Journal Pre-proof

An intelligent healthcare system with data priority based on multi vital biosignals

Rongjun Xie, Shahriar Badsha, Ibrahim Khalil, Mohammed Atiquzzaman

PII: S0169-2607(19)31006-5
DOI: https://doi.org/10.1016/j.cmpb.2019.105126
Reference: COMM 105126

To appear in: *Computer Methods and Programs in Biomedicine*

Received date: 25 June 2019
Revised date: 14 September 2019
Accepted date: 8 October 2019

Please cite this article as: Rongjun Xie, Shahriar Badsha, Ibrahim Khalil, Mohammed Atiquzzaman, An intelligent healthcare system with data priority based on multi vital biosignals, *Computer Methods and Programs in Biomedicine* (2019), doi: https://doi.org/10.1016/j.cmpb.2019.105126

Highlights

- The data collected from patients are organized based on a time window, which is grouped into a Time-Window Data Chunk.
- The proposed method can decide which TWDC needs to be transmitted based on the priority from data assessment.
- The method is efficient in data collection and optimizes the processing order in prediction

An intelligent healthcare system with data priority based on multi vital biosignals

Rongjun Xie[a], Shahriar Badsha[b], Ibrahim Khalil[c], Mohammed Atiquzzaman[d]

[a] *Medibank, Australia*
[b] *University of Nevada, Reno 1664 N. Virginia St. M/S 171 Reno, NV 89557 United States*
[c] *RMIT University, Australia*
[d] *University of Oklahoma, Norman, USA*

Abstract

Background and Objective: Home-based personal healthcare systems are becoming popular and affordable due to the development of Internet of Things (IoT) devices. However, with an increasing number of users, such healthcare systems are challenged to store and process enormous volumes of data. For instance, multi-biosignal data are collected continuously from patients using IoT device like body sensors and are sent to the server by portable devices for further analysis (e.g. knowledge discovery or the clinical event prediction). These enormous amount of data from large number of patients are causing the transmission overhead and high latency in network which are responsible for inefficiency issues in clinical event prediction. To address these problems, in this paper, data assessment method is introduced to improve the efficiency in data collection and data prediction.

Methods: The assessment algorithm is inspired by National Early Warning Score (NEWS) used in Emergency Department. In our method, only the abnormal time-sequence data for analysis are sent to the server. Thus, the waiting time of data before prediction can be optimized because data with higher priority are processed in front of those with lower priority, which helps our system to provide diagnostic decisions in a proper time according to patients' urgency.

Results: Our experiments show that the proposed model ideally can save

Email address: `sbadsha@unr.edu` (Shahriar Badsha)

20% volume of data in the collection and can reduce 75% waiting time of data with the highest priority before predicting. In addition, the waiting time of data for further analysis is optimized compared to the normal processing flow.

Conclusion: The paper introduces an enhanced healthcare system with assessing data priority in order to optimize the data collection and the prediction in terms of data size and waiting time.

Keywords: Healthcare system, Data priority, vital biosignals

1. Introduction

Different kinds of healthcare systems are becoming prevalent since the rapidly changing landscape of health information technology and machine-learning technology. A variety of health information collected from patients are used to guide decisions from clinicians [1] and train those smart healthcare systems. As one of the common healthcare systems, the clinical decision support system (CDSS) is used to improve decision-making and thereby ensures care quality and safety [2].

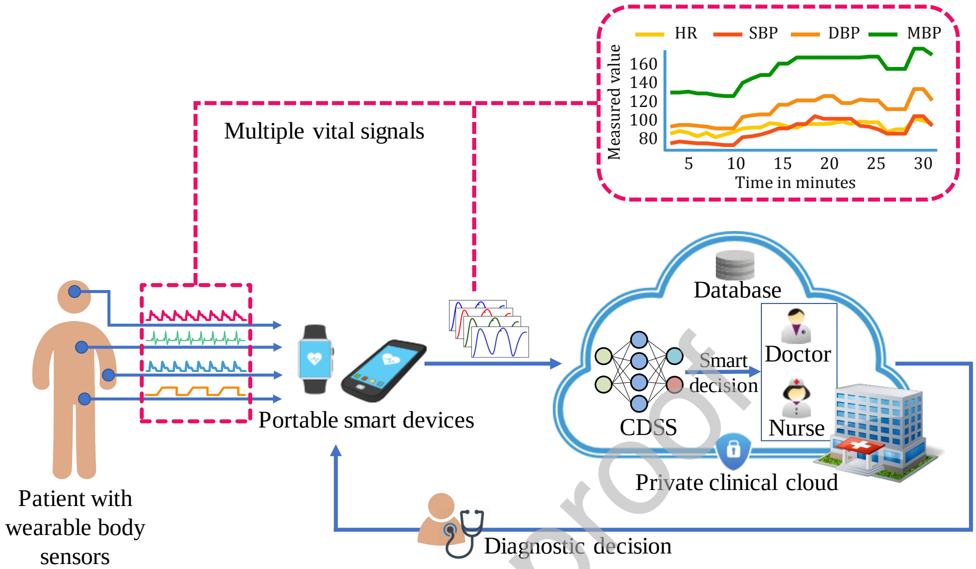


Figure 1: The typical architecture of remote healthcare systems. Portable smart devices collect multi biologicals of a patient with body sensors and sent these data to the server. The clinical institution can host the server in the Cloud or locally. A CDSS is used to do the smart prediction which supports the diagnosis from clinicians. If an abnormal event is detected, the system can notify the patient properly.

Figure 1 shows the typical architecture of a remote healthcare system, which includes two main components: the data collection module and a data processing module. As shown in the figure, the smart IoT device/ portable smart device continually collects the bio-signal data from patients by their wearable sensors and then sends the data to CDSS. The CDSS can be hosted in a Cloud-based platform or a local server employed by a medical institution. The system can not only provide smart decision (classification or prediction) to help clinicians (e.g. doctors or nurses), but also can send a proper notification to the patient once an abnormal clinical event is detected.

In addition, a variety of health information collected from patients are used to guide decisions from clinicians [1] and train those smart healthcare systems. For example, the advanced technology in wearable sensors has made it possible to monitor multiple vital signs of a patient anytime, anywhere. As multiple vital signs from a large number of patients are accumulated, the issue of big data is evolved. For instance, vital signs such as Heart Rate (HR), Blood Pressure (BP), Respiratory Rate (RR) and Oxygen Saturation

(SPO_2) are a crucial part of big medical data [3]. If the numerical value of each vital sign contains 4 bytes and the frequency of data collection is 1 minute, then for 6 vital signs total 24 bytes data gathered per minute, which is equivalent to 33.75 Kilobyte (KB) per day, or 12 Megabyte (MB) per year. If such data are gathered from 5 million patients, then the data amount will be 57.3 Petabyte (PB) per year. Processing these large amounts of data among different medical institutions is not feasible at all.

1.1. Motivation

Huge amounts of raw healthcare data are generated everyday [4], which is challenging the efficiency in the data collection. These data generated by the sensors have the 3Vs characteristics of big data: volume, velocity, and variety [5]. The case becomes more critical with the more elderly population for continuous monitoring. So an efficient method of data collection is required in healthcare systems.

On the other hand, the existing healthcare systems [6][7][8] predict data equally without considering urgency, which leads to delay in treatment of severe conditions of patients. In [9], the authors define clinical decision support as "providing clinicians with computer-generated clinical knowledge and patient-related information which is intelligently filtered and presented at appropriate times to enhance patient care". So embedding the system with a standardized method to optimize the waiting time of the data is essential. In a real medical institution, in order to make sure fair access to services and avoid confusion [10], clinical priority settings are used to sort the flow of patients so that the patients with more urgent conditions can be diagnosed or treated before those with less urgent conditions [11].

1.2. Tasks and goals

In this paper, an enhanced healthcare system with data assessment is designed to improve the efficiency of data collection and optimize the processing flow of data. Our goals are achieved by introducing data priority based on patients' urgency. Through our algorithms, the data volume is reduced in data collection, which can make data transmission faster. In addition, the system can adjust the waiting time of data before predicting based on the patients' urgency, which makes the system more practical and optimal.

