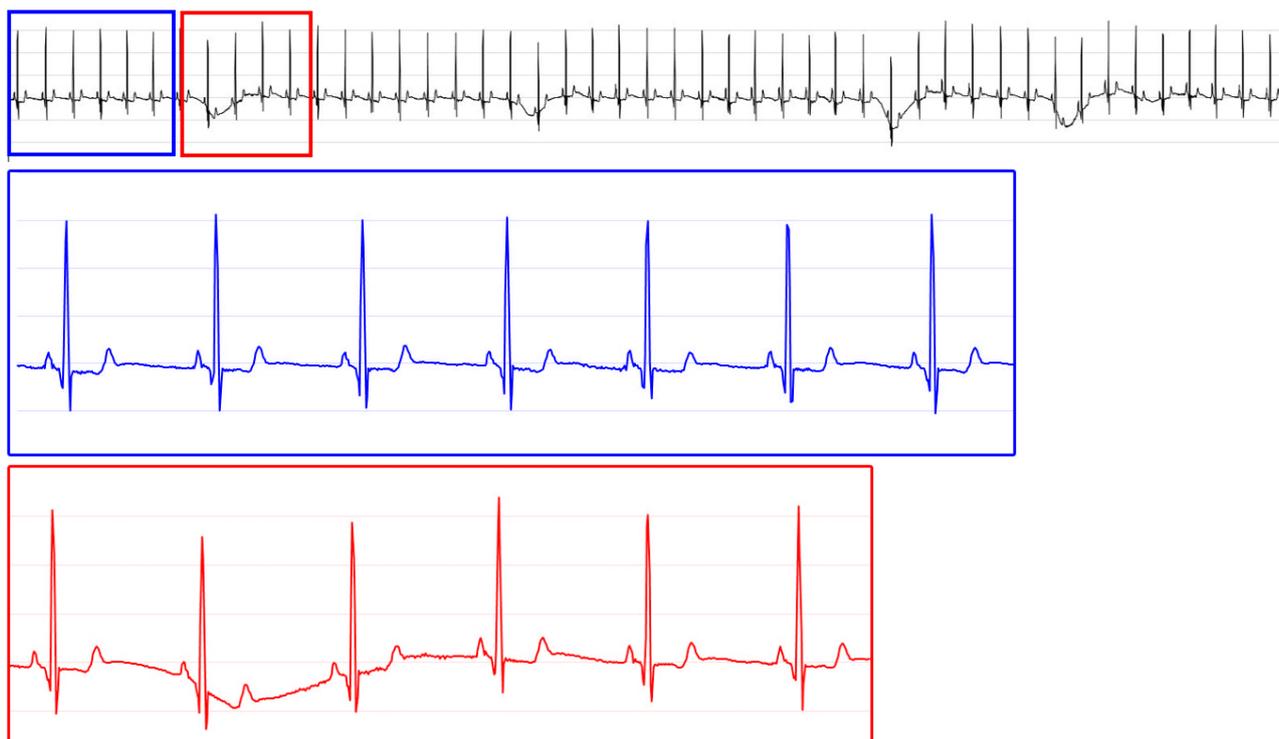


Figure 2. An abnormal ECG signal and two sampling examples.

An ideal sampling method for the home-use IoT ECG monitoring system should have the following advantages.

- Be efficient with different user case, have the ability to adapt to various data.
- Can detect the abnormal part from the original ECG data.
- Keep the original formation, maintain the readability for the experts.
- Improve the classification efficiency for classification algorithms.

2.2. The brick-up metaheuristic optimization algorithm

The task of ECG data down-sampling can be implemented as a search problem, which selectively looks for the abnormal part from the original ECG signal data. Computationally it can be treated as a optimization problem which picks a segment of ECG time series with the most significant abnormality.

As a well developed algorithm family, the metaheuristic optimization algorithm could be an excellent solution for this problem. There are a comprehensive collection of algorithms of different strengths.

Metaheuristic algorithms received so much attention in the computer science and mathematical optimization in recent years. Metaheuristic algorithms are also known as high-level procedure by the prefix ‘meta-‘ that guides how an optimization which is basically a search function runs in a search space. It is famous for its low required capacity for computation and is also able to solve NP-hard problems that generates a near-perfect result at the end ,without exhaustive executions.

Metaheuristic algorithms operate in stochastic manner, which means its executions are all based on random variables, randomly selected candidate data and random search over solution space. Sometimes it may only find a local optimum instead of the global optimum. Some improved metaheuristic algorithms can overcome the local optima in order to find a global optima solution by searching over a large area or feasible solutions. The upper layer of logics called meta is to guide the search out of local optima for chances of finding better solutions elsewhere in the solution space.

The metaheuristic algorithms have similar structure, and the core algorithm framework are not quite different [15]. Some algorithms improved in certain parts of the search, and they almost have the same structure as the classic ones except some modification on the search rules and movement logics. Based on these rules and movement logics, we can split the algorithms into several parts, each part is into some interchangeable function which is called “brick” in BMOS. We also need to make sure that each part has the same input and output variables. The combination of each part of the algorithm may cooperatively yield a better chance to form a powerful metaheuristic algorithm for solving some optimization problems.

This algorithm is proposed to optimize the IoT system by selecting the optimal segment of ECG time series. Based on the IoT network structure, the proposed algorithm should be able to adapt to different data collected from the user ends. As no single metaheuristic algorithm can be guaranteed to be the best for all data [16]. The brick-up metaheuristic algorithm would be an adaptive and excellent solution.

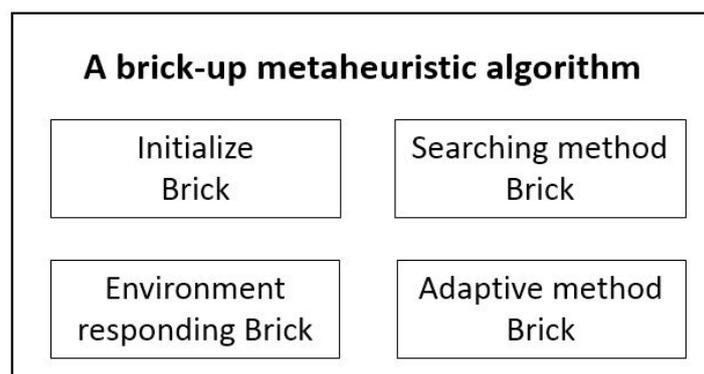


Figure 3. The bricks of a brick-up metaheuristic algorithm.

