Heliyon 8 (2022) e11372

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

Research article

A hypoglycemia early alarm method for patients with type 1 diabetes based on multi-dimensional sequential pattern mining

Ning Ma, Xia Yu *, Tao Yang, Yuhang Zhao, Hongru Li

College of Information Sciences and Engineering, Northeastern University, Shenyang, 110819, China

ARTICLE INFO

Dataset link: http://smarthealth.cs.ohio.edu/OhioT1DMdataset.html

Keywords:
Multi-dimensional sequential pattern mining
UniSeq algorithm
Hypoglycemia early alarm
Type 1 diabetes

ABSTRACT

Hypoglycemia is a limiting factor for blood glucose management. Serious symptoms such as seizures, and coma may occur during severe hypoglycemia, and nocturnal hypoglycemia is particularly dangerous for patients with type 1 diabetes (T1D). An effective early alarm method is essential for hypoglycemia prevention but challenging, as blood glucose is affected by many factors and the hypoglycemia sequence patterns vary from person to person. In this paper, we proposed a hypoglycemia early alarm method for mining the hidden information in blood glucose based on multi-dimensional sequential pattern mining. The blood glucose, meal, and insulin time series information were used to construct a multi-dimensional database, then the UniSeq algorithm was used to extract multi-dimensional hypoglycemia sequence patterns. Hypoglycemia early alarm was realized through pattern matching with real-time blood glucose. The public OhioT1DM dataset was used for performance evaluation. The experiment results were: 75.76% Sensitivity, 75% Precision, 75.38% F1 score, and 25.17 minutes early alarm time. The result verified that multi-dimensional sequential pattern mining can extract more hidden information and demonstrate more potential significance in providing comprehensive diagnosis support for personalized treatment. Furthermore, early alarms for potential hypoglycemia can also reserve sufficient time for blood glucose management.

1. Introduction

Diabetes is a chronic metabolic disease [1] and hypoglycemia is a potential risk for T1D patients. T1D patients suffer thousands of symptomatic hypoglycemia episodes over a lifetime, and one or more episodes of severe, temporarily disabling hypoglycemia. Previous studies have demonstrated that frequent hypoglycemia events lead to a high incidence of diabetic retinopathy for T1D patients [2]. Dead-in-bed syndrome is the worst complication of severe hypoglycemia [3]. More than 50% of severe hypoglycemia episodes at night time make the patients especially fear nocturnal hypoglycemia [4, 5]. It is founded that repeated and severe hypoglycemia can increase the death risk of diabetes [6, 7]. Therefore, an effective hypoglycemia early alarm method is an urgent problem in blood glucose management for reducing and avoiding hypoglycemia events [8, 9].

In recent years, many methods have been proposed for hypoglycemia early alarm [10]. Palerm et al. [11] used real-time blood glucose sensor signals and Kalman filtering to predict hypoglycemia. Cameron et al. [12] employed multiple statistical linear predictions with regression windows as a hypoglycemia detection algorithm. Turksoy et al. [13] introduced a subject-specific recursive linear time series model to capture blood glucose variations and used it in hypoglycemia early alarm systems. Bayrak et al. [14] employed the recursive autoregressive partial least squares algorithm to model the continuous glucose monitoring (CGM) data and predict future blood glucose in hypoglycemia early alarm systems. Yang et al. [15] proposed a prediction framework by the autoregressive moving average model with an identification algorithm. The hypoglycemia early alarm methods depend on the accuracy of the prediction model. Because blood glucose is affected by many factors, the prediction model is difficult to accurately predict especially for hypoglycemia.

The sequence pattern mining is to discover frequent patterns contained in sequences, which are widely used in customer purchasing, weather forecasts, and production processes [16, 17]. Aileen P et al. [18] applied the CSPADE algorithm to mine sequence patterns of diabetes medication prescriptions. And ranked the drug class and generic drug level by the support statistic. W Lee et al. [19] proposed a sequential pattern mining-based framework–FuzzyGap for extracting dis-

* Corresponding author.
E-mail address: yuxia@ise.neu.edu.cn (X. Yu).

https://doi.org/10.1016/j.heliyon.2022.e11372
Received 30 June 2022; Received in revised form 12 August 2022; Accepted 26 October 2022

criminative and representative clinical pathways from EHRs. Wu et al. [20] applied a complete algorithm based on the Net-tree structure to mine the closed patterns in SARS-CoV-2 and SARS virus. Sequential pattern mining predictions by mining frequent sequence patterns in events avoid the inherent and undeniable uncertainty of traditional prediction methods. However, single-dimensional sequential pattern mining does not consider the potential impact of multi-dimensional data correlation.

Multi-dimensional sequential pattern mining is proposed by Han et al. [21, 22, 23], which aims to discover more relevant and valuable patterns by considering the correlation between multi-dimensional data. Petitjean et al. [24] proposed a mining framework for discovering satellite image time series patterns. Yusof et al. [25] presented a novel approach for mapping frequent wind profile patterns using multi-dimensional sequential pattern mining and identified four frequent wind profile patterns. Sakuma et al. [26] developed a method for extracting interesting animal behaviors from multi-dimensional time series and applied it to several animal trajectory datasets for demonstrating effectiveness. Reasonable use of more information can provide a more meaningful reference for sequence pattern mining and obtain better application value.

Recent studies have revealed that blood glucose is affected by many factors [27, 28, 29] and can be indirectly reflected through electroencephalogram [30], heart rate [31], etc. This paper proposed a multi-dimensional hypoglycemia early alarm method based on the idea of sequence pattern mining. It can avoid the inherent uncertainty of the prediction model and extract more valuable blood glucose information. First, the blood glucose sequence is integrated with related factors (such as meal, insulin, exercise, etc.) to construct a multi-dimensional sequence database. Secondly, the UniSeq algorithm is employed to mine hypoglycemic frequent sequence patterns for constructing the multi-dimensional sequential pattern library. Finally, real-time blood glucose is matched through pattern matching to realize hypoglycemia early alarm. A multi-dimensional blood glucose sequence that contains more information can extract potential information more comprehensively. The consideration of multiple factors is conducive to individual pattern mining and better reflects the individual patients' characteristic. Compared with the prediction model, it avoids prediction errors and provides comprehensive diagnosis support for decision-makers.

Based on the above ideas, the main content of this paper is shown as follows: Section 2 introduces the method of multi-dimensional sequential pattern mining; Section 3 describes multi-dimensional hypoglycemia early alarm; Section 4 shows the results and Section 5 analyzes the method performance; Section 6 finally gives the conclusion.