One of the challenges of our work is to assess the priority based on time sequential data. Biosignal data are commonly collected from patients and are used for medical prediction in healthcare systems. As shown in Figure

1, these biosignal data are time-sequential data, which usually are collected in a specific frequency (e.g. every minute or 10 minutes). Based on clinical priority settings, these data can be provided different priority levels which can used to sort the flow of patients so patients with more urgent conditions should be diagnosed or treated before those with less urgent conditions [11]. But these time-sequential data have different numbers of increasing and decreasing trends, which makes patient urgency change over time. The accuracy and efficiency to assess patients' urgency with time-sequential biosignal data are considered in this challenge. In addition, when trying to collect less biosignal data from the patients, the system needs to provide complete data information to clinicians for accurate diagnoses. For example, doctors need to observe data in a specific time window (e.g. 1 hour or 1 day) in order to diagnose some chronic diseases correctly.

1.3. Contribution

The main contributions of this paper are summarized as follows.

- A time-window-based method for data collection is introduced in our system. The data collected from patients are organized based on a time window, which is grouped into a Time-Window Data Chunk (TWDC) in our method. Then our method decides which TWDC needs to be transmitted based on the priority from data assessment.
- A data assessment algorithm is proposed based on the clinical priority setting to identify patients' urgency. Our algorithms can not only assess each TWDC, but also can evaluate the priority based on time sequential priorities of TWDCs, which provides complete medical information for diagnoses.
- An enhanced healthcare system with data assessment is explored, which is efficient in data collection and optimizes the processing order in prediction. With patients' urgency, the system can reduce the medical data collected from patients by filtering normal data and can adjust the waiting time before predicting based on different priority levels.

2. Background

With significant advances in body sensors and Internet of Thing (IoT) devices, remote healthcare monitoring systems embedded with CDSS are proposed by different studies [6][7]. These systems are targeting the elderly with chronic diseases [12]. These diseases caused by irregular lifestyle, improper diet, and congenital genetic problems [13] are the main reason of many deaths in Australia and other western countries [6].

Different studies [6] [14] [8] consider multiple biosignals (e.g. ECG, blood pressure, heart rate, respiration and O_2 saturation) for future abnormality prediction. But the majority are at theoretical level and still far behind to be widely used in public.

Even though [6] and [7] introduce their frameworks at the application level, some improvement can still be considered to make the healthcare system better. The [7] mainly focuses on personal state estimation based on Hidden Markov Model (HMM). Specific rules are used to decide which data need to be transmitted. The [6] explores the Cloud-based framework to deal with the pressure of data storage and processing due to a huge amount of data. Both [6] and [7] use the mobile device to collect and transmit the raw bio-signals into the server continuously. Similar to [7], HMM is adopted to perform the clinical event prediction. Another practical example is the BioSign device [15] that can minimize the time of occurrence of critical clinical situation. But there is no predictive capability in the system.

The above systems show good practical solutions for healthcare, but all of them process the medical data from patients' with the first-in-first-out principle. However, in real clinical cases, a patient's urgency is commonly considered at the beginning in order to decide the order of medical services. The [16] introduces the effective triage system used in the Emergency Department (ED) when predicting Intensive Care Unit (ICU) admission or in-hospital mortality. The National Early Warning Score (NEWS) is explored in [17], which is a good predictor of patient outcomes and can provide additional value to monitor patients in the ED and in the hospital. According to the existing clinical settings, a patient's urgency plays a significant role in monitoring patients in the hospital. Therefore, when processing patients' medical data, the smart healthcare system is required to consider patients' urgency.

Compared to other state of the art literature, our study explores to build an enhanced healthcare system to improve the efficiency of data transmission and optimize the prediction flow based on the data priority.

3. Preliminary

3.1. Extreme Learning Machine (ELM)

The ELM [18] is a Single Hidden Layer Feed Forward Neural Network (SLFN) without tuning the hidden layer. The main advantage of ELM is that it overcomes limitations of backpropagation algorithms which are commonly used in artificial neural networks by randomly generating input weights and analytically calculating output weights. The main limitations of backpropagation include over fitting, high computation cost of the learning process and local minima. Moreover, the ELM's learning speed and performance are also significantly better than other conventional learning algorithms.

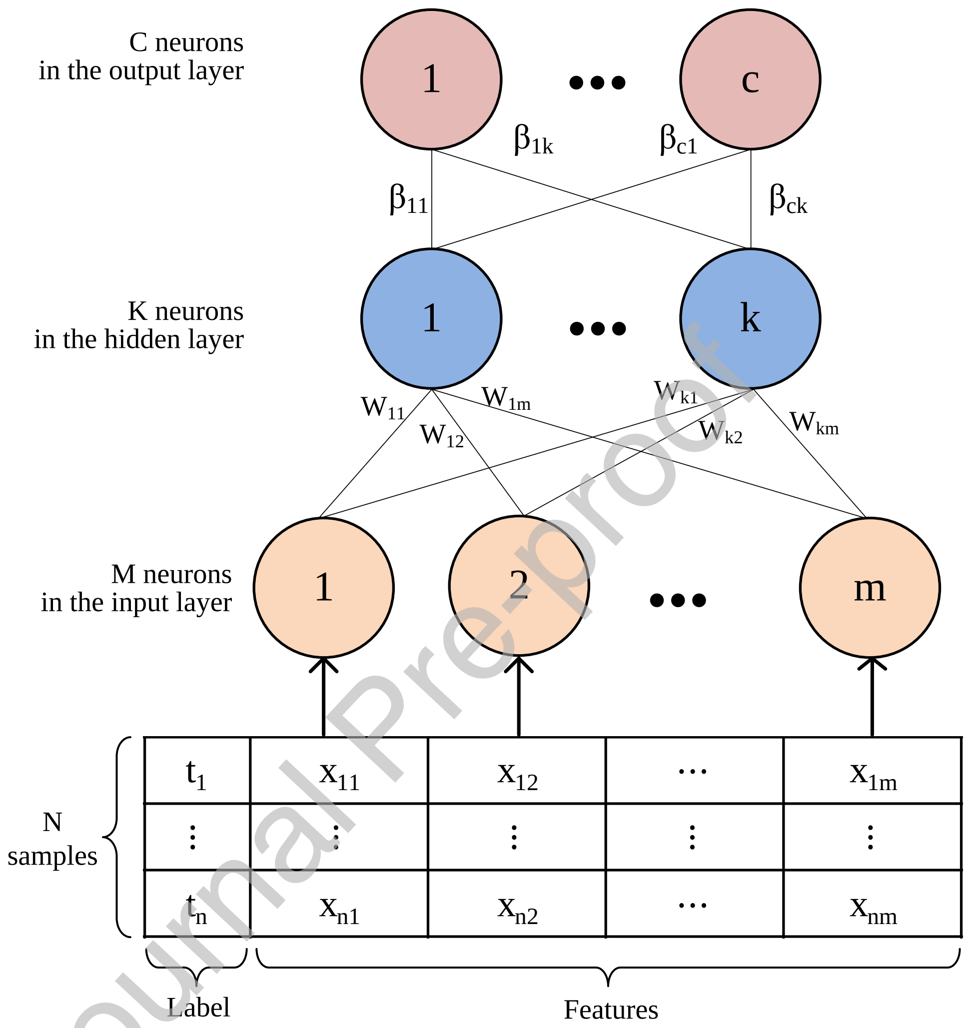


Figure 2: The structure of ELM. Samples with labels and features are fed into the input layer in ELM and then are operated with random input weights and an activation function in the hidden layer. The final model is calculated based on output matrices from the hidden layer.