Brick-Up metaheuristic optimization system (BMOS) is designed to split the existing metaheuristic algorithms into several parts and recombining them into a new brick-up metaheuristic algorithm. It offers the flexibility of using brick-up methods to provide users to choose a combination of fundamental parts of metaheuristic algorithms which has already been proposed and proved with good performance.

The algorithms can be rebuilt by a brick-up method is of the popular movement functions in a search algorithm. For example, there are similar movement functions in genetic algorithm [17], bat algorithm [18], ant algorithm [19], wolf search algorithm [20], hybrid PSO [21] and so on. Some of these algorithms have been used in industry and finance to solve real-life optimization problems.

A final brick-up metaheuristic optimization algorithm (BMOA) is generated by the brick-up metaheuristic optimization system (BMOS), which is an expert system that offers advice on how to choose the best composite metaheuristic algorithm from many alternatives. In BMOS, the metaheuristic is broken down to several bricks which are components to build an algorithm: initialized brick, searching method brick, environment responding brick and the adaptive method brick based on the function, as shown in Figure 3. The system maintain 4 brick pools, each contains bricks that are collected from published metaheuristic algorithms. These 4 brick pools are the initialized brick pool, searching method brick pool, environment responding brick pool and the adaptive method brick pool. BMOS picks up the bricks in sequence given an optimization problem from the user case. The procedure of BMOS is shown in Figure 4.

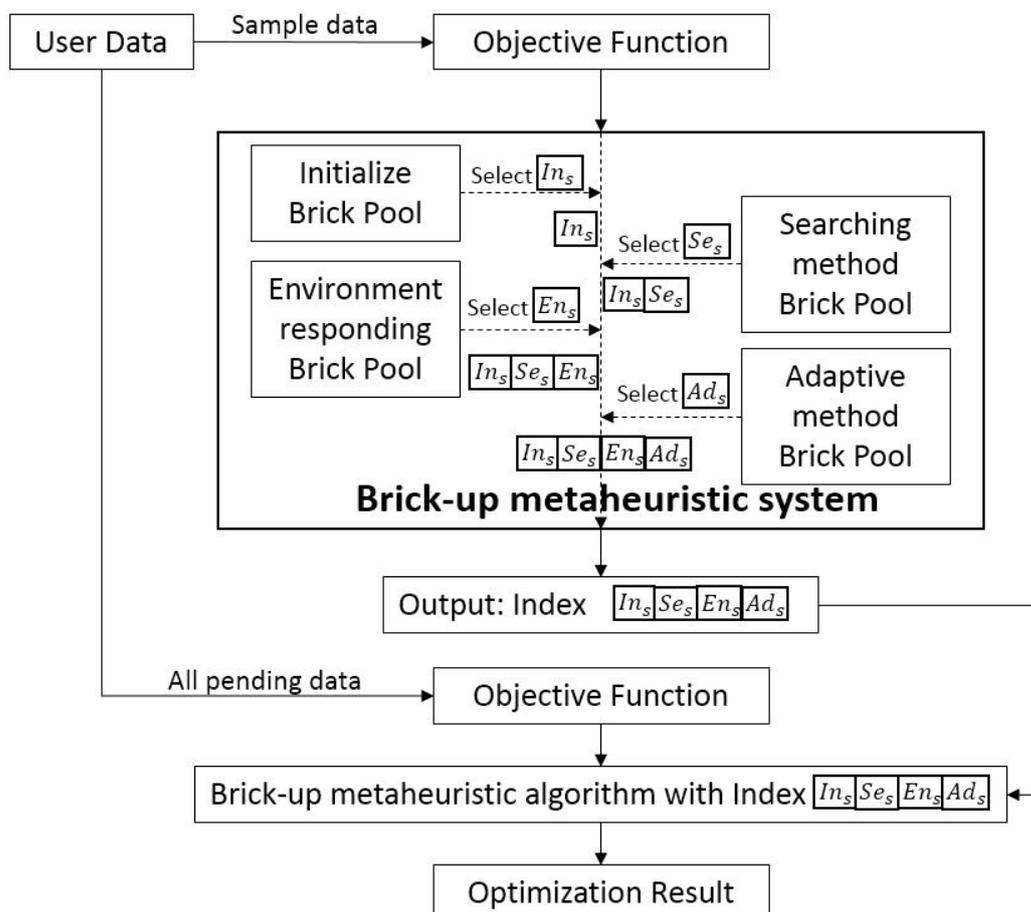


Figure 4. The flow chart of selecting brick procedure for brick-up metaheuristic optimization system.

Earlier on, a naïve version of BMOS was designed to run by Monte Carol decision model. It takes a very long time to go through thousands of simulation runs for fetching a suitable set of bricks. This

brick-up metaheuristic system is therefore improved in mind by enhancing the efficiency in fast picking the appropriate bricks using Analytic Hierarchy Process (AHP) decision model. Using BMOS, the selection of an optimization algorithm is fast, which in turn is used for extracting the optimal segment of ECG data. Collectively, BMOS maximizes the performance of an IoT health monitoring system. By processing a simplified version of ECG sample data at the server using a well-selected combination of optimization algorithms from resource pools of bricks, the system can advise each user in an efficient manner. For each user, they do not need to maintain or run the algorithm at the user end. Only the selected brick index should be stored, which is generated from the previous BMOS run which represents the best combination of optimization algorithm found for this particular user. Users can reuse the stored index which locates a suitable algorithm next time for its optimizing his ECG data.

Using this BMOS, a large workload of finding and testing suitable optimization algorithms can be reduced. However, it is important to formulate an appropriate objective function for BMOS to tackle the ECG sampling problem.

3. Proposed ECG data sampling method and the in-home IoT ECG monitor system optimization

As discussed earlier, the main purposes of this ECG data sampling are first, save the storage cost so the data transfer will become more efficient; second, improve the classification precision rate for the classification algorithm; third, keep the readability feature from the original ECG signal. To achieve these goals, instead of using time-frequency transform method such as the widely used time-series sample algorithms, here a more directly cutting method is explored.