2. Material and methods

2.1. Multi-dimensional sequential pattern mining

The multi-dimensional sequence database consists of a base task dimension and task-related dimensions. The base task dimension is one or more ordered information dimensions that express the progress of a transaction over time. The task-relevant dimensions are one or more unordered information dimensions that provide background information. Data is recorded as the schema $(RID, S, A_1, \cdots, A_m)$ in a multi-dimensional sequence database [32], where RID is the primary key; $A_1, \cdots, A_m$ are dimensions and S is in the domain of sequences. $(s, a_1, a_2, \cdots, a_m)$ is defined as a multi-dimensional sequence, where $a_i \in (A_i \cup \{*\})$ $(1 \le i \le m)$, s is a sequence.

Definition 1. A multi-dimensional sequence $P = (s, a_1, \cdots, a_m)$ is said to match a tuple $t = (s_t, x_1, x_m)$ in the multi-dimensional sequence database if and only if, for $a_i = x_i$ (or $a_i = *$) and $s \subseteq s_t$ $(1 \le i \le m)$. The number of tuples in the database matching multi-dimensional sequence P is called the support of P, denoted as $support(P)$. Given a minimum support threshold of $min\_support$, a multi-dimensional sequence P is called a multi-dimensional sequential pattern if and only if $support(P) \ge min\_support$.

Definition 2. The multi-dimensional information $M = (a_1, \cdots, a_m)$ in the multi-dimensional sequence pattern $P = (s, a_1, \cdots, a_m)$ is called multi-dimensional patterns or MD-patterns. If $a_1, \cdots, a_m$ consists $(n \le m)$ of which number is n instead of $*$, M is called n-dimensional patterns. There are i-dimensional patterns $M_i = (a_1, \cdots, a_m)$ and j-dimensional patterns $M_j = (b_1, \cdots, b_m)$. If and only if $a_k = b_k$ for all a_k $(1 \le k \le m)$ not $*$, M_i is a sub-pattern of M_j and M_j is a super-pattern of M_i.

Multi-dimensional sequential pattern mining refers to the mining of one or more disordered information dimensions and an ordered information dimension. UniSeq algorithm [33], as a multi-dimensional sequential pattern mining, merges multi-dimensional information into the original sequence database to form a multi-dimensional sequence database. The multi-dimensional information can be embedded in the first or last element in the extended sequence. The PrefixSpan [34] which is based on the idea of pattern growth is a classic sequential pattern mining algorithm. It uses different prefixes to project the target dataset and performs sequence pattern mining on the obtained data subset. Then, PrefixSpan is used to mine the multi-dimensional sequence database. Based on the PrefixSpan, UniSeq does not require the cost of data structure conversion, and mining process and has higher efficiency when the dimensionality is low.

2.2. Sequence pattern matching

Definition 3. Suppose a longest common subsequence $Z = \langle z_1, z_2, \cdots, z_k \rangle$ of sequence $X = \langle x_1, x_2, \cdots, x_m \rangle$ and $Y = \langle y_1, y_2, \cdots, y_n \rangle$: if $x_m = y_m$, then $z_k = x_m = y_m$ and Z_{K-1} is the longest common subsequence of X_{m-1} and Y_{n-1}; if $x_m \ne y_m$ and $z_k \ne x_m$, then Z is the longest common subsequence of X_{m-1} and Y; if $x_m \ne y_m$ and $z_k \ne y_m$, then Z is the longest common subsequence of X and Y_{n-1}; among them $X_{m-1} = \langle x_1, x_2, \cdots, x_{m-1} \rangle$, $Y_{n-1} = \langle y_1, y_2, \cdots, y_{n-1} \rangle$, and $Z_{k-1} = \langle z_1, z_2, \cdots, z_{k-1} \rangle$.

The longest common subsequence algorithm (LCSS) [35] does not change the sequence order and obtains a new sequence by removing elements in the sequence. Exhaustive methods and dynamic programming algorithms can be used for LCSS solution. Dynamic programming first finds the optimal solution of the sub-problem, then constructs the optimal solution of the original problem. Establish the two-dimensional array $C[i, j]$ as the longest common subsequence of record sequence X and Y. When $i = 0$ or $j = 0$, the empty sequence is the longest common subsequence of X and Y, so $C[i, j] = 0$. In other cases, the recursive relationship can be established by the theorem as equation (1):

$$C[i,j] = \begin{cases} 0 & i = 0 \text{ or } j = 0 \\ C[i-1, j-1] + 1 & i, j > 0 \text{ and } x_i = y_j \\ \max\left(C[i, j-1], C[i-1, j]\right) & i, j > 0 \text{ and } x_i \ne y_j \end{cases} \quad (1)$$

By recursion from the bottom right corner of the matrix achieves the solution of all the longest common subsequences.

3. Multi-dimensional hypoglycemia early alarm

3.1. Data preprocessing

3.1.1. Missing data filling

Here we use the OhioT1DM Dataset [36] including the training set and test set data of 12 T1D patients. The information in the dataset is from a CSII-CGM therapy and a fitness tracker band. Because the data has outliers and missing etc., it needs to be preprocessed. The outliers in CGM use Gaussian process regression to detect and correct. The basal insulin dose has daily periodicity, so the data of the previous day is applied for filling. The first-order Taylor series extrapolation method

Table 1. Symbolic mapping rules for blood glucose.

Range of Blood Glucose	Symbolic Representation	Range of Blood Glucose	Symbolic Representation
(70,75]	'a'	(120,130]	'h'
(75,80]	'b'	(130,140]	'i'
(80,85]	'c'	(140,150]	'j'
(85,90]	'd'	(150,200]	'k'
(90.100]	'e'	(200,300]	'l'
(100,110]	'f'	(300,400]	'm'
(110,120]	'g'	[40,70]	'z'

and historical average are used for filling when the consecutive missing values are less than 12. The missing CGM in the training dataset is not filled to avoid introducing additional noise [37]. The training set is used to mine the hypoglycemia sequential pattern library and the test set verifies the performance of the hypoglycemia early alarm method.

3.1.2. Physiological information conversion

Insulin can be divided into basal and high-dose insulin for T1D patients. The physiological information conversion model is introduced for reflecting the insulin effect on blood glucose.

The insulin remaining active within the body can be represented using a two-compartment model that estimates the insulin on board (IOB)[38] as equation (2). The IOB is an estimation of the residual insulin accumulated in the subcutaneous tissue.

$$\begin{aligned} \frac{dC_1(t)}{dt} &= u(t) - K_{DIA}C_1(t) \\ \frac{dC_2(t)}{dt} &= K_{DIA}\left(C_1(t) - C_2(t)\right) \\ IOB(t) &= C_1(t) + C_2(t) \end{aligned} \tag{2}$$

where t is the time instant, the compartments C_1, and C_2 insulin mass (mU) in the accessible and non-accessible subcutaneous compartments, and $u(t)$ (mU $\min^{-1}$) is the insulin dose. $K_{DIA} = 0.0195$ ($\min^{-1}$) is a constant related to the duration of insulin action (DIA), which characterizes the patient's insulin activity dynamics.