As a structure of ELM shown in Figure 2, there are three different layers in ELM: the input layer, the hidden layer and the output layer. Suppose that ELM has M neurons in the input layer, K neurons in the hidden layer and C neurons in the output layer, for N arbitrary distinct samples $(\vec{x}_i, \vec{t}_i)$, where $\vec{x}_i = [x_{i1}, x_{i2}, x_{i3}, \cdots, x_{im}]^T \in \mathbb{R}^M$ and $\vec{t}_i = [t_{i1}, t_{i2}, t_{i3}, \cdots, t_{ic}] \in \mathbb{R}^C$. The

m, i and c represent the index of features, samples and neurons respectively. ELM with K hidden neurons is mathematically modeled as

$$\begin{gathered}\vec{o}_i = \sum_{k=1}^{K}[\vec{\beta}_k \cdot G(\vec{w}_k, b_k, \vec{x}_i)] = \sum_{k=1}^{K}[\vec{\beta}_k \cdot G(\vec{w}_k \cdot \vec{x}_i + b_k)] \\ (\vec{o}_k, \vec{\beta}_k \in \mathbb{R}^C, \vec{w}_k \in \mathbb{R}^M, b_k \in \mathbb{R}, i = 1, 2, \cdots, N)\end{gathered} \tag{1}$$

where $\vec{w}_k = [w_{k1}, w_{k2}, w_{k3}, \cdots, w_{km}]^T$ and b_k are random input weights in the k-th hidden node, $\vec{\beta}_k = [\beta_{k1}, \beta_{k2}, \beta_{k3}, \cdots, \beta_{kc}]^T$ is the weight vector connecting the k-th hidden node and the output nodes, $\vec{o}_i = [o_{i1}, o_{i2}, o_{i3}, \cdots, o_{ic}]$ is the i-th output vector of ELM, and finally $G(*)$ corresponds to an output of an activation function used in neurons of the hidden layer. Particularly, the value of elements in $\vec{t}_i$ is 1 when the output of neuron belongs to the sample class and the rest are -1. ELM can evaluate these N samples with zero error, which is the basic principle of least squares algorithm. The evaluation is shown in

$$\sum_{n=1}^{N} \|\vec{o}_i - \vec{t}_i\| = 0 \tag{2}$$

and can be expressed as

$$H \cdot \beta = T \tag{3}$$

where

$$H = \begin{bmatrix} G(\vec{w}_1, b_1, \vec{x}_1) & \cdots & G(\vec{w}_k, b_k, \vec{x}_1) \\ \vdots & \ddots & \vdots \\ G(\vec{w}_1, b_1, \vec{x}_n) & \cdots & G(\vec{w}_k, b_k, \vec{x}_n) \end{bmatrix}_{N \times K} = \begin{bmatrix} g_{11} & \cdots & g_{1k} \\ \vdots & \ddots & \vdots \\ g_{n1} & \cdots & g_{nk} \end{bmatrix}_{N \times K} \tag{4}$$

$$\beta = \begin{bmatrix} \vec{\beta}_1 \\ \vdots \\ \vec{\beta}_k \end{bmatrix}_{K \times C} \quad and \quad T = \begin{bmatrix} \vec{t}_1 \\ \vdots \\ \vec{t}_n \end{bmatrix}_{N \times C} \tag{5}$$

H is named as the hidden layer output matrix of ELM with a specific input dataset $X = [\vec{x}_1, \vec{x}_2, \vec{x}_3, \cdots, \vec{x}_n]$. The smallest norm least-squares solution of above linear system can be expressed as:

$$\hat{\beta} = H^{\dagger} \cdot T = (H^T H)^{-1} H^T T \tag{6}$$

$$H^T H = \begin{bmatrix} u_{11} & \cdots & u_{1k} \\ \vdots & \ddots & \vdots \\ u_{k1} & \cdots & u_{kk} \end{bmatrix}_{K \times K} \quad (7)$$

and

$$H^T T = \begin{bmatrix} v_{11} & \cdots & v_{1c} \\ \vdots & \ddots & \vdots \\ v_{k1} & \cdots & v_{kc} \end{bmatrix}_{K \times C} \quad (8)$$

where $H^\dagger$ is the pseudo inverse which extrapolates the inverse of matrix H in Equation (3).

4. The Proposed Design for Healthcare System

The proposed system model is shown in Figure 3. There are mainly two entities in our system: the client working on data collection and the server performing the data analytics.

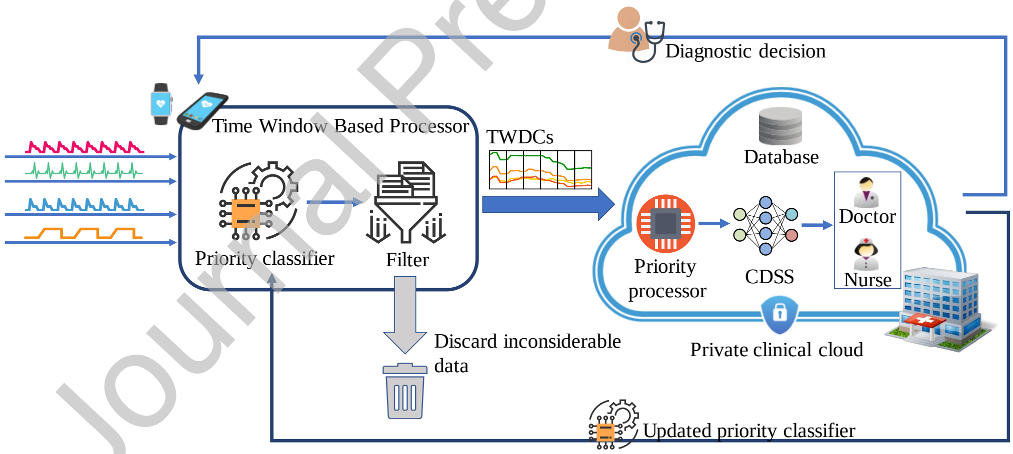


Figure 3: The proposed remote healthcare system. TWBP in a portable device is used to assess the priority of every TWDC and filter TWDCs with a specific time window (e.g. 10 minutes). The selected TWDCs are transmitted to the server. The general priority of sequential TWDCs in a specific observation window (e.g. 60 minutes) is calculated by the priority processor. Then the CDSS and clinicians can diagnose the data which are ordered based on their priority. Then a proper notification is sent once an abnormal event is detected. In addition, the data store in the database for further study and the system can synchronize the updated priority classifier to the TWBP.

- In the client, data collection is implemented in three steps.
  - The portable smart device collects the patient's bio-signal data continually from his or her body sensors. As shown in later sections, 6 biosignals (HR, SBP, DBP, MBP, RR and SPO_2) are considered in our system, which are used to detect 5 clinical events.
  - The processor in the device deals with the data based on a specific window (e.g. 10 minutes). The data in a window are defined as Time-Window Data Chunk (TWDC) and the processor a Time-Window Based Processor (TWBP). There are two main functions achieved in the processor: classifying the priority of TWDC and filtering TWDCs.
  - After filtering, the processor transmits time-series TWDCs and their priority to the server employed by a medical institution.
- In the server, data processing is achieved in three steps.
  - After receiving enough TWDCs within a specific observation window (e.g. 60 minutes), the server groups these TWDCs together and calculates the general priority by considering all TWDCs in the observation window. Then the server orders these grouped data into the waiting list based on the priority. Higher priority has a higher index, which means less time to wait for processing.
  - CDSS predicts the medical conditions using these grouped TWDCs and sends the data to clinicians once an abnormal clinical event is detected. Then proper diagnosis decisions are provided to the patient.
  - The TWDCs are stored in the database for backup and further knowledge discovery. The data are also used to update the priority classifier and generate medical knowledge which can be shared and improve the diagnosis prediction, explaining in [19].

4.1. The overview of system participants

Both the client and server play an important role in our proposed system.

- The client: It uses portable devices (e.g. smartphone or smartwatch) to collect patients' data continuously from body sensors and to evaluate data priority for these medical data based on a specific time window (e.g. 10 minutes). The data priority represents patients' urgency, which means that a more severe patient has a higher data priority. In addition, these time-window medical data and their priorities are organized in a Time-Window Data Chunk (TWDC). The client selects which data chunks need to be transmitted to the server based on their priorities. Practically, some wearable products like Omron HeartGuide[1] and Asus VivoWatch BP[2] have the capacity of monitor blood pressure.