In Figure 2, the red box contains the abnormal signal that needs to be kept. Shown as Eq 3.1, the *samplefunc()* should let $ECGdata_{sampled}$ contain as much abnormal signal as possible.

$$ECGdata_{sampled} = samplefunc(ECGdata_{raw}) \quad (3.1)$$

By reviewing the ECG data analysis research, the statistical analysis can be applied to mark the outliers of the ECG signal [22]. As marked in Figure 1, The P wave represents atrial depolarization, the QRS complex represents ventricular depolarization, the T wave represents ventricular repolarization. The QT interval is measured from the beginning of the QRS complex to the end of the T wave. One widely used statistical way to filter the outlier signal is based on the average of QT interval data. $aver_{3cycl}$ is calculated by averaging three consecutive heartbeats cycles. Compared with the sample average QT interval data, Q in Eq 3.2 is calculated as the outlier value. If $Q > 5\%$, the marked part will be treated as an outlier.

$$Q = \frac{|aver - aver_{3cycl}|}{aver} \quad (3.2)$$

With the knowledge of the outlier of ECG data can be marked with statistical analysis, the objective function of ECG data sampling is summarized. The purpose of this sampling method is to sample a long monitored ECG data. As for in-home monitoring, the behavior of the user may not be stable during the monitoring time. The user behavior can lead to rapid change of heart rate and ECG data. Only using the average of QT interval data, erroneous marking of the outlier may happen. Instead of using the average QT interval data as the outlier marker, the coefficient of variation of a certain length

ECG data is considered. According to the experiment from cardiologist, this certain length is set to 5 heartbeats per cycle in our study. ECG represents the raw ECG data, and the average time in judging the normality of $ECGs$ is 5 heartbeat cycles.

$$S = \left| \frac{\sqrt{\frac{\sum_{i=1}^n (ECG_i - \bar{ECG})^2}{n-1}}}{\bar{ECG}} - \frac{\sqrt{\frac{\sum_{j=1}^{n_s} (ECG_{s_j} - \bar{ECG}_s)^2}{n_s-1}}}{\bar{ECG}_s} \right| \quad (3.3)$$

S in Eq 3.3 is the coefficient of variation absolute difference of the raw ECG data and the 5 heartbeat cycle ECG fragment data. For a normal ECG data recorded by user in a stable environment, the ideal S would be close to 0. If the S of a fragment has a sudden increase, the current fragment has a higher chance to contain an abnormal signal than other fragments. Adapting this method, the ECG sample objective function should mark a suitable number of fragments that can offer the largest S .

Assume that the ideal number of fragments is N , $frag_i$ is the starting position of the i th fragment from the raw ECG data. The 5 heartbeat cycle length is set as $fragl$. ECG_i is the i th data on the raw ECG signal, and the ECG raw data length is $ECGL$. $h(x)$ summarized in Eq 3.4 is the final function for calculating the outlier value for sampled ECG which contain n fragments. The input x_i is an index of the fragments starting position $x_i = (frag_1, frag_2, \dots, frag_n)$.

$$h(x) = \sum_{w=1}^n \left| \frac{\sqrt{\frac{\sum_{i=1}^{ECGL} \left(ECG_i - \frac{\sum_{j=1}^{ECGL} ECG_j}{ECGL} \right)^2}{ECGL-1}}}{\frac{\sum_{j=1}^{ECGL} ECG_j}{ECGL}} - \frac{\sqrt{\frac{\sum_{k=x_w}^{x_w+fragl} \left(ECG_k - \frac{\sum_{h=x_w}^{x_w+fragl} ECG_h}{fragl} \right)^2}{fragl-1}}}{\frac{\sum_{h=x_w}^{x_w+fragl} ECG_h}{fragl}} \right| \quad (3.4)$$

If only the outlier value taken into consideration, the sample may be concentrated within certain time range. To make sure the sampled ECG data can be more representative, function $g(x)$ in Eq 3.5 is introduced. $\sum_{i=1}^{n-1} (sort(x_g)_{i+1} - sort(x_g)_i)$ calculate the total distance of each fragment starting position. The outcome value of $g(x)$ can represent the distribution divergence of the sample fragment combination.

$$g(x) = \sum_{g=1}^n \frac{\sum_{i=1}^{n-1} (sort(x_g)_{i+1} - sort(x_g)_i)}{ECGL} \quad (3.5)$$

Combining $h(x)$ and $g(x)$, the objective function for the optimization problem is $f(x) = h(x) + g(x)$. The proposed metaheuristic algorithm should find x , and we let $f(x)$ be the maximum value.

As shown in Figure 2, the ideal sample result should maintain the readability for data review. In this research the suggested length of fragment is 5 heart beat cycles. With the objective function defined, the steps of the proposed ECG data sample method could be summarized as follow.

- Define the number of the fragment windows n .
- Initialize the starting location for each fragment window l_1, l_2, \dots, l_n .
- Use the brick-up metaheuristic optimization algorithm to select the most fitting result for objective function $f(x)$, $x = (l_1, l_2, \dots, l_n)$. The pseudo-code for this brick-up metaheuristic algorithm is shown in Algorithm 1.

- Combine the selected fragments, use the combined ECG data as the sampled data.

Algorithm 1 Brick-up metaheuristic optimization algorithm for ECG data sampling

Input: $f(x)$ the objective function, n user defined fragment window number, initialized brick In_s , searching method brick Se_s , environment responding brick En_s , adaptive method brick Ad_s , raw ECG data ECG , max iteration number $maxiter$;

- 1: Initialize $x \leftarrow l_1, l_2, \dots, l_n$ based on In_s .
- 2: **repeat**
- 3: $n \leftarrow n + 1$
- 4: Update x based on Se_s , En_s and Ad_s , with objective function $f(x)$.
- 5: **until** $n > maxiter$
- 6: Update $ECG_{sampled}$ with x .

Output: $ECG_{sampled}$.

In Figure 5, an example of a 5 window sample from a raw ECG data is shown. By defining the number of the fragment windows, the user can decide the sampling rate. Less fragment windows will lead to a higher sampling rate. By adapting this ECG data sampling method, the system structure for in-home IoT ECG monitor system can be optimized as those block diagrams shown in Figures 6 and 7.