Carbohydrates on board (COB) represents the remaining CHO amount of a meal that has not yet appeared in the blood glucose. It is an extension of the model which describes the appearance rate (R_a) of glucose in the blood due to CHO intake[39] as equation (3).

$$\begin{aligned} R_a(t) &= \frac{C_{in}C_{bio}te^{(-t/t_{\max})}}{t_{\max}^2} \\ COB(t) &= C_{in}C_{bio} - \int_{t_{meal}}^{t} R_a(t)dt \end{aligned} \tag{3}$$

where C_{in} is the amount of CHO ingested and $C_{bio} = 0.8$ is the bioavailability. $t_{\max} = 60$ denotes the maximum appearance rate time of glucose in the accessible glucose compartment and t_{meal} is the time instant in which a meal is consumed.

3.1.3. Symbolization of multi-dimensional blood glucose data

Multi-dimensional hypoglycemia early alarm aims to discover frequently sequence patterns from the blood glucose time series. The variable sequence in a continuous numerical form is not easy to describe and search. According to the characteristics of the blood glucose time series, the symbol space is divided into different variables based on the value. Finally, the data sequence is discretized into a sequence composed of several distinct symbols. The "coarse-grained" can capture large-scale features and reduce the impact of measurement noise. It is more conducive to mining the hidden sequence patterns for the blood glucose time series. The symbolic mapping rules for blood glucose are shown in Table 1. And the symbolic mapping of IOB and COB is similar to that of blood glucose.

Table 2. Multi-dimensional hypoglycemic sequence database.

RID (Hypoglycemia number)	IOB	COB	Sequence
1	A	M	ihgfeaz
...	...	...	...
k	D	Z	ffeecba

Table 3. Multi-dimensional hypoglycemic sequence database SDB^{md}.

RID (Hypoglycemia number)	Extended sequence
1	$\langle(A, M)(ihgfeaz)\rangle$
...	...
k	$\langle(D, Z)(ffeecba)\rangle$

3.2. Definition of hypoglycemia early alarm problem

The UniSeq algorithm is applied to realize multi-dimensional hypoglycemia sequential pattern mining and the longest common subsequence is used to match the real-time multi-dimensional blood glucose. The related concepts of multi-dimensional hypoglycemia sequential pattern and early alarm sequence are defined as shown in Fig. 1.

Definition 4. An alarm sequence. According to the alarm rule (the blood glucose ≤70 mg/dL [40]), the alarm sequence is determined as the sequence index $S_i \sim (S_i + L_w)$.

Definition 5. Early alarm sequence. Given the length L_e of the fixed window, the early alarm sequence is the blood glucose before the alarm sequence. That is, the corresponding value in the sequence index $(S_i - L_e) \sim S_i$ composes the early alarm sequence.

Definition 6. Non-alarm sequence. Non-alarm sequence refers to the blood glucose sequence between the alarm sequence and the next alarm sequence, defined as $(S_{i-1} + L_p) \sim (S_i - L_e)$.

Definition 7. Multi-dimensional hypoglycemia sequential pattern. The corresponding timestamp IOB and COB merge into the blood glucose sequence as the multi-dimensional hypoglycemia sequential pattern.

3.3. Hypoglycemia early alarm based on multi-dimensional sequence mining

The overall process of hypoglycemia early alarm based on multi-dimensional sequential pattern mining is shown in Fig. 2. The main steps are listed as follows and the algorithm 1.

Step 1. In the original time series, the alarm event is screened out according to the blood glucose threshold, and the dataset is divided into multi-dimensional early alarm/non-alarm sequence set. To ensure the early alarm/non-alarm sequences are in the same sequence length for sequence pattern mining, the non-alarm sequence needs to be processed into subsequences with the same length as the early alarm sequence.

Step 2. Symbolize the database and shown in Table 2. The hypoglycemia number RID is used to mark each sequence. IOB and COB are two dimensions. $a, b, \cdots, h$ is the blood glucose value, $A, B, \cdots, H$ is IOB, $M, N, \cdots, Z$ is COB, and sequence is used to record the historical value before hypoglycemia.

Step 3. The multi-dimensional information is embedded as the first element in the extended sequence in converting the multi-dimensional sequence database, as shown in Table 3. The number of non-alarm