- The server: It is employed by a medical institution, which is used to provide the accurate prediction of patients' medical conditions using their medical data collected from the client. The prediction order of patients' data is optimized by the data priority, which makes sure that diagnosis decisions are provided at appropriate times based on patients' urgency. In addition, with more and more patients' data, the server can provide a more accurate solution to assess patients' data and to predict different medical conditions by introducing P2P learning from our earlier work [19, 20].

4.2. The overview of system components

In our proposed system, the Time-Window Based Processor (TWBP) is the main component in the client, which includes a priority classifier for data assessment and a filter for data selection. Below, the process of priority classifier and filter are described briefly,

- Priority classifier: It uses a Machine-Learning (ML) based method to classify TWDCs into different priorities automatically. The data priority plays an important role in our system optimization. More specifically, the ML classifier is trained with medical samples containing four vital biosignals and the labels of training data are identified by our data assessment algorithm based on a real-life clinical setting.

- Filter: It helps the system to collect patients' medical data more efficiently by considering different data priorities. In our healthcare system, the abnormal data are more valuable and considerable than the

[1] `https://omronhealthcare.com/products/heartguide-wearable-blood-pressure-monitor-bp8000m/?utm_source=cj&utm_medium=affiliate&cjevent=ac174e01c62411e9816500450a1c0e13`

[2] `https://www.asus.com/VivoWatch/ASUS-VivoWatch-BP-HC-A04/`

normal data since our system is required to provide corresponding diagnosis decisions when detecting the abnormal data. So the filter targets the abnormal data which have higher data priority.

In addition, there are three more components in the server of our model: database for data storage, priority processor for prediction queue management and CDSS for data prediction.

- Database: It is used for data storage to record TWDCs from different patients. The records of different TWDCs can be used to provide reliable long-term diagnoses for patients and to discovery useful medical knowledge which can improve the smart healthcare system.

- Priority processor: It helps to arrange the prediction order of patients' medical data based on data priority, which can reduce the waiting time of urgent patients' data and can help our system to provide assistance to patients' at the appropriate time based on their urgency. An algorithm to assess long-term data priority is designed based on the priorities of time-series TWDCs. For example, with the calculation of a data priority in an hour, the priorities of 6 time-series TWDCs using a 10-minute window are required.

- CDSS: It is a Machine-Leaning based classifier which can predict patients' medical conditions accurately using medical data collected from patients in real time. In order to meet the changing in the clinical environment (e.g. new diseases), the CDSS also has the ability to do P2P learning[19], which can improve the diagnosis accuracy efficiently and effectively.

5. Design details

Similar to [6], our system targets various medical information, including 6 vital biosignals shown in Table 1 and 5 clinical events listed in Table 2. Then criteria for assessing patient's urgency are explained. Finally, two key components of our system are described in details: the TWBP in the client and priority processor in the server.

5.1. Description of vital biosignals and clinical events

Our system considers numerical trend data of six vital biosignals shown in Table 1 to identify the early sign of clinical deterioration and assess treatment effects. The values of vital biosignals are various since different conditions (e.g. age and sex) of patients have impacts on these values. In order to provide a basic diagnosis, medical science defines a common normality range of each biosignals which are shown in Table 1.

Table 1: Vital biosignals and their normal range.

Biosignal	Acronym	Normal range
Respiratory rate	RR	12-18 breaths per min
Blood oxygen saturation	SPO_2	95-100%
Heart rate	HR	60-100 beats per min
Systolic blood pressure	SBP	90-120 mmHg
Diastolic blood pressure	DBP	60-90 mmHg
Mean blood pressure	MBP	60-110 mmHg

Tachycardia and bradycardia are defined as rising and fall in HR respectively. The rise in blood pressure is known as hypertension and fall is called hypotension. Rise and fall in RR are called as tachypena and bradypena respectively. In addition, deficiency in SPO_2 is named hypoxia. Our system detects these clinical conditions happen at the same time and last for a specific time period, shown in Table 2. According to [6], it is reliable to use 1-hour data to predict the coming clinical event.

Table 2: Targeted clinical events.

Clinical event	Acronym
Simultaneous Tachycardia, Hypotension, Tachypena and Hypoxia for more 30 minutes	THTH
Simultaneous Bradycardia, Hypotension, Tachypena and Hypoxia for more 30 minutes	BHTH
Simultaneous Tachycardia, Hypertension, Tachypena and Hypoxia for more 30 minutes	TTTH
Simultaneous Tachycardia, Hypotension, Bradypena and Hypoxia for more 30 minutes	THBH
All six biosignals are in normal range	NNNN

5.2. Criteria for data assessment

Our system employs a supervised learning algorithm to assess patients' urgency and achieve a proper priority of the data. In order to label all samples, our system develops a method similar to National Early Warning Score (NEWS) [17] which is used in Emergency Department. As it is difficult to obtain the level of consciousness (LOC) of the patients automatically by smart devices, our system removes LOC to simplify our assessment method. In addition, since [21] shows the weak contribution of the systolic blood pressure (SBP) and temperature parameters to NEWS performance and suggests to remove the temperature, our method considers 4 out of 6 vital biosignals to assess the data urgency.

As shown in Table 3, these biosignals are RR, SPO_2, HR and SBP. The related scores are given based on their value thresholds. For example, if the value of RR is from 9 to 11, score 1 is provided by our method. It is important to note that all values are integer. In addition, the thresholds of SBP are different from NEWS. NEWS results 0 when the SBP value is from 110 to 219, but a normal patient SBP value should be always within 80 and 120. The score 0 of other 3 vital biosignals means the value is within the normal ranges. So in order to maintain the consistency, our system modifies the score thresholds of SBP when its value is larger than 110.

Table 3: Vital biosignals defining the triage.

Score / Sign	3	2	1	0	1	2	3
RR	≤ 8		9-11	12-20		21-24	≥ 25
SpO_2	≤ 91	91-92	94-95	≥ 96			
HR	≤ 40		41-50	51-90	91-110	111-130	≥ 131
SBP	≤ 90	91-100	101-110	110-120	121-170	171-219	≥ 220

Similar to NEWS [22], our assessment method calculates the total score of all vital biosignals and classifies it into 4 priorities. The classification criteria are based on the waiting time which is defined in NEWS. The detail definition of all priorities is shown in in Table 4. In particular, if the total score is 0, the system considers discarding the data since all vital biosignals are normal. The range of each priority is used in evaluating a time sequential priority group, explained in the later section.

Table 4: Priority definition by score of vital biosignals.

Priority	0	1	2	3
Label	Normal	Elective	Urgent	Emergency
Waiting time	-	≤ 1 h	≤ 0.5 h	0
Range	0	(0,1]	(1,2]	(2, 3]
Total Score	0	1-3	4-6	≥ 7 or Score of 3 in any sign

5.3. *Time-Window Based Processor (TWBP)*

Instead of sending data continually from the client to the server, our system deals with the TWDC in the TWBP and considers which TWDC is required for diagnoses. Figure 4 shows the workflow of a TWBP with four time-series biosignal data (HR, SBP, RR and SPO_2). The TWBP goes through 3 steps as follows.

- The features of a TWDC of all vital biosignals are extracted.
- The priority of the TWDC is detected by the priority classifier.

- According to its priority, the filter decides which TWDC is considerable for diagnoses.