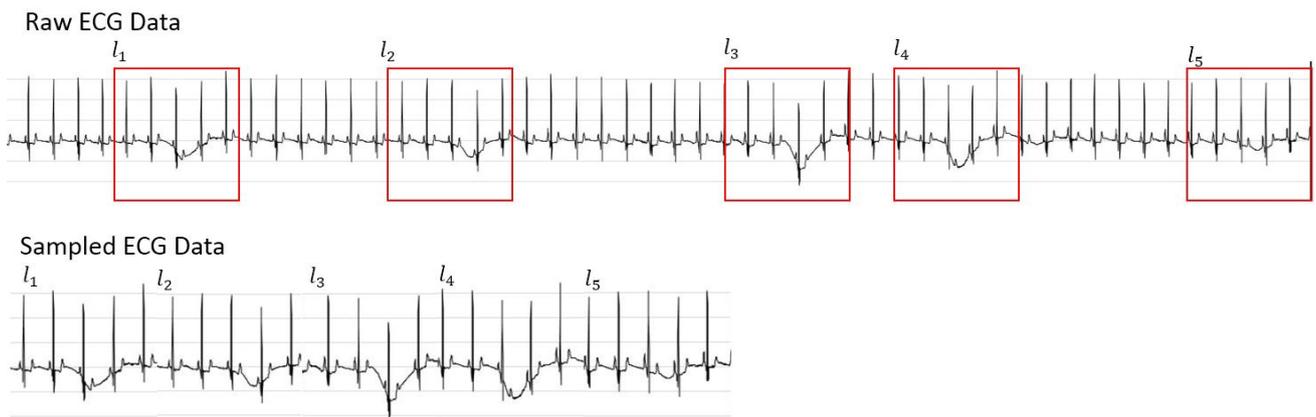


Figure 5. An example for ECG data sampling.

Figure 6 is the boot loader phase for the proposed in-home IoT ECG monitoring system. In this stage, the in-home monitor transfer collected raw data to the server. The server runs the brick-up metaheuristic system and selected the suitable bricks for certain user and sends the selected bricks back to the user database. After the booting stage, the user database will maintain the selected bricks for the customized BMOA and the raw user data (this raw data is optional for user). The server will maintain the received raw user data in this stage.

Figure 7 is the run phase for proposed in-home IoT ECG monitoring system. In this phase, after the monitor collected the user ECG data, the local computer will sample the data with selected brick-up metaheuristic algorithm, and transfer the sampled user data to server. User can choose whether to store the user data or no. The user data is decided by the user, which could be raw ECG data or sampled

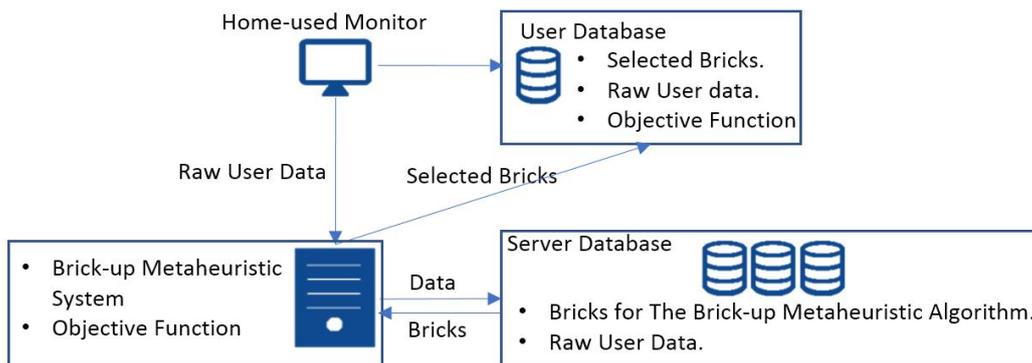


Figure 6. The boot loader phase for in-home IoT ECG monitor system.

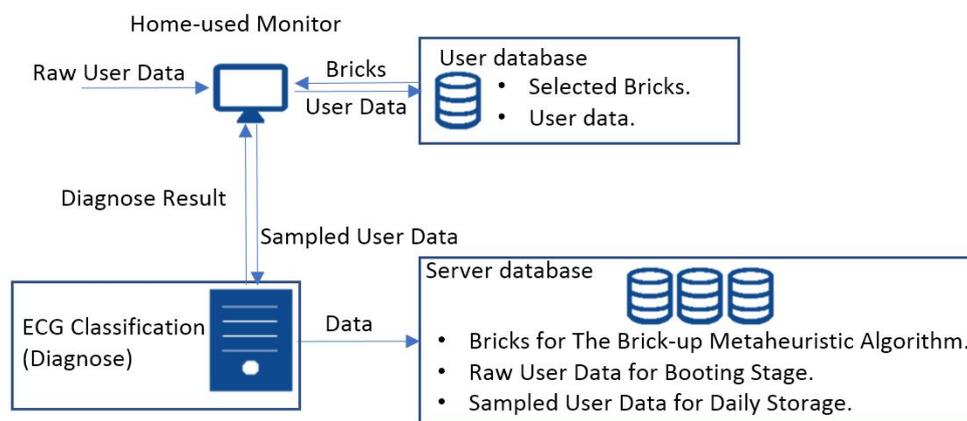


Figure 7. The run phase stage for in-home IoT ECG monitor system.

ECG data. Server will run the ECG classification algorithm to return the diagnose result. The server database will store the sampled user data for expert or user's review.

4. Simulation experiment and result

4.1. Simulation experiment

To verify the efficiency of the introduced ECG sample method, simulation experiment is built. To test if the sample method can improve the classification efficiency and accuracy, we first run the classification network with raw ECG data, then compare with the classification result with sampled ECG data. PSO is also adopted into the introduced ECG sampling method as a control experiment of the BMOA.

This experiment is run with database ECG200 [23]. The ECG200 dataset contains raw ECG data with 6000 data points in 2 classes. The two classes are a normal heartbeat and a Myocardial Infarction. To build a simulation experiment, 4000 ECG series are randomly selected from the dataset as our raw ECG data. In each training and testing experiment, 400 from the selected 4000 series are randomly detected as the test data, other 3600 are used as the training data. The most widely used classification algorithm for ECG data is the deep learning algorithm family [24]. In this simulation experiment we

use the long short-term memory networks.

The long short-term memory networks (LSTMs) is one of the most widely used neural networks for ECG classification, published in 1997 by Hochreiter and Schmidhuber [25]. It was built to solve the long-term dependencies problem, and it has been proved a very efficiency solution for this problem. This feature made the LSTMs a popular tool for ECG data anomaly detection [26].

The LSTMs is developed based on the structure of RNNs. It shares the same chain structure with RNNs. The biggest difference of these two networks is the cell state. For RNNs, the cell state for each module is simple linear updates. For LSTMs, the cell state controlled by gate units. Each gate unit contains a sigmoid network and a pointwise operation [27].

