Trend Estimation of Blood Glucose Level Fluctuations Based on Data Mining

Masaki Yamaguchi, Shigenori Kambe
Department of Material Systems Engineering and Life Science, Faculty of Engineering, Toyama University
3190 Gofuku, Toyama 930-8555, Japan

Karin Wårdell
Department of Biomedical Engineering, Institute of Technology, Linköping University
SE-581 85 Linköping, Sweden

Katsuya Yamazaki, Masashi Kobayashi
The First Department of Internal Medicine, Faculty of Medicine, Toyama Medical & Pharmaceutical University
2630 Sugitani, Toyama 930-0194, Japan

Nobuaki Honda, Hiroaki Tsutsui, Chosei Kaseda
Research & Development Headquarters, Yamatake Corporation
1-12-2 Kawana, Fujisawa 251-8522, Japan

ABSTRACT

We have fabricated calorie-calculating software that calculates and records the total calorific food intake by choosing a meal menu selected using a computer mouse. The purpose of this software was to simplify data collection throughout a person's normal life, even if they were inexperienced computer operators. Three portable commercial devices have also been prepared a blood glucose monitor, a metabolic rate monitor and a mobile-computer, and linked into the calorie-calculating software. Time-course changes of the blood glucose level, metabolic rate and food intake were measured using these devices during a 3 month period. Based on the data collected in this study we could predict blood glucose levels of the next morning (FBG) by modeling using data mining. Although a large error rate was found for predicting the absolute value, conditions could be found that improved the accuracy of the predicting trends in blood glucose level fluctuations by up to 90 %. However, in order to further improve the accuracy of estimation it was necessary to obtain further details about the patients' life style or to optimise the input variables that were dependent on each patient rather than collecting data over longer periods.

Keywords : Blood Glucose, Data Mining, Diabetic Patient, Food Intake, Metabolic Rate, Estimation

1. INTRODUCTION

Many diabetic patients collect their own blood and carry a portable-type blood glucose monitor to examine their blood glucose levels daily (Self Monitoring of Blood Glucose, SMBG). Patients using SMBG require not only a clinical understanding of its use in order to maintain blood glucose levels within a normal range (glycemic control), but also an educational understanding to realise the importance of diet and exercise. Although the above two requirements can be controlled by the patients, it is not easy for them to estimate their future carbohydrate metabolism conditions[1]-[4]. Because of this, glycemic control ultimately depends on the judgment of medical specialists.

By efficiently utilizing *in-vivo* data, such as blood glucose levels collected over a long period of time, the authors have been studying the possibility of predicting blood glucose levels using a data mining method. This method involves exploratory data analysis to establish a technology which can support glycemic control[5]. In a previous study, we reported that it was difficult to predict absolute blood glucose levels from non-continuous data, for example samples collected over a period of four months but taken only once or twice a day[6].

In this present research, we have fabricated calorie-calculating software that calculates and records the total calorific food intake by choosing a meal menu selected using a computer mouse. The purpose of this software is to simplify data collection throughout a person's normal life, even if they are inexperienced computer operators. Three portable commercial devices have also been prepared a blood glucose monitor, a metabolic rate monitor and a mobile-computer, and linked into the calorie-calculating software. Time-course changes of the blood glucose level, metabolic rate and food intake were measured using these devices during a 3 month period. Finally, based on the three data sets collected, we discussed the trends in blood glucose level to fluctuation, namely the increasing and decreasing tendencies observed from data mining.

2. MATERIALS AND METHODS

The subjects were four ambulatory diabetic patients (two male and two female, aged between 19 and 42 years old) who continuously measure their SMBG (Table1). The body mass indexes were 25.2, 17.5, 22.4, and 20.3 kg/m^2 for subject a, b, c, and d respectively. The daily insulin doses of the subjects were 46, 25, 30, and 50 Units respectively. The insulin dose was divided into four times, such as before breakfast, before lunch, before supper and pre-sleep time. The ratio of the insulin dose was different in the every individual. The aim of the experiment was explained to the subjects and consent was obtained after confirmation that they fully understood the experiment.

The fasting blood glucose level (FBG, mg/dL) was measured using a portable blood glucose monitor (ARKRAY, Inc., Japan, 45g, W51.0×L87.8×H14.5mm) (Fig.1). The subjects were attached to a portable metabolic rate monitor (Suzuken Co., Ltd., Japan, 40g, W62.5×L46.5×H26mm) in the lumbar region to enable measurement of their metabolic rate (calorie, cal) in every two minutes.

Calorie-calculating software was fabricated for the measurement of food intake. The food intake was automatically calculated by using a mouse operation to select a meal and its quantity from images of multiple meals displayed on the screen. 245 different menus were located into the software. It was classified into four sections of staple food (38), main dish (83), sub-dish (85), and fruit and favorite food (39). The calorie of these menus was mainly decided by the Standard Tables