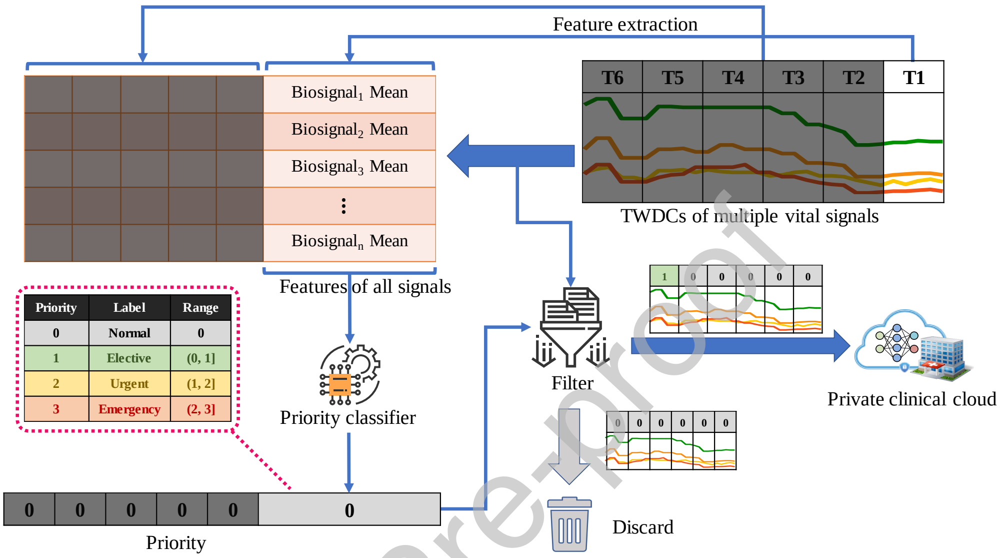


Figure 4: The workflow of the time-window based processor. The priority classifier assesses the priority of a TWDC based on the mean feature vector from multi biosignals. Then the filter can decide whether the TWDC should be sent or discarded according to the sequential priories.

5.3.1. Time-Window Data Chunk (TWDC) and its priority

The main role of the priority classifier in TWBP is to label TWDC with a corresponding data priority. The TWDC and its priority are introduced as follows, which are the foundation for our proposed system processing patients' data.

Suppose the collection of discrete time-series data (X) of time length (T) is split into K windows (Ws) with equal size. There are N samples in each window (W) where N is equal to T/K. So the time-series data can be considered as a sequence of a TWDC—$X_1(t), X_1(t), X_1(t), \cdots, X_K(t)$. Then the samples in each window are used to construct features. For example, if a 10-minute window is employed in the system, 60-minute data can be divided into a sequence of 6 TWDCs (from T1 to T6 shown in Figure 4). In order to detect the TWDC priority, the mean value of each biosignal is obtained from samples in a window based on Equation 9.

$$f = \frac{\sum_{K=0}^{K=N}(X_K(t)))}{N} \tag{9}$$

Then all mean values from all vital biosignals are grouped as a feature vector—$(f_1, f_2, f_3, \cdots, f_s)$ where s is the total number of biosignals. As our assessment algorithm described above only works with integers, all components of the feature vector have to be changed to the value of a number rounded to the nearest integer $(f^{'})$ which shown in

$$f^{'} = \lfloor f + 0.5 \rfloor \tag{10}$$

Similar to [6], a 10-minute window is adopted in our system. After extracting features of a TWDC, the trained priority classifier can detect the priority of the TWDC. More specifically, any kinds of machine-learning algorithms can be used as the priority classifier.

5.3.2. Data collection with the filter

In our proposed system, the filter in TWBP decides the considerable medical data based on its priority, which can optimize the data collection by reducing the data volume. Our proposed data collection processor is explained as follows.

Considering priority 0 means the values of all biosignals in a TWDC are in the normal range, if an abnormal TWDC existing in a buffer, the whole sequential TWDCs need to be transmitted for further prediction. Figure 5 shows an example of data collection. With a new TWDC coming into the TWBP every time period (t), the blue buffer window slips to the left as the increase of t. There are three statuses of a TWDC: send, pending and discard. Once there is the priority of a TWDC larger than 0 in the buffer, these 6 TWDCs are sent to the server. But if the priorities of the new TWDC and the rest are all 0s, the new TWDC is marked as pending. When the buffer moves out of the pending TWDC, the pending TWDC is discarded. Clearly, instead of simply sending TWDCs one by one, our algorithm adjusts the data collection process based on abnormal TWDCs.

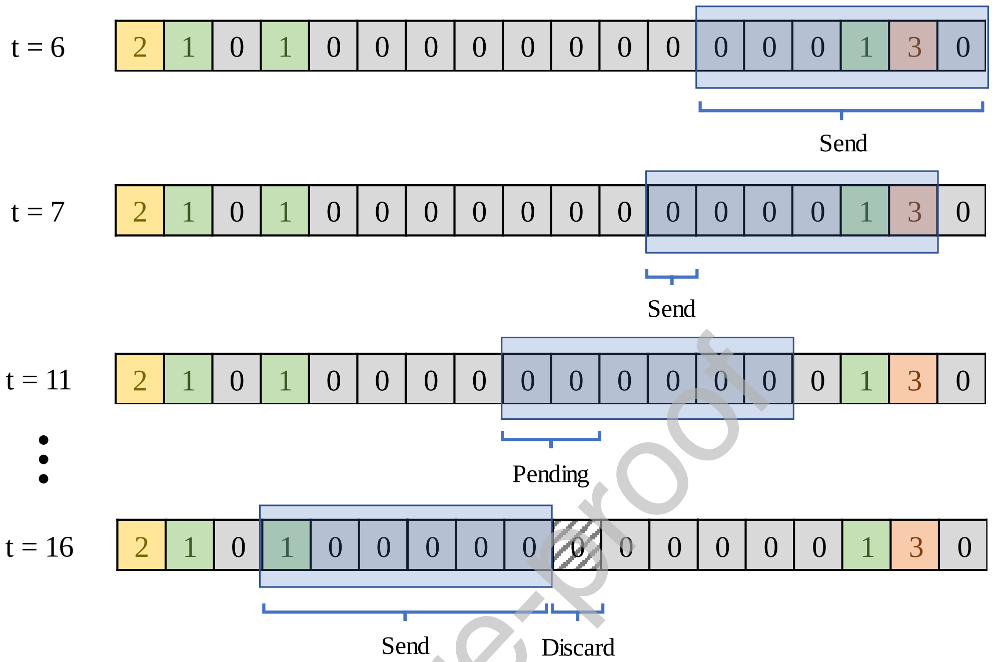


Figure 5: An example of data transmission. One new TWDC comes to the list as the increase of t. The blue buffer window includes all TWDCs processed by the system. In every t, the status of TWDCs are shown. In particular, pending of a TWDC means the system need to obtain more TWDCs to decide whether the TWDC should be sent or discarded.

5.4. *Priority processor*

Similar to [6], in order to diagnose different clinical events, our system needs to consider sequential TWDCs instead of just a single TWDC, which is achieved in the priority processor shown in Figure 6.

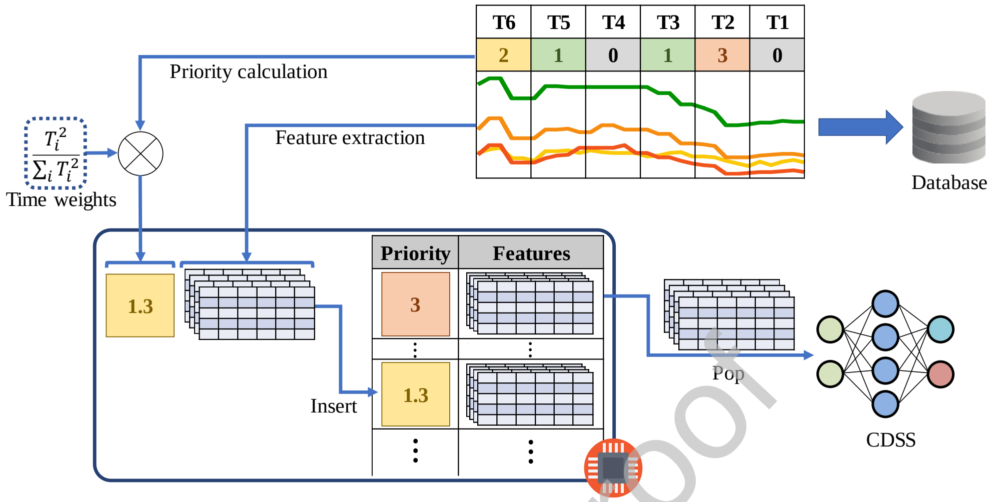


Figure 6: The workflow of the priority processor. The general priority of sequential TWDCs is calculated by TWDC priorities and the time weights. Based on the priority, the features of the whole TWDCs are inserted to the waiting list for predicting. Data with the highest priority in the list is processed by the CDSS.