The forget gate layer can decides whether the information need to remove from the cell state or not. The input gate layer decide which values need to be updated, then the tanh layer generate the candidate vector which can be updated to the cell state. The output will be a filtered cell state of the LSTMs.

With the introduced feature, the LSTMs have been proved as a particularly useful network for learning sequences containing longer term memory. This made the LSTMs a unique advantage for ECG signal, so that the ECG signal can be directly fed into the network without any elaborate preprocessing as required by other techniques [28]. The research also proved that LSTMs shows a higher potential to produce a better result than other traditional deep learning networks with ECG data [29].

In this experiment, the LSTMs is used as the classification network in the system. If the sample method can improve the accuracy and efficiency for the LSTMS. The sample method can be verified as a success. The following experiment is simulated on Dell standalone workstation with processor Intel(R) Xeon(R) CPU E5-2670 v2 @ 2.50 Ghz, GPU NVIDIA GeForce GTX 980 Ti, installed memory 128 GB and 64-bit Windows 7 operation system. The experiment is implemented with Matlab 2020a, CUDA 10.0.

As the used database that contains the raw ECG data with 6000 data points. The fragment window is set with 500 data points, which contain around 5 heart beat cycle. The fragment number n is set to 8, 6, 4. When $n = 8$, the sampled ECG data will contain 4000 data points, the sampling rate is 66.67%. When $n = 6$, the sampled ECG data will contain 3000 data points, the sampling rate is 50%. When $n = 4$, the sampled ECG data will contain 2000 data points, the sampling rate is 33.33%.

For each experiment with different sampling rate, the discussed ECG sample method is compared with PSO sampleing method. Also compared with similar sample function $direct_{cut}$ and $random_{cut}$. $direct_{cut}$ sampled the raw ECG data from the starting point to the user set length, $random_{cut}$ set the fragment window location randomly.

The discussed sample method is also compared with time-frequency domain transformation feature selection method. Fourier transforms is used to calculate the transform. Instantaneous frequency and spectral entropy for each ECG data is used as the feature for classification.

4.2. Experiment result

Table 1 shows the experiment result comparison. Raw Data column contain the training and testing phase accuracy, error, sensitivity, specificity, precision, false positive rate (FPR), F1 score, Matthews correlation coefficient (MCC) and Kappa value. By comparing these values from the Fourier transforms feature selection and the discussed BMOA-based ECG sampling method, the LSTMs classification algorithm offers a better result with brick-up metaheuristic algorithm base ECG sampling method. The accuracy increased with fragment window number n . However even when $n = 4$ (sample rate

33.33%), the brick-up metaheuristic algorithm base ECG sampling method still offer a better result than non-sampled data and the Fourier transforms feature selection.

Table 1. Experiment result for LSTMs ECG classification with raw ECG data, Fourier transforms feature selection and BMOA-based ECG data sampling method (n = 8, 6 and 4).

	Raw Data		Fourier trans.		Brick-up n = 8		Brick-up n = 6		Brick-up n = 4	
	Train	Test	Train	Test	Train	Test	Train	Test	Train	Test
Accuracy	0.7664	0.7525	0.8425	0.7825	0.8919	0.8600	0.8233	0.8575	0.8356	0.8075
Error	0.2336	0.2475	0.1575	0.2175	0.1081	0.1400	0.1767	0.1425	0.1644	0.1925
Sensitivity	0.6794	0.6250	0.7828	0.6900	0.9172	0.8700	0.8022	0.8550	0.8250	0.7900
Specificity	0.8533	0.8800	0.9022	0.8750	0.8667	0.8500	0.8444	0.8600	0.8461	0.8250
Precision	0.8225	0.8389	0.8890	0.8466	0.8731	0.8529	0.8376	0.8593	0.8428	0.8187
FPR	0.1467	0.1200	0.0978	0.1250	0.1333	0.1500	0.1556	0.1400	0.1539	0.1750
F1_score	0.7441	0.7163	0.8325	0.7603	0.8946	0.8614	0.8195	0.8571	0.8338	0.8041
MCC	0.5410	0.5223	0.6899	0.5749	0.7849	0.7201	0.6472	0.7150	0.6713	0.6154
Kappa	0.5328	0.5050	0.6850	0.5650	0.7839	0.7200	0.6467	0.7150	0.6711	0.6150

A two-class confusion matrix is summarized in Table 2. In Table 2, TP is the True Positive, FP is the False Positive, FN is False Negative and TN is True Negative.

Table 2. Two-class confusion matrix for comparison experiment.

		Raw ECG Data		Fourier trans.		Brick-up n = 8		Brick-up n = 6		Brick-up n = 4	
		TP	FP	TP	FP	TP	FP	TP	FP	TP	FP
Train	FN	1223	577	1409	391	1651	149	1444	356	1485	315
	TN	264	1536	176	1624	240	1560	280	1520	277	1523
Test	FN	125	75	138	62	174	26	171	29	158	42
	TN	24	176	25	175	30	170	28	172	35	165

The result from Tables 1 and 2 prove the efficiency of the proposed sampling method. To prove this efficiency is not only offered by the fragment cut-re-assemble, but also the brick-up metaheuristic algorithm another group of comparison experiment is setup. In Tables 3, 4 and 5, ECG data sampling method with BMOA is compared with the sampling method with *direct_{cut}*, *random_{cut}* and the PSO. The experiment result with brick-up metaheuristic algorithm always is the best compared with *direct_{cut}*, *random_{cut}* and the PSO.

The two-class confusion chart for test result with BMOA and PSO is summarized in Figure 8, and result with *direct_{cut}*, *random_{cut}* is in Figure 9.

5. Conclusions

In this paper, an ECG data sampling method is discussed to meet the increasing need of in-home IoT ECG monitoring system. The motivation of this sampling method is to reduce the data transfer and storage load for the IoT system. By designing the objective function Equations 3.5 and 3.4 for optimization algorithm, the proposed sampling method can be implemented with metaheuristic algorithms. By analysing the IoT system features, the brick-up metaheuristic system is adapted on the

Table 3. Experiment result with different sampling method with fragment windows.