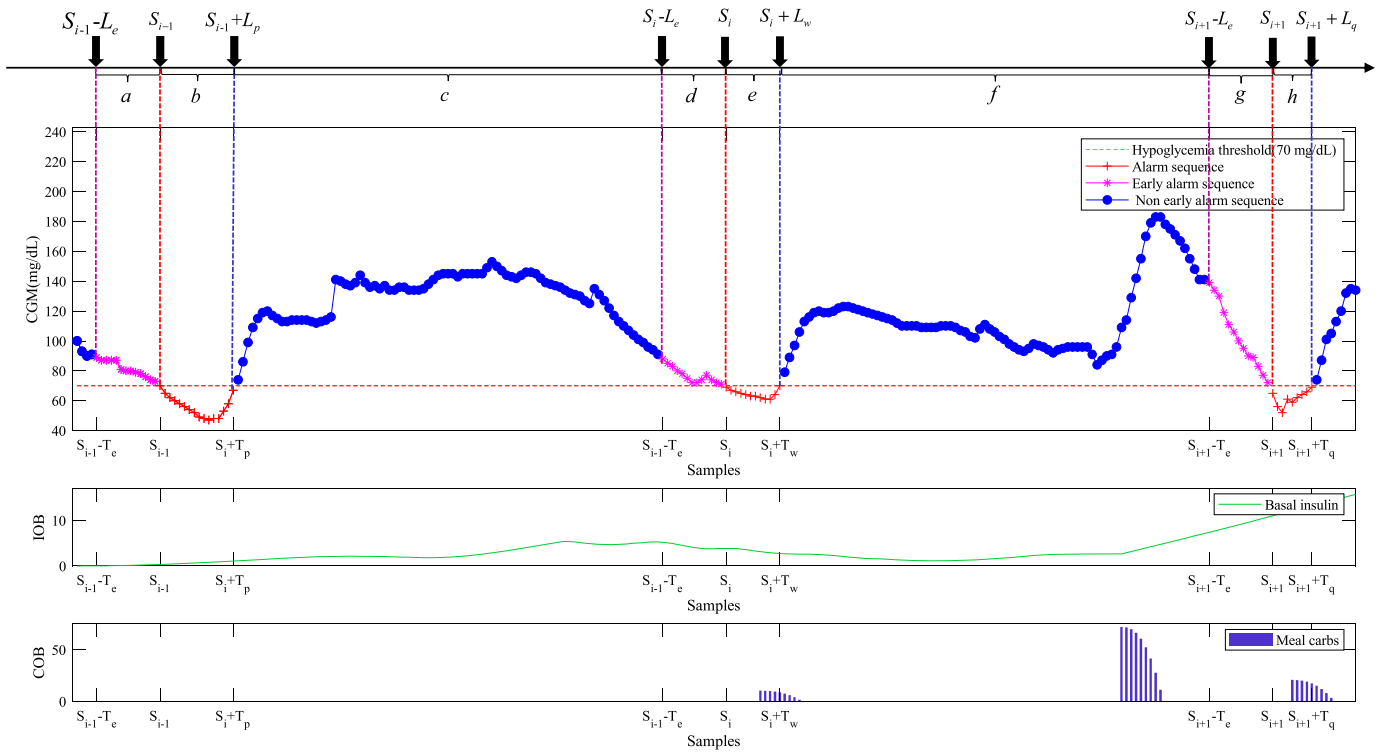


Fig. 1. Multi-dimensional hypoglycemia early alarm/non-alarm sequence (where a, d, g are early alarm number $i-1$, i, $i+1$; b, e, h are alarm number $i-1$, i, $i+1$; c, f are non-alarm number i, $i+1$)

sequence set is much larger than that of early alarm. So the support threshold setting will be different for the alarm and non-alarm sequence.

Step 4. Set the support threshold and sequence length for the algorithm, and use the Prefix Span algorithm to find out the frequent sequences of all values when scanning the database for the first time. Then, find the single frequent sequence in the projection library, and obtain the preliminary multi-dimensional hypoglycemia sequential pattern library through continuous regression mining.

Step 5. Use the longest common subsequence algorithm to delete frequent subsequences that are lower than the minimum length threshold from the early alarm/non-alarm sequence in the multi-dimensional hypoglycemia sequential pattern library.

Step 6. The longest common subsequence algorithm is used to remove the common frequent sequence in the hypoglycemia early alarm sequential pattern library. The redundant patterns are eliminated to complete the construction of multi-dimensional hypoglycemia early alarm sequential pattern library.

Step 7. Real-time blood glucose uses the same slide window size as the early alarm sequence, and symbolizes the real-time multi-dimensional blood glucose.

Step 8. The longest common subsequence algorithm matches the real-time multi-dimensional blood glucose sequence with the multi-dimensional hypoglycemia early alarm sequential pattern library. If the matching is successful, a hypoglycemia early alarm signal is issued.

Algorithm 1. Early alarm of hypoglycemia sequence mode based on UniSeq algorithm.

Input: Early alarm sequence set S obtained by dynamic sliding window, minimum support is α

Output: low blood sugar warning label 0 (non-hypoglycemia) or 1 (hypoglycemia)

Find all prefixes of length and the corresponding projection database HyD;

$i = 1$: Obtain the l-item frequent sequence whose support degree is less than the threshold value, and put it into the dataset FS;

For recursive mining for each $prefix_i$ in FS:

  Find the suffix projection database HyD_i corresponding to $prefix_i$

  If HyD_i is empty:

    End the current cycle, return to continue the next cycle

  For $event_i$ in HyD_i:

    Calculate the number of support $c_i = count(event_i)$

  If $c_i < \alpha$:

    End the current cycle, return to continue the next cycle

  Combine items that meet the support number with the current prefix to update FS

  $i = i + 1$

End for

For j in length (CGM):

  Current blood glucose sequence s obtained by sliding window s perform physiological information conversion and symbolic conversion

  Use the longest common subsequence s to match with FS

  Output alarm label 0, 1

  $j = j + 1$

End for

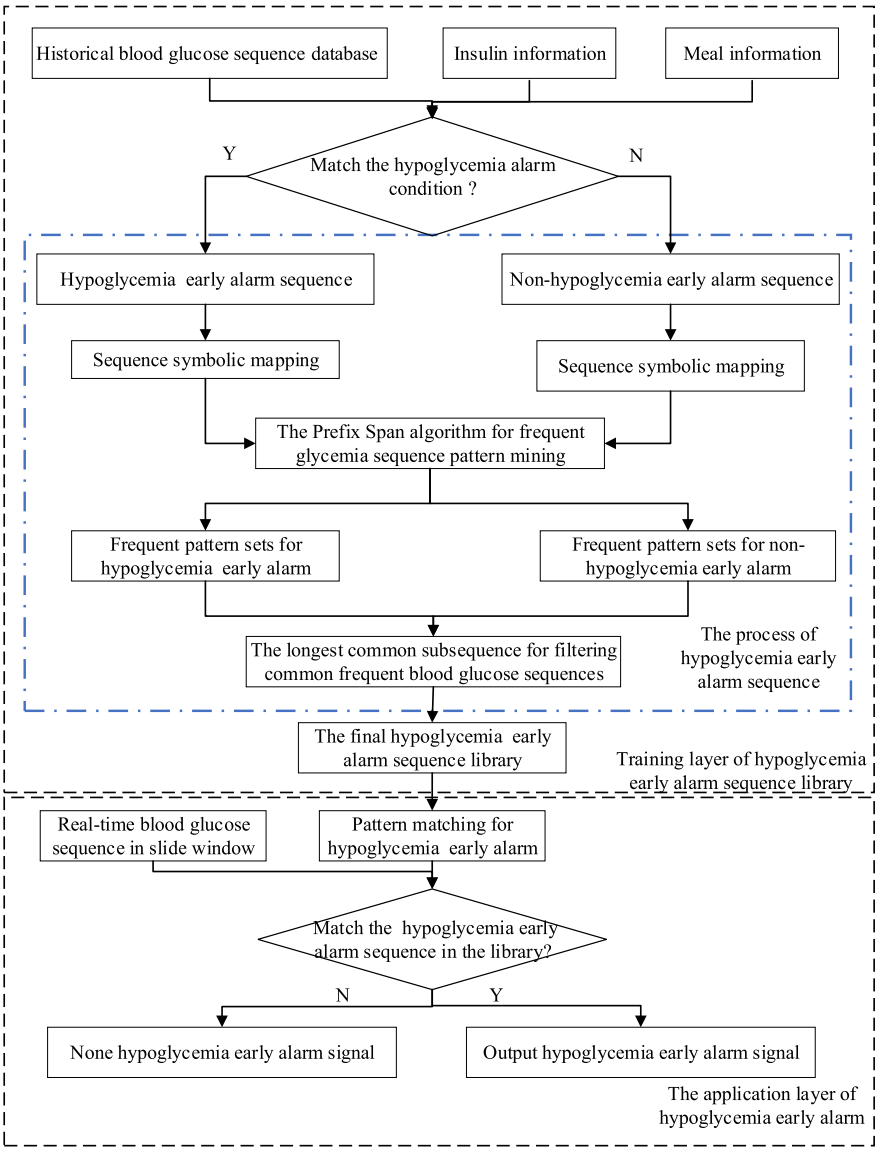


Fig. 2. Flow chart of multi-dimensional hypoglycemia early alarm.