As mentioned in the previous section, totally K time-series priorities within the window W are used to decide the general priority of these TWDCs. Instead of averaging all priorities, time weight factors defined in Equation 11 are introduced in our methods to maintain the importance of time.

$$w_i = \frac{T_i^2}{\sum_{i=1}^{i=K} T_i^2} \quad \text{and} \quad T_i = i * W \tag{11}$$

Denote P as the set of all priorities, then $P = (p_1, p_2, \cdots, p_K)$. The general priority $p^{'}$ of continuous TWDCs is calculated as follows:

$$p^{'} = P \cdot W = \sum_{i=1}^{i=K} p_i * w_i \tag{12}$$

The meaning of $p^{'}$ within a specific range is shown in Table 4. Then our system puts all extracted features of sequential TWDCs into the waiting list based on their priority, shown in Figure 6. The data with the highest priority are selected to do the prediction. The feature extraction process is similar to [6]. In each biosignal of a TWDC, 5 features are extracted, which are mean, standard deviation, median, the number of increasing trends and decreasing trends. A feature matrix which includes all extracted features of sequential TWDCs is used to predict the coming clinical event.

6. Results

The experimental section is categorized in four sections.

- Data preprocessing: explaining how the datasets are preprocessed for experimental evaluation.
- The accuracy of priority classification: showing the performance of three different classifiers categorizing data into four priorities.
- The efficiency of data collection: evaluating the performance of data collection in our proposed system in terms of the data volume and the data sending frequency. The experiments are conducted with 10 different ratios (from 0.1 to 1) of abnormal data in the system.
- The average waiting time: It is used to measure the performance in the waiting queue. The time-series data priorities are generated using two different discrete distributions: uniform distribution and normal distribution.

6.1. Data preprocessing

In order to evaluate the accuracy of data priority assessment, 6 vital biosignals from MIMIC-II numeric dataset of MIT physiobank are used in the experiment. More specifically, only the records containing at least 24 hours numerical trend data of these 6 biosignals are adopted. Most of the biosignals are sampled in one minute. Data sampled per second are converted to per minute sampling by averaging all values in a minute. The data missing values over a long period and the noisy data are also filtered. Finally, 1023 records are obtained for the experiments.

As mentioned in the previous section, 4 out of 6 vital biosignals are used to identify the priority. These signals are HR, SBP, RR and SPO_2. In every biosignal, after averaging all values in a 10-minute window, each sample has 4 mean values. Then the algorithm described above are used to label all samples. In order to balance the dataset, 1500 samples of each priority are randomly selected. Considering totally 4 priorities are targeted, there are 6000 samples in the dataset. The dataset is normalized by the z-score linearly transformation. 70% of the samples form the training dataset and the rest the testing dataset, which are shown in Table 5.

Table 5: The preprocessed data with 4 data priorities.

Data priority	0	1	2	3
Type	Normal	Abnormal		
The number of training samples	1050	1050	1050	1050
The number of testing samples	450	450	450	450
Total	1500	1500	1500	1500

6.2. Priority classification

The accuracies of different neural network classifier are shown in Table 6. In particular, the classifiers based on the decision tree are not considered in our experiments because they are good at classifying the data labels generated from rules. In our test, they can provide over 99% accuracy. Extreme learning machine (ELM) [18] is run in Matlab, which has 500 hidden neurons with the sigmoid activation function. The result classifiers are run in Weka 3.8 [23] with default settings.

Table 6: The accuracy comparison among different learning algorithm.

	Multilayer Peceptron	SMO with RBF kernal	ELM
Accuracy (%)	71.89	66.83	80.6
Training Time (s)	2.82	13.53	1.09

ELM shows the best accuracy 80.6% among all candidate neural network classifier. The confusion matrix from the classification result is shown in Table 7. Compare to the baseline (25%), ELM shows a significant improvement in classification accuracy. But the model is not good to distinguish the normal and elective data.

Table 7: The confusion matrix after performing the classification. Here 80% accuracy is obtained.

	Normal	Elective	Urgent	Emergency
Normal	445	136	10	1
Elective	4	248	18	8
Urgent	1	66	380	63
Emergency	0	0	42	378

Except the overall accuracy and confusion matrix, different accuracy measures (precision, sensitivity, and specificity) for each priority are applied shown in Table 8. From this observation, the classification is not sensible to the data with the elective priority.

Table 8: The performance measure of each priority using ELM.

	Normal	Elective	Urgent	Emergency
Precision (%)	75.169	88.889	74.510	90.000
Sensitivity (%)	98.670	55.111	84.444	84.000
Specificity (%)	89.111	97.705	90.377	96.891

6.3. Data collection efficiency

In this experiment, the data transmission process is stimulated with Python. In order to simplify the stimulation, the priorities from 1 to 3 are regarded as the abnormal priority 1. Based on 10 different abnormal data ratios from 0.1 to 1, a list with 6000 binary values is generated. The abnormal data ratio is calculated as

$$R_{abnormal} = \frac{\text{The number of 1s in the list}}{\text{The length of the list}} \tag{13}$$

. And the ratio of transmission data is calculated as

$$P_{data} = \frac{\text{The number of data sent}}{\text{The length of the list}} \tag{14}$$

. And the ratio of data sending requests is calculated as

$$F_{sr} = \frac{\text{The number of sending requests}}{\text{The length of the list}} \tag{15}$$

.

As shown in Figure 7, over 20% of the TWDC can be discarded in our proposed system when 10% TWDCs are abnormal. 99.8% of TWDCs are transmitted when the abnormal ratio is 40%. If over 40% abnormal TWDCs exist in the list, all TWDCs need to be transmitted.

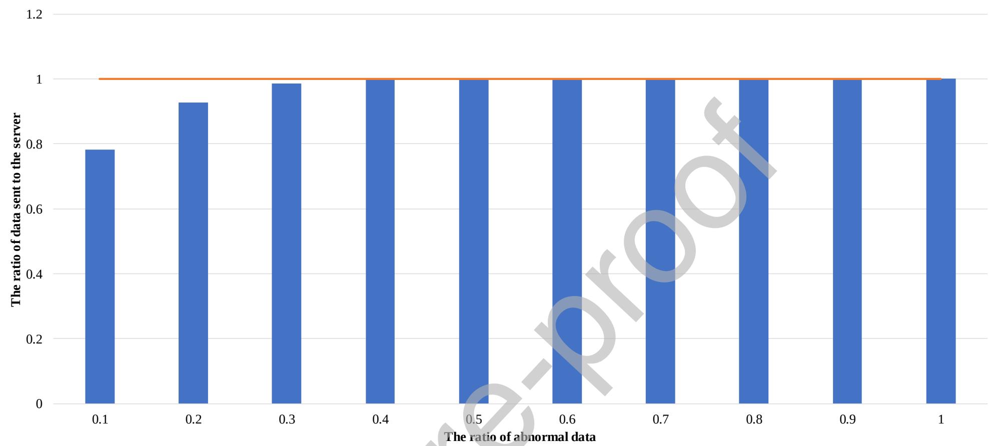


Figure 7: The ratio of data sent to the server with different abnormal data ratios.

Similar to the data transmission percentage, when the list has 10% abnormal TWDCs, over half of the data sending requests can be saved, because the system groups the sequential TWDCs and sends once. And 0.2% of the requests are saved with 60% data abnormal ratio. When there are over 60% abnormal TWDC in the list, each TWDC is sent to the server one by one, which means the ration of sending requests is 1.