	Brick-up n = 8		Direct n = 8		Random n = 8		PSO n = 8	
	Train	Test	Train	Test	Train	Test	Train	Test
Accuracy	0.8919	0.8600	0.6094	0.6500	0.5950	0.6125	0.7631	0.7450
Error	0.1081	0.1400	0.3906	0.3500	0.4050	0.3875	0.2369	0.2550
Sensitivity	0.9172	0.8700	0.4956	0.5250	0.2122	0.2500	0.7444	0.7550
Specificity	0.8667	0.8500	0.7233	0.7750	0.9778	0.9750	0.7817	0.7350
Precision	0.8731	0.8529	0.6417	0.7000	0.9052	0.9091	0.7732	0.7402
FPR	0.1333	0.1500	0.2767	0.2250	0.0222	0.0250	0.2183	0.2650
F1_score	0.8946	0.8614	0.5592	0.6000	0.3438	0.3922	0.7586	0.7475
MCC	0.7849	0.7201	0.2248	0.3098	0.2953	0.3267	0.5265	0.4901
Kappa	0.7839	0.7200	0.2189	0.3000	0.1900	0.2250	0.5261	0.4900

Table 4. Experiment result with different sampling method with 6 fragment windows.

	Brick-up n = 6		Direct n = 6		Random n = 6		PSO n = 6	
	Train	Test	Train	Test	Train	Test	Train	Test
Accuracy	0.8233	0.8575	0.5989	0.5875	0.6047	0.6125	0.8067	0.8350
Error	0.1767	0.1425	0.4011	0.4125	0.3953	0.3875	0.1933	0.1650
Sensitivity	0.8022	0.8550	0.2711	0.3100	0.2306	0.2400	0.7978	0.8350
Specificity	0.8444	0.8600	0.9267	0.8650	0.9789	0.9850	0.8156	0.8350
Precision	0.8376	0.8593	0.7871	0.6966	0.9161	0.9412	0.8122	0.8350
FPR	0.1556	0.1400	0.0733	0.1350	0.0211	0.0150	0.1844	0.1650
F1_score	0.8195	0.8571	0.4033	0.4291	0.3684	0.3825	0.8049	0.8350
MCC	0.6472	0.7150	0.2619	0.2104	0.3158	0.3373	0.6134	0.6700
Kappa	0.6467	0.7150	0.1978	0.1750	0.2094	0.2250	0.6133	0.6700

Table 5. Experiment result with different sampling method with 4 fragment windows.

	Brick-up n = 4		Direct n = 4		Random n = 4		PSO n = 4	
	Train	Test	Train	Test	Train	Test	Train	Test
Accuracy	0.8356	0.8075	0.5597	0.6025	0.6139	0.5825	0.7172	0.7100
Error	0.1644	0.1925	0.4403	0.3975	0.3861	0.4175	0.2828	0.2900
Sensitivity	0.8250	0.7900	0.5389	0.5950	0.2533	0.1900	0.8422	0.8200
Specificity	0.8461	0.8250	0.5806	0.6100	0.9744	0.9750	0.5922	0.6000
Precision	0.8428	0.8187	0.5623	0.6041	0.9084	0.8837	0.6738	0.6721
FPR	0.1539	0.1750	0.4194	0.3900	0.0256	0.0250	0.4078	0.4000
F1_score	0.8338	0.8041	0.5504	0.5995	0.3962	0.3128	0.7486	0.7387
MCC	0.6713	0.6154	0.1195	0.2050	0.3288	0.2663	0.4487	0.4305
Kappa	0.6711	0.6150	0.1194	0.2050	0.2278	0.1650	0.4344	0.4200

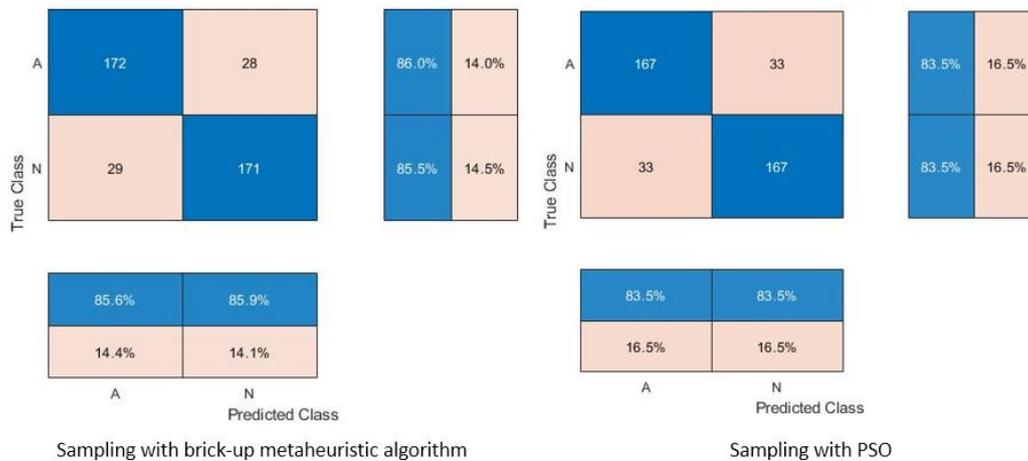


Figure 8. Two-class confusion matrix for brick-up method and PSO method with $n = 6$.

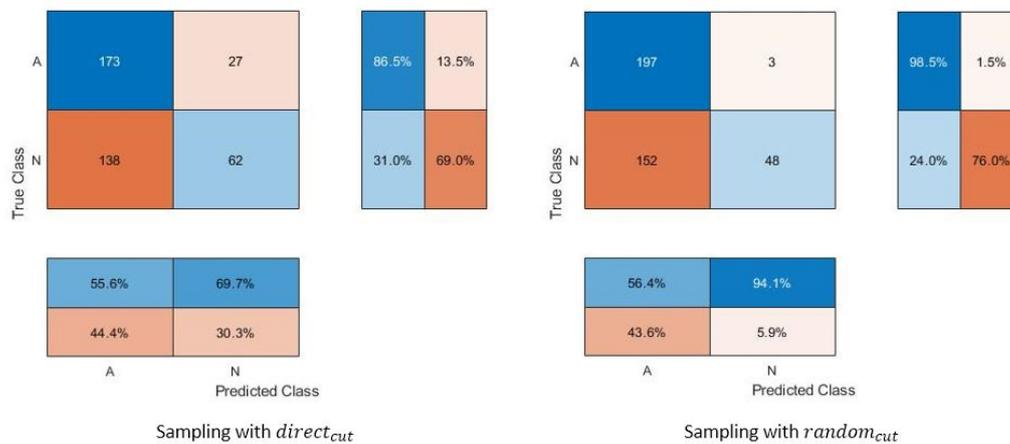


Figure 9. Two-class confusion matrix for $direct_{cut}$, $random_{cut}$ with $n = 6$.

server to generate brick-up metaheuristic algorithms for certain user. Based on the solution, the in-home IoT ECG monitoring system is optimized in Figures 6 and 7. The efficiency of the discussed sampling method and system is simulated with long short-term memory networks.