4. Experimental results and discussion

4.1. Evaluation indicators

In the paper, we use sensitivity, precision, F1 score, and early alarm time to evaluate the method's performance. The true positive (TP) represents the number of cases predicted correctly to have a hypoglycemia alarm and the true negative (TN) indicates the number of cases predicted correctly to not have a hypoglycemia alarm. The false positive (FP) is the number of cases predicted falsely to have a hypoglycemia alarm and the false negative (FN)is the number of cases predicted incorrectly to have a hypoglycemia alarm. The time of hypoglycemia is defined as T_h, and the time of the early alarm sequence endpoint is defined as T_{end}. And the formulas are defined as equations (4)–(7):

$$\text{Sensitivity} = \frac{TP}{TP + FN} * 100\% \tag{4}$$

$$\text{Precision} = \frac{TP}{TP + FP} * 100\% \tag{5}$$

$$\text{F1 score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{6}$$

$$\text{Early Alarm Time} = T_h - T_{end} \tag{7}$$

4.2. Results

4.2.1. Analysis of hypoglycemia early alarm method factors

The support threshold and sequence length are key indicators for the hypoglycemia early alarm based on sequence pattern. To assess the relationship between support threshold, sequence length, and the method performance, we selected Subject 540, 552, and 567 for verification. Set the support threshold between 0.2–0.4, and the sequence length between 10–14. The influence on the hypoglycemia early alarm is shown in Fig. 3 and Table 4.

Fig. 3 clearly illustrated that a few sequence patterns are mined at the support 0.3, while the "explosion mode" occurs at the support 0.1. The higher the support threshold is, the fewer frequent sequence patterns are mined, and some low-frequency sequence patterns are ignored. As shown in Table 4, the sequence length of Subject 540 increased the early alarm, false alarm, and average early alarm time. Subjects 540, 552, and 567 had a better performance with sequence lengths 10, 12, and 12 respectively. The introduced blood glucose information could improve the early alarm and reduce the false alarm. Too much ineffective blood glucose information will dilute the effective early alarm information while leaving only high-frequency early alarm information will lead to many missed alarms. Thus, the support thresh-

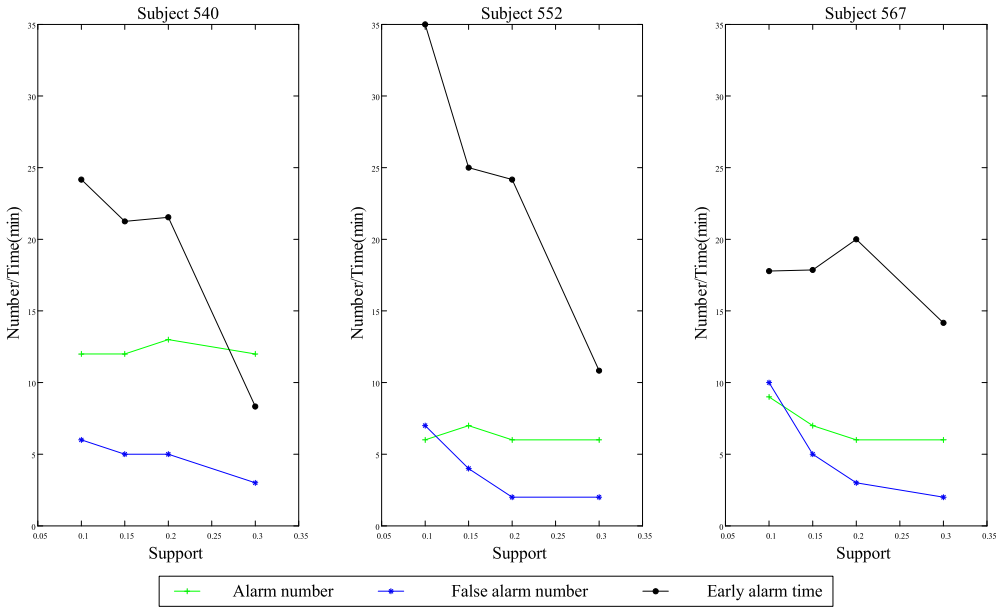


Fig. 3. Hypoglycemia early alarm with different support threshold (sequence length 10).

Table 4. Hypoglycemia early alarm with different sequence length (support 0.2).

Research Object	Subject 540			Subject 552			Subject 567		
Sequence Length	Alarm	False	Time	Alarm	False	Time	Alarm	False	Time
10	13	5	21.54	6	2	24.17	6	6	20
12	12	3	18.75	7	4	35	8	6	21.86
14	11	2	6.82	6	2	17.5	6	4	15.83

In the table, alarm represents early alarm; false is a false alarm; time means early alarm time.

Table 5. Comparison of single and multi-dimension hypoglycemia early alarm.

Method	Sensitivity	Precision	F1 score	Early alarm time
Single-Dimension	70%	70%	70%	20.81
Multi-Dimension (IOB)	71.84%	74%	72.91%	19.54
Multi-Dimension (IOB, COB)	75.76%	75%	75.38%	25.17

old and sequence length need to be set according to the size and specific conditions of the subjects.

Meanwhile, we expanded the CGM database to verify the influence on the hypoglycemia early alarm. Keep the same sequence length and support threshold for the subjects, as shown in Figs. 4 (a) and (b).

As illustrated in Fig. 4 (a) that the overall fluctuations of Subjects 540, 552, and 563 are relatively smaller than other subjects. It can be seen from Fig. 4 (b) that Subject 591 fluctuated relatively small compared with other samples. The hypoglycemia early alarm has been greatly improved for Subject 584 and there is no change in other indicators. For Subject 540, 552, and 563, the expanded library has improved the false alarm. The CGM database expansion can improve the method's performance. Some subjects have worse performance due to the differences between subjects.