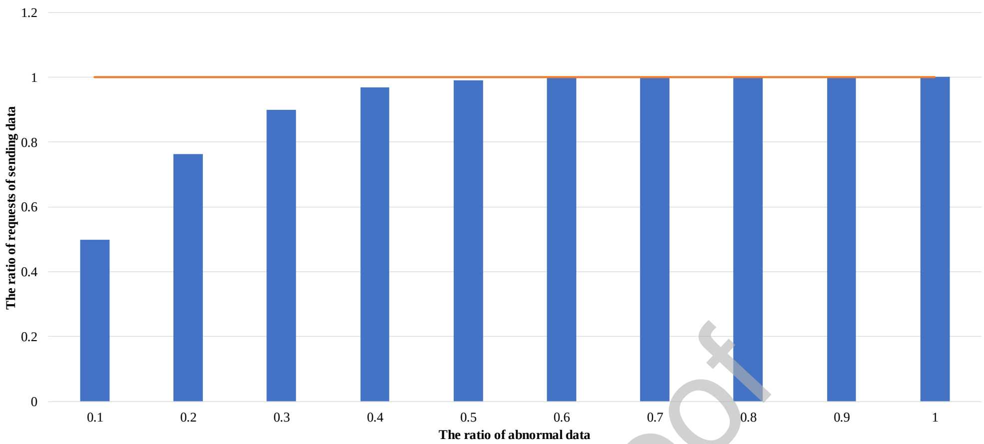
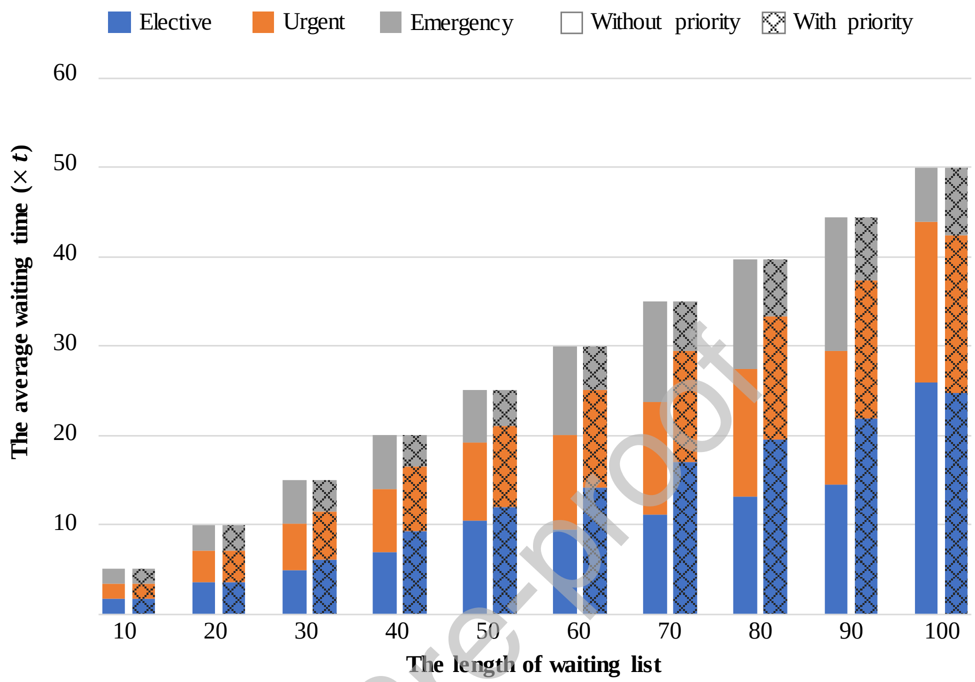


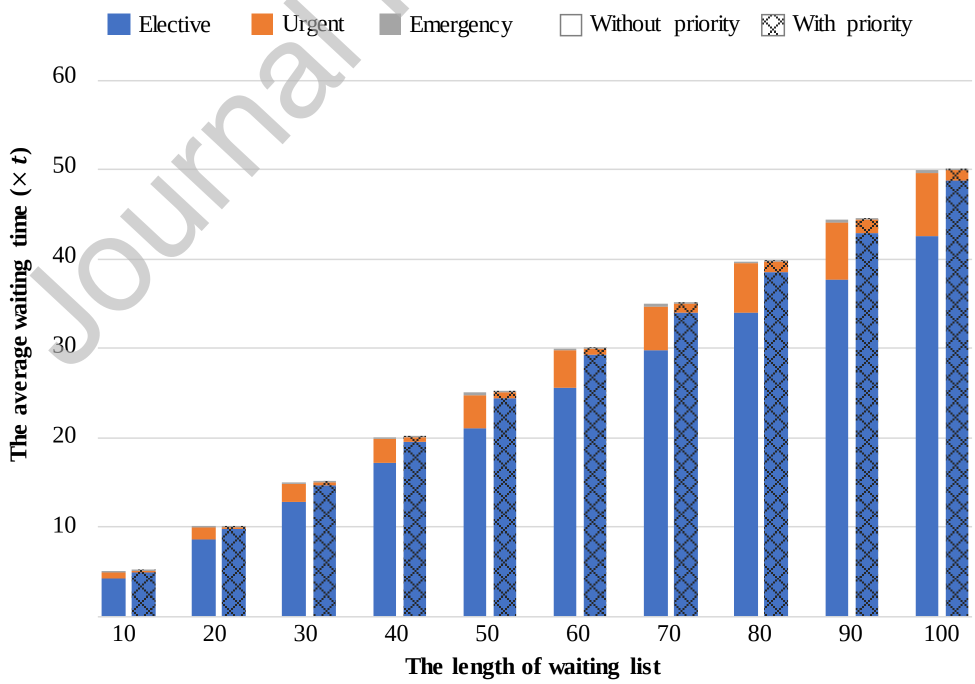
Figure 8: The ratio of requests of sending data with different abnormal data ratios.

6.4. The average waiting time

In this part, the waiting time of patients' records before prediction are evaluated with the setting that the waiting list can contain Q records for further prediction. After the prediction of all Q records is finished, new Q records come to the waiting list. The CDSS consumes only 1 records every time and spends t on predicting the record. The stimulation is developed with Python. 3000 records with 3 abnormal priories are generated from the uniform distribution and the standard normal distribution respectively. Thus, $3000/Q$ batches are used for prediction and the waiting time is calculated by averaging the waiting time of records with different priorities in all batches. When considering no priority, the system deals with every data based on First In First Out (FIFO). Otherwise, the system orders the data in the waiting list according to their priories and then processes them one by one.



(a) Priorities of all data follow the uniform distribution.



(b) Priorities of all data follow the standard normal distribution.

Figure 9: The waiting time of records with and without priority. Every record is processed in Time t.

The results with different data distributions are shown in Figure 9. From the figures, as the waiting list becomes longer our system increases the waiting time of the data with the lower priority (Elective).

In Figure 9(a), the waiting time without considering priority is fluctuant with the increase of the waiting list length, while the results with priority are stable regardless the changing of the length of the waiting list. In Figure 9(b), compared to the results without considering the data priority, our proposed method can significantly reduce the waiting time of urgent data.

Table 9 shows the comparison of the waiting time of the system with and without priority when the length of the waiting list is 10 and all data priories follow the standard normal distribution. As an assumption, the system and clinicians spend 10 minutes on diagnosing each data. Clearly, the waiting time of all data with different priority is less than the maximum waiting time from NEWS clinical definition. But the system can optimize the processing flow of data based on data priority. 75% waiting time of urgent data can be reduced by our proposed method.

Table 9: Waiting time comparison. Assume that every data requires 10 min to get the prediction result. The length of the waiting list is 10 and data priories follow the standard normal distribution.

	Normal	Elective	Urgent	Emergency
NEWS suggestion	-	≤ 1 h	≤ 0.5 h	0
CDSS with priority	-	10 min	1 min	0
CDSS without priority	-	26 min	4 min	0.2 min

7. Discussion

In this section, some of the limitations of this research are discussed and envisioned to address in the future to improve the system. This paper mainly focuses on improving efficiency in the smart healthcare system in terms of data transmission and data prediction. The impact of different standards like the Modified Early Warning Score (MEWS) and Paediatric Early Warning Score(PEWS) are not considered. Another shortcoming is that the proposed system is not tested in a real-life environment while considering a simulation and producing meaningful results. Also, our experiments rely on simulated prototypes and focus on system development, validation, and performance evaluation.