Table 1 shows that, using proposed method, the classification accuracy improves from 0.7525 to 0.8600 while the data size is reduced by 33.33%. With brick-up metaheuristic algorithm based ECG data sampling method when set the number of fragment window to 4, classification accuracy improved from 0.7525 to 0.8075. Data size of raw data is 203339KB, after sampling the data size becomes 67398KB. This result shows brick-up metaheuristic algorithm based ECG data sampling method can perfectly reduce the load for data transfer and storage, while improving the classification efficiency. This method didn't introduce time-frequency domain transformation or other data format transformation, the sampled data still maintain ECG fragments that can be read by experts. The discussed BMOA-based ECG data sampling method is an excellent solution to improve the efficiency for in-home IoT.

This result also proved the efficiency of brick-up metaheuristic system. In this case, the brick-up metaheuristic system sent the bricks to the user-end, each user only needs to maintain his own

brick-up metaheuristic algorithm. The user can get a suitable algorithm without any knowledge of the metaheuristic algorithms. The feature of brick-up metaheuristic algorithm can ensure the user obtain a candidate algorithm that produces accuracy results on par with the best algorithm available from the system. Result in Tables 3, 4 and 5 prove the efficiency of brick-up metaheuristic algorithm. Brick-up metaheuristic system can meet the distributed structure of the IoT system, it has a high potential to be used in other IoT based system that deals with time-series.

6. Further works

This paper introduced a brick-up metaheuristic algorithm based ECG data sampling method. To test this method, a very simple IoT system is simulated. With limited resource, this method is tested with a 4000 series dataset, and simulated a distributed system. Consider the commercial in-home IoT ECG monitoring system that will have a more complex structure and limitation, this BMOA-based ECG data sampling method and the proposed system structure may need to be further refined.

Firstly, to improve the performance of brick-up metaheuristic algorithm, more bricks are needed to be added into the brick-up metaheuristic system brick pool. The performance of suggested brick-up metaheuristic algorithm largely depends on the brick variety. However, having more bricks not only improve the efficiency but also cost a lot of computing resources, which may lead to certain computing overheads to the distributed IoT system. Further works for this problem should find a suitable way to add new and diverse bricks to the pool and remove bricks which have high similarity or lower efficiency to the others.

Secondly, as a simple simulation, in this experiment, the latency between the user end and the server is not considered. The run-time of the pre-classification data processing is not calculated. But in real-life scenarios, the user-end computer may have a very large disparity from the server. Taking this situation into consideration, whether putting the data sampling algorithm on the user end would be a problem. So the further work for this problem should be: acquiring a real in-home IoT ECG monitor system, testing the normally used user-end computer efficiency, recording the processing time, taking data transfer time into consideration and improving the in-home IoT ECG monitoring system based on the real use case.

Thirdly, after improve the system structure and sampling method, more classification method and feature extraction method should be taken into consideration. The LSTMs and Fourier transform based feature extraction used in this paper is a very basic solution. With the development of time series data research more and more efficient methods can be used for ECG classification. The matching learning algorithm for revised network is proved very efficiency for the highly varying time-series data, such as ECG [30]. For feature selection, the characteristic features of the heartwave signal comprises of the P-wave, QRS-complex, Twave, onset and offset of P-wave and T-wave, Discrete Wavelet Transformation coupled with adaptive parameters can extract more accurate features from the signal [31]. Other method such as adaptive heartrate-dependent heartwave-signal-based biometric identification can also provide excellent ECG single processing result [32]. With all these developments, the suggested ECG data sampling method should be tested with these new published methods to show its adaptability.

Another important further work is, as a suggested in-home IoT system, the user may not have the proper user guide as in a hospital. So the data may be collected under some extreme physical duress. This may lead to the collected ECG data containing extreme morphological variations. In the

literature, this problem can solve by a heartwave based authentication method with an ensemble of deep belief networks (DBNs). This method under different parameters can increase the reliability of feature extraction, can classify ECG signals from 50 bpm to 180 bpm [33]. For an in-home IoT ECG monitor system, this is a very important part to be added.

The introduced ECG data sampling method for in-home IoT IOT ECG monitoring system optimization based on brick-up metaheuristic algorithm is shown by experiments to be very efficient. New methods to reduce the ECG data still emerge recently. For example, researchers try to compress the ECG data with Douglas- Peucker algorithm and fractal interpolation [34]. By combining these new ideas into our proposed sampling method, the system could be further improved and complete.

Conflict of interest

We declare that we do not have any commercial or associative interest that represents a conflict of interest in connection with the work submitted.

Acknowledgment

The authors are thankfully acknowledging with gratitude for the supports of these grants: the 2020 Key Technology R&D Program of GuangDong Province, grant number ZH01110405180056PWC, Zhuhai Basic and Application Research Project, grant number ZH22017003200011PWC.

References

1. L.-Y. Ma, W.-W. Chen, R.-L. Gao, L.-S. Liu, M.-L. Zhu, Y.-J. Wang, et al., China cardiovascular diseases report 2018: an updated summary, *J. Geriatr. Cardiol.*, **17** (2020), 1–8.
2. M. Sanz, A. M. del Castillo, S. Jepsen, J. Gonzalez-Juanatey, F. D’Aiuto, P. Bouchard, et al., Periodontitis and cardiovascular diseases: Consensus report, *J. Clin. Periodontol.*, **47** (2020), 268–288.
3. S. Baumann, Evaluation of data usability generated by wearables & iot-enabled home use medical devices via telehealth to identify if blockchain can solve potential challenges, 2020.
4. M. H. Nornaim, N. A. Abdul-Kadir, F. K. Harun, M. A. A. Razak, A wireless ecg device with mobile applications for android, In *7th Int. Conf. Electr. Eng. Comput. Sci. Inf.*, pages 168–171. IEEE, 2020.
5. A. K. Sangaiah, M. Arumugam, G.-B. Bian, An intelligent learning approach for improving ecg signal classification and arrhythmia analysis, *Artif. Intell. Med.*, **103** (2020), 101788.
6. P. M. Rautaharju, S. H. Zhou, E. W. Hancock, B. M. Hor, D. Q. Feild, J. M. Lindauer, et al., Comparability of 12-lead ecgs derived from easi leads with standard 12-lead ecgs in the classification of acute myocardial ischemia and old myocardial infarction, *J. Electrocardiol.*, **35** (2002), 35–39.
7. Y. Zou, J. Han, X. Weng, X. Zeng, An ultra-low power qrs complex detection algorithm based on down-sampling wavelet transform, *IEEE Signal Process. Lett.*, **20** (2013), 515–518.
8. L. Mesin. Heartbeat monitoring from adaptively down-sampled electrocardiogram, *Comput. Biol. Med.*, **84** (2017), 217–225.