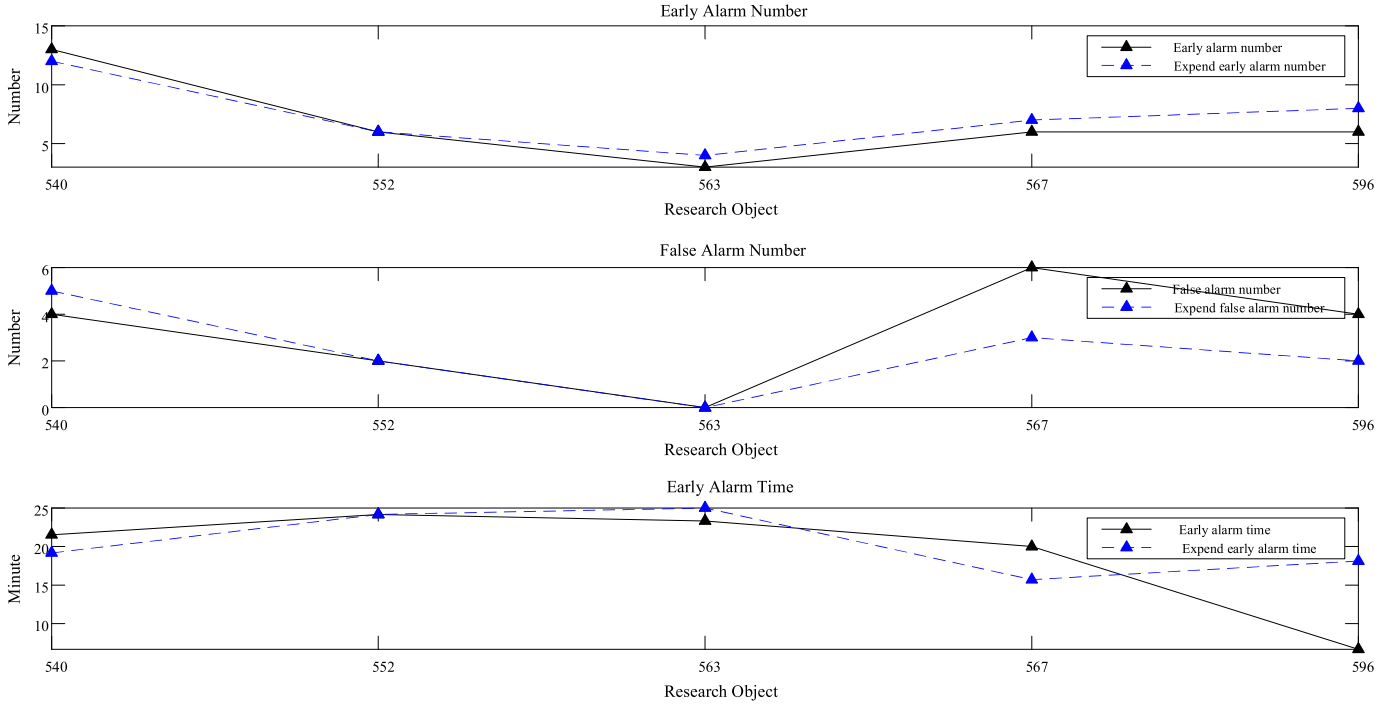
4.2.2. Multi-dimensional sequential pattern mining of hypoglycemia early alarm

Compare single and multi-dimensional sequential pattern mining with the OhioT1DM Dataset. Based on the analysis of the support threshold and sequence length, set the support threshold as 0.2, 0.15, and the sequence length as 12. The multi-dimensional hypoglycemia early alarm for Subject 540 is shown in Fig. 5, and the average result is shown in Table 5.

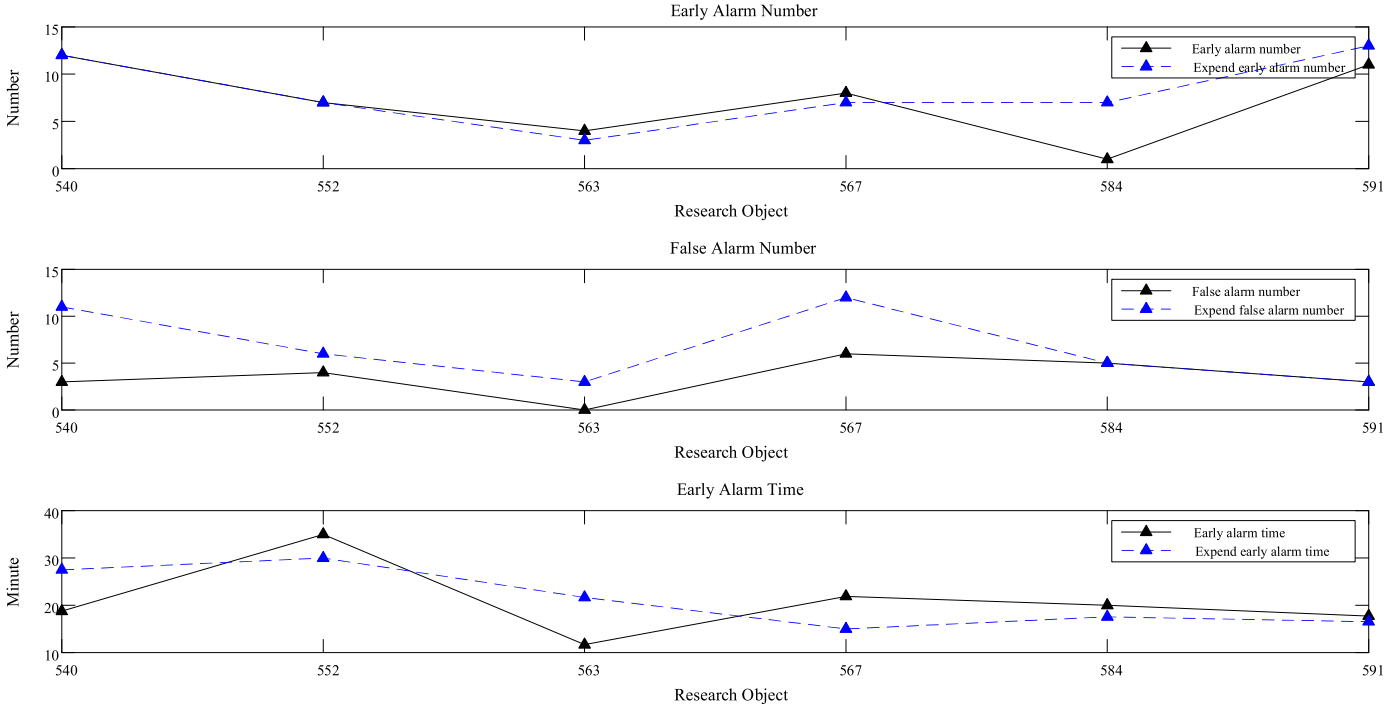
Fig. 5 clearly illustrated that the multi-dimensional hypoglycemia early alarm for Subject 540 can provide early alarm of steadily and rapidly hypoglycemia events. Due to the frequent blood glucose fluctuations, there is a missing alarm between 1500–2000 steps. But it can early alarm potential serious hypoglycemia events in time, which is beneficial to take remedial measures. As shown in Table 5, the sensitivity and specificity of the multi-dimensional early alarm with IOB improved compared with the single. Although the early alarm time declined, it is still possible to give an appropriate early alarm. Secondly, the multi-dimensional early alarm with IOB and COB is better than the single-dimensional in all evaluation indicators, especially in the early alarm.

5. Discussion

Compared to the multi-dimensional early alarm with IOB, the accuracy of the early alarm improved slightly, but the early alarm time improved greatly. The multi-dimensional sequential pattern mining can dig out some potential knowledge information, and improve the method ability in hypoglycemia early alarm. The performance has been improved by merging food and insulin information. However, there are certain differences in the effect shown for different patients. Using only the training set data of the research object itself can better extract the patient's multi-dimensional characteristics and avoid interference between different research objects. Due to the large difference between the training set and the test set, the lack of abundant early alarm sequence patterns is one reason for the high rate of the missing alarm.