Publicly available data of patients monitored in hospital beds are used, presuming a similar nature for real-life data collected in a controlled manner from wearable sensors. Note that, the system does not collaborate with medical institutions to test the system in a test-bed environment for collecting real-time patients' data and to evaluate the performance in the real-life environment.

In order to address the above limitations, different studies and methods need to be investigated. By proposing an enhanced method to evaluate patients' urgency based on their personal situations, the healthcare system can be improved significantly. For example, it is supposed that most of the time the BP value of a hypertensive patient is higher than the normal. Content-aware techniques [24] can be introduced to evaluate the urgency more correctly since different contents of patients (e.g. running or sleeping) have a significant impact on the value of their biosignals.

8. Conclusion

In this paper, an enhanced healthcare system with assessing data priority is introduced in order to optimize the data collection and the prediction in terms of data size and waiting time. Novel algorithms inspired by real-world clinical settings are developed to evaluate data priority which can represent patients' urgency. Through the data priority algorithms, not only fewer data are collected from patients, but also complete medical information required in long-term accurate diagnoses is provided in our proposed system. Considering different distributions of data priority in the real case, our extensive experiments show that our proposed method can improve the efficiency of data collection and can perform optimization of waiting times according to patients' urgency. The system can ideally save 20% volumes of data in the transmission and reduce 75% waiting time of urgent data before predicting.

In our research, other autonomous functional requirements of real-time healthcare systems, such as low-level infrastructure of sensors, sensor failures, the reliability of communication between sensors and mobile devices, noise in sensor data and network fault management, are ignored. Such requirements need an independent research investigation. Our proposed model is a foundation to expand the scope of multiple research directions.

In the future, different medical standards can be investigated with our system architecture and the system should be tested in a real-life environment. Our system can also leverage more enhanced data evaluation methods for patients like content-aware techniques, improving prediction accuracy.

9. Ethics Approval

The data used in this study was sourced from published literature and no new clinical data was used. There is thus no need for ethics approval.

10. Declaration of Competing Interest

The authors declare no competing or financial interest in this work.

11. Funding

No funding was received for this work.

12. References

References

[1] R. Briere, Institute of medicine: Crossing the quality chasm, Washington, DC (2001).

[2] K. Dunn Lopez, S. M. Gephart, R. Raszewski, V. Sousa, L. E. Shehorn, J. Abraham, Integrative review of clinical decision support for registered nurses in acute care settings, Journal of the American Medical Informatics Association 24 (2017) 441–450.

[3] Y. L. Zheng, X. R. Ding, C. C. Y. Poon, B. P. L. Lo, H. Zhang, X. L. Zhou, G. Z. Yang, N. Zhao, Y. T. Zhang, Unobtrusive sensing and wearable devices for health informatics, IEEE Transactions on Biomedical Engineering 61 (2014) 1538–1554.

[4] A. Pantelopoulos, N. G. Bourbakis, A survey on wearable sensor-based systems for health monitoring and prognosis, IEEE Trans. Systems, Man, and Cybernetics, Part C 40 (2010) 1–12.

[5] M. A. Beyer, D. Laney, The importance of big data: a definition, Stamford, CT: Gartner (2012) 2014–2018.

[6] A. R. M. Forkan, I. Khalil, Peace-home: Probabilistic estimation of abnormal clinical events using vital sign correlations for reliable home-based monitoring, Pervasive and Mobile Computing 38 (2017) 296–311.

[7] P. Jiang, J. Winkley, C. Zhao, R. Munnoch, G. Min, L. T. Yang, An intelligent information forwarder for healthcare big data systems with distributed wearable sensors, IEEE Systems Journal 10 (2016) 1147–1159.

[8] O. Salem, Y. Liu, A. Mehaoua, R. Boutaba, Online anomaly detection in wireless body area networks for reliable healthcare monitoring, IEEE J. Biomedical and Health Informatics 18 (2014) 1541–1551.

[9] J. M. Teich, J. A. Osheroff, E. A. Pifer, D. F. Sittig, R. A. Jenders, Clinical decision support in electronic prescribing: recommendations and an action plan: report of the joint clinical decision support workgroup, Journal of the American Medical Informatics Association 12 (2005) 365–376.

[10] D. C. Hador, S. C. of the Western Canada Waiting List Project, et al., Setting priorities for waiting lists: defining our terms, Canadian Medical Association Journal 163 (2000) 857–860.

[11] O. F. Norheim, Clinical priority setting, BMJ 337 (2008).

[12] M. Von Korff, E. H. Wagner, K. Saunders, A chronic disease score from automated pharmacy data, Journal of clinical epidemiology 45 (1992) 197–203.

[13] W. C. Willett, J. P. Koplan, R. Nugent, C. Dusenbury, P. Puska, T. A. Gaziano, Prevention of chronic disease by means of diet and lifestyle changes, in: D. T. Jamison, J. G. Breman, A. R. Measham, G. Alleyne, M. Claeson, D. B. Evans, P. Jha, A. Mills, P. Musgrove (Eds.), Disease Control Priorities in Developing Countries, The International Bank for Reconstruction and Development/The World Bank, 2006, pp. 833–850.

[14] H. Cao, L. Eshelman, N. Chbat, L. Nielsen, B. Gross, M. Saeed, Predicting icu hemodynamic instability using continuous multiparameter trends, in: Engineering in Medicine and Biology Society, 2008. EMBS 2008. 30th Annual International Conference of the IEEE, IEEE, pp. 3803–3806.

[15] L. Tarassenko, A. Hann, A. Patterson, E. Braithwaite, K. Davidson, V. Barber, D. Young, Biosign : multi-parameter monitoring for early warning of patient deterioration (2005) 71 – 76.

[16] C. Barfod, M. M. P. Lauritzen, J. K. Danker, G. Sölétormos, J. L. Forberg, P. A. Berlac, F. Lippert, L. H. Lundstrøm, K. Antonsen, K. H. W. Lange, Abnormal vital signs are strong predictors for intensive care unit admission and in-hospital mortality in adults triaged in the emergency department-a prospective cohort study, Scandinavian journal of trauma, resuscitation and emergency medicine 20 (2012) 28.

[17] N. Alam, I. Vegting, E. Houben, B. van Berkel, L. Vaughan, M. Kramer, P. Nanayakkara, Exploring the performance of the national early warning score (news) in a european emergency department, Resuscitation 90 (2015) 111–115.

[18] G.-B. Huang, H. Zhou, X. Ding, R. Zhang, Extreme learning machine for regression and multiclass classification, IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics) 42 (2012) 513–529.

[19] R. Xie, I. Khalil, S. Badsha, M. Atiquzzaman, Fast and peer-to-peer vital signal learning system for cloud-based healthcare, Future Generation Computer Systems 88 (2018) 220–233.

[20] X. Rongjun, I. Khalil, S. Badsha, M. Atiquzzaman, Collaborative extreme learning machine with a confidence interval for p2p learning in healthcare, Computer Networks 149 (2019) 127–143.

[21] L. Lus, C. Nunes, Short national early warning score developing a modified early warning score, Australian Critical Care (2017).

[22] N. C. E. Committee, et al., National early warning score national clinical guideline no. 1 (2013).

[23] I. H. Witten, E. Frank, M. A. Hall, C. J. Pal, Data Mining: Practical machine learning tools and techniques, Morgan Kaufmann, 2016.

[24] A. Forkan, I. Khalil, Z. Tari, Cocamaal: A cloud-oriented context-aware middleware in ambient assisted living, Future Generation Computer Systems 35 (2014) 114 – 127. Special Section: Integration of Cloud Computing and Body Sensor Networks; Guest Editors: Giancarlo Fortino and Mukaddim Pathan.