9. Q. Song, S. Fong, Brick-up metaheuristic algorithms, In *5th IIAI Int. Congress Adv. Appl. Inf.*, pages 583–587. IEEE, 2016.
10. L. S. Lilly, *Pathophysiology of heart disease: a collaborative project of medical students and faculty*, Lippincott Williams & Wilkins, 2012.
11. C. Zhang, Y. Chen, A. Yin, X. Wang, Anomaly detection in ecg based on trend symbolic aggregate approximation, *Math. Biosci. Eng.*, **16** (2019), 2154–2167.
12. S. Mitra, M. Mitra, B. B. Chaudhuri, Generation of digital time database from paper ecg records and fourier transform-based analysis for disease identification, *Comput. Biol. Med.*, **34** (2004), 551–560.
13. R. J. Martis, U. R. Acharya, L. C. Min, Ecg beat classification using pca, lda, ica and discrete wavelet transform, *Biomed. Signal Process. Control*, **8** (2013), 437–448.
14. J.-J. Wei, C.-J. Chang, N.-K. Chou, G.-J. Jan, Ecg data compression using truncated singular value decomposition, *IEEE Trans. Inf. Technol. Biomed.*, **5** (2001), 290–299.
15. S. Fong, X. Wang, Q. Xu, R. Wong, J. Fiaidhi, S. Mohammed, Recent advances in metaheuristic algorithms: Does the makara dragon exist?, *J. Supercomput.*, **72** (2016), 3764–3786.
16. W. Li, G.-G. Wang, A. H. Gandomi, A survey of learning-based intelligent optimization algorithms, *Arch. Comput. Method. E.*, (2021), pages 1–19, 2021.
17. S. Mirjalili, Genetic algorithm, In *Evolutionary algorithms and neural networks*, pages 43–55. Springer, 2019.
18. X.-S. Yang, X. He, Bat algorithm: literature review and applications, *Int. J. Bio-inspir. Com.*, **5** (2013), 141–149.
19. D. J. Li, C. Z. Qiang, Y. Z. Zhi, On the combination of genetic algorithm and ant algorithm, *J. Comput. Inf. Syst.*, **9** (2003), 10.
20. R. Tang, S. Fong, X.-S. Yang, S. Deb, Wolf search algorithm with ephemeral memory, In *7th Int. Conf. Digit. Inf. Management*, pages 165–172. IEEE, 2012.
21. K. Premalatha, A. Natarajan, Hybrid pso and ga for global maximization, *Int. J. Open Problems Compt. Math*, **2** (2009), 597–608.
22. B. Mendzelevski, C. S. Spencer, A. Freier, D. Camilleri, C. Graff, J. Täubel, Comparing the consistency of electrocardiogram interval measurements by resting ecg versus 12-lead holter, *Ann. Noninvas. Electro.*, page e12851, 2021.
23. R. T. Olszewski, *Generalized feature extraction for structural pattern recognition in time-series data*, Carnegie Mellon University, 2001.
24. E. K. Wang, L. Xi, R. P. Sun, F. Wang, L. Y. Pan, C. X. Cheng, et al., A new deep learning model for assisted diagnosis on electrocardiogram, *Math. Biosci. Eng.*, **16** (2019), 2481–2491.
25. S. Hochreiter, J. Schmidhuber, Long short-term memory, *Neural Comput.*, **9** (1997), 1735–1780.
26. P. Malhotra, L. Vig, G. Shroff, P. Agarwal, Long short term memory networks for anomaly detection in time series, In *Proceedings*, volume 89, pages 89–94. Presses universitaires de Louvain, 2015.

27. K. Greff, R. K. Srivastava, J. Koutník, B. R. Steunebrink, J. Schmidhuber, Lstm: A search space odyssey, *IEEE Trans. Neural Netw. Learn. Syst.*, **28** (2016), 2222–2232.
28. S. Chauhan, L. Vig, Anomaly detection in ecg time signals via deep long short-term memory networks, In *IEEE Int. Conf. Data Sci. Adv. Anal.*, pages 1–7. IEEE, 2015.
29. M. Liu, Y. Kim, Classification of heart diseases based on ecg signals using long short-term memory, In *40th Ann. Int. Conf. IEEE Eng. Med. Biol. Soc.*, pages 2707–2710. IEEE, 2018.
30. B. H. D. Koh, C. L. P. Lim, H. Rahimi, W. L. Woo, B. Gao. Deep temporal convolution network for time series classification, *Sensors*, **21** (2021), 603.
31. C. L. P. Lim, W. L. Woo, S. S. Dlay, Enhanced wavelet transformation for feature extraction in highly variated ecg signal, In *2nd IET Int. Conf. Intell. Signal Process*, pages 1–6. IET, 2015.
32. C. L. P. Lim, W. L. Woo, S. S. Dlay, B. Gao, Heartrate-dependent heartwave biometric identification with thresholding-based gmm–hmm methodology, *IEEE Trans. Ind. Inf.*, **15** (2018), 45–53.
33. C. L. P. Lim, W. L. Woo, S. S. Dlay, D. Wu, B. Gao, Deep multiview heartwave authentication, *IEEE Trans. Ind. Inf.*, **15** (2018), 777–786.
34. H. Guedri, A. Bajahzar, H. Belmabrouk, Ecg compression with douglas-peucker algorithm and fractal interpolation, *Math. Biosci. Eng.*, **18** (2021), 3502–3520.



AIMS Press

©2021 the Author(s), licensee AIMS Press. This is an open access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0>)