(a)



(b)

Fig. 4. Hypoglycemia early alarm with database expansion (a: support 0.2, length 10; b: support 0.2, length 12).

6. Conclusions

In this paper, a multi-dimensional hypoglycemia sequential pattern mining method is proposed for hypoglycemia early alarm. Compared with the single-dimensional sequential pattern mining, the multi-dimensional method can better mine the potential information of glucose dynamics, and achieve improved detection performance of hypoglycemia events. The experiment results show that the proposed multi-dimensional sequential pattern mining can predict future hypoglycemia events with relatively high accuracy in a short time, and could provide comprehensive diagnosis support for decision marking. Our future work will concentrate to develop a personalized model for hypoglycemia detection to reduce the negative effect of specificity. And conduct a cluster analysis of related research objects for achieving better data expansion to improve the indicators.

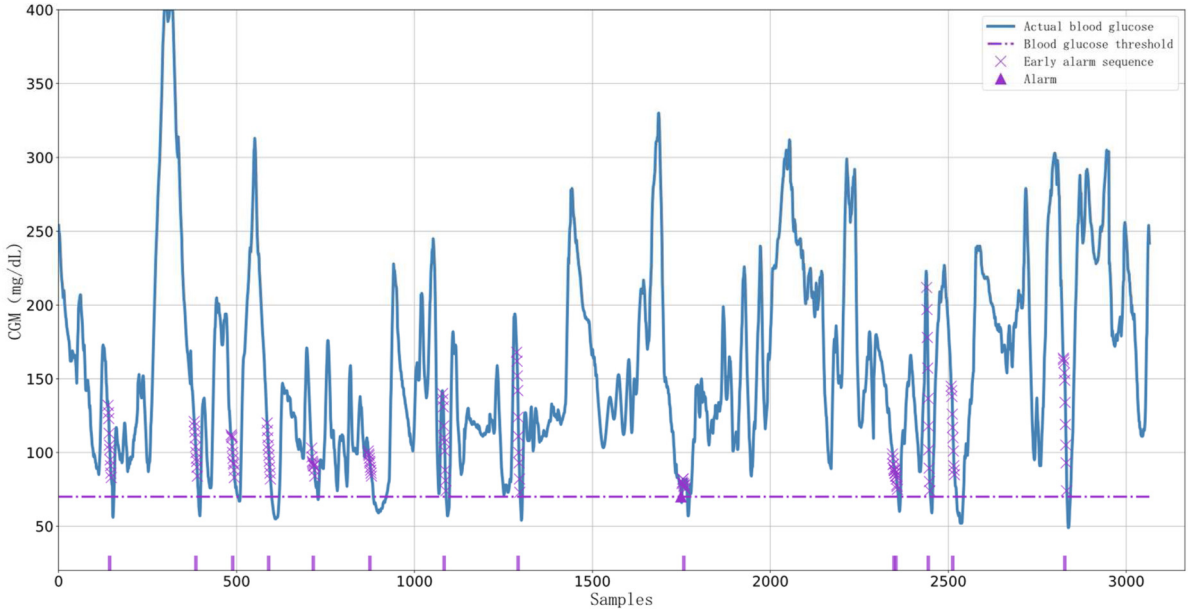


Fig. 5. Multi-dimensional hypoglycemia early alarm with IOB and COB for subject 540.

Declarations

Author contribution statement

Ning Ma: Conceived and designed the experiments; Performed the experiments; Analyzed and interpreted the data; Wrote the paper. **Xia Yu:** Conceived and designed the experiments; Contributed reagents, materials, analysis tools or data; Wrote the paper. **Tao Yang, Yuhang Zhao:** Performed the experiments; Analyzed and interpreted the data. **Hongru Li:** Contributed reagents, materials, analysis tools or data.

Funding statement

This work was supported by National Natural Science Foundation of China (61903071 & 61973067) and Foundation Research Funds for the Central Universities (N2104031).

Data availability statement

Data associated with this study has been deposited at OhioT1DM Dataset under http://smarthealth.cs.ohio.edu/OhioT1DMdataset.html.

Declaration of interests statement

The authors declare no conflict of interest.

Additional information

No additional information is available for this paper.

References

[1] G. Zvilichovsky, M. David, Definition, classification, and diagnosis of diabetes mellitus, Exp. Clin. Endocrinol. Diabetes 126 (07) (2018) 406–410.
[2] C. Ning, M. Paul, W. Yin, Diabetic Retinopathy, Lancet 376 (9735) (2010) 124–136.
[3] D.C. Klonoff, The need for hypoglycemia detection and prevention in type 1 diabetes, Diabetes Technol. Ther. 3 (4) (2001) 567–570.
[4] Epidemiology of severe hypoglycemia in the diabetes control and complications trial, Am. J. Med. (1991).
[5] A.J. Graveling, B.M. Frier, The risks of nocturnal hypoglycaemia in insulin-treated diabetes, Diabetes Res. Clin. Pract. 133 (2017) 30–39.
[6] S.A. Amiel, P. Aschner, Hypoglycaemia, cardiovascular disease, and mortality in diabetes: epidemiology, pathogenesis, and management, Lancet Diabetes Endocrinol. 7 (5) (May 2019) 385–396.
[7] A. Amer Diabet, Older adults: standards of medical care in diabetes-2020, Diabetes Care 43 (Jan 2020) S152–S162.
[8] E. Bekiari, K. Kitsios, Artificial pancreas treatment for outpatients with type 1 diabetes: systematic review and meta-analysis, BMJ, Br. Med. J. 361 (2018).
[9] S. Ma, P.J. Schreiner, E.R. Seaquist, M. Ugurbil, R. Zmora, L.S. Chow, Multiple predictively equivalent risk models for handling missing data at time of prediction: with an application in severe hypoglycemia risk prediction for type 2 diabetes, J. Biomed. Inform. 103 (2020).
[10] Y.K. Lin, G. Danielle, O. Chan, et al., Alarm settings of continuous glucose monitoring systems and associations to glucose outcomes in type 1 diabetes, J. Endocr. Soc. 1 (1) (2019) 1.
[11] C.C. Palerm, B.W. Bequette, Hypoglycemia detection and prediction using continuous glucose monitoring–a study on hypoglycemic clamp data, J. Diabetes Sci. Technol. 1 (5) (2007) 624–629.
[12] F. Cameron, G. Niemeyer, K. Gundy-Burlet, B. Buckingham, Statistical hypoglycemia prediction, J. Diabetes Sci. Technol. 2 (4) (2008) 612–621.
[13] K. Turksoy, E.S. Bayrak, L. Quinn, E. Littlejohn, D. Rollins, A. Cinar, Hypoglycemia early alarm systems based on multivariable models, Ind. Eng. Chem. Res. 52 (35) (2013) 12329–12336.
[14] E.S. Bayrak, K. Turksoy, A. Cinar, L. Quinn, E. Littlejohn, D. Rollins, Hypoglycemia early alarm systems based on recursive autoregressive partial least squares models, J. Diabetes Sci. Technol. 7 (1) (2013).
[15] Y. Jun, L. Lei, S. Yimeng, X. Xiaolei, An ARIMA model with adaptive orders for predicting blood glucose concentrations and hypoglycemia, IEEE J. Biomed. Health Inform. 23 (3) (2019) 1251–1260.
[16] L. Li, X. Wang, X. Huang, G. Wang, Y. Sun, Enterprise lean catering material management information system based on sequence pattern data mining, in: 2018 IEEE 4th International Conference on Computer and Communications (ICCC), 2018.
[17] A.A. Ramaki, M. Amini, R.E. Atani, RTECA: real time episode correlation algorithm for multi-step attack scenarios detection, Comput. Secur. 49 (2015) 206–219.
[18] A.P. Wright, A.T. Wright, A.B. Mccoy, D.F. Sittig, The use of sequential pattern mining to predict next prescribed medications, J. Biomed. Inform. 53 (2015).
[19] E.W. Lee, J.C. Ho, FuzzyGap: sequential pattern mining for predicting chronic heart failure in clinical pathways, AMIA Jt. Summits Transl. Sci. Proc. 2019 (2019) 222–231 [Online]. Available: https://www.ncbi.nlm.nih.gov/pubmed/31258974.
[20] Y. Wu, C. Zhu, Y. Li, L. Guo, X. Wu, NetNCSP: nonoverlapping closed sequential pattern mining, Knowl.-Based Syst. 196 (2020) 105812.
[21] R.C. Xiao, Efficient algorithm of mining multi-dimensional sequential patterns, Comput. Eng. Appl. 44 (6) (2008) 187–190.
[22] J.W. Han, J. Pei, X.F. Yan, From sequential pattern mining to structured pattern mining: a pattern-growth approach, J. Comput. Sci. Technol. 19 (3) (2004) 257–279.
[23] H. Pinto, J. Han, P. Jian, W. Ke, U. Dayal, Multi-dimensional sequential pattern mining, in: Proceedings of the 2001 ACM CIKM International Conference on Information and Knowledge Management, Atlanta, Georgia, USA, November 5-10, 2001, 2001.
[24] F. Petitjean, F. Masseglia, P. Gancarski, G. Forestier, Discovering significant evolution patterns from satellite image time series, Int. J. Neural Syst. 21 (6) (2011) 475–489.
[25] N. Yusof, R. Zurita-Milla, Mapping frequent spatio-temporal wind profile patterns using multi-dimensional sequential pattern mining, Int. J. Digit. Earth 10 (3) (2017) 238–256.

[26] T. Sakuma, et al., Finding discriminative animal behaviors from sequential biologging trajectory data, in: N. Streitz, S. Konomi (Eds.), Distributed, Ambient and Pervasive Interactions: Technologies and Contexts, Dapi 2018, Pt Ii, in: Lecture Notes in Computer Science, vol. 10922, 2018, pp. 125–138.
[27] R. Reddy, N. Resalat, L.M. Wilson, J.R. Castle, J.E. Youssef, P.G. Jacobs, Prediction of hypoglycemia during aerobic exercise in adults with type 1 diabetes, J. Diabetes Sci. Technol. 13 (5) (2019).
[28] A.B. Evert, et al., Nutrition therapy for adults with diabetes or prediabetes: a consensus report, Diabetes Care 42 (5) (2019) 731–754.
[29] F. Imamura, et al., Effects of saturated fat, polyunsaturated fat, monounsaturated fat, and carbohydrate on glucose-insulin homeostasis: a systematic review and meta-analysis of randomised controlled feeding trials, PLoS Med. 13 (7) (2016).
[30] L.S.N. Snogdal, et al., Detection of hypoglycemia associated EEG changes during sleep in type 1 diabetes mellitus, Diabetes Res. Clin. Pract. 98 (1) (2012).
[31] M.D. Breton, S.A. Brown, C.H. Karvetski, L. Kollar, B.P. Kovatchev, Adding heart rate signal to a control-to-range artificial pancreas system improves the protection against hypoglycemia during exercise in type 1 diabetes, Diabetes Technol. Ther. 16 (8) (2014) 506.
[32] D.Q. Luan, X.S. Qin, Application of algorithm PrefixSpan in multi-dimensional sequential pattern mining, Microcomput. Develop. (2003).
[33] J. Yang, W. Zuo, Multi-dimensional sequential pattern mining based on concept lattice, in: Advanced Data Mining and Applications, Second International Conference ADMA 2006, Xi'an, China, August 14–16, 2006, Proceedings, 2006.
[34] J. Pei, J. Han, B. Mortazavi-Asl, H. Pinto, M. Hsu, PrefixSpan: mining sequential patterns efficiently by prefix-projected pattern growth, in: International Conference on Data Engineering, 2001.
[35] M. Vlachos, G. Kollios, D. Gunopulos, Discovering similar multidimensional trajectories, in: Proceedings 18th International Conference on Data Engineering, 2002.
[36] C. Marling, R. Bunescu, The OhioT1DM Dataset for Blood Glucose Level Prediction: Update 2020, presented at the Proceedings of the 5th International Workshop on Knowledge Discovery in Healthcare Data, Santiago de Compostela, Spain & Virtually, August 29–30, 2020, presented at the.
[37] T. Yang, et al., Multi-Scale Long Short-Term Memory Network with Multi-Lag Structure for Blood Glucose Prediction, presented at the Proceeding of 5th International Workshop on Knowledge Discovery in Healthcare Data, Santiago de Compostela, Spain & Virtually, August 29–30, 2020, presented at the.
[38] M.E. Wilinska, L.J. Chassin, H.C. Schaller, L. Schaupp, R. Hovorka, Insulin kinetics in type-I diabetes: continuous and bolus delivery of rapid acting insulin, IEEE Trans. Biomed. Eng. 52 (1) (2005) 3–12.
[39] A. Bertachi, L. Biagi, I.N. Contreras, N. Luo, J. Veh´I, Prediction of blood glucose levels and nocturnal hypoglycemia using physiological models and artificial neural networks, presented at the Proceeding of 5th International Workshop on Knowledge Discovery in Healthcare Data, Santiago de Compostela, Spain & Virtually, 2018.
[40] A. American Diabetes, 6. Glycemic targets: standards of medical care in diabetes-2020, Diabetes Care 43 (Suppl. 1) (2020) S66–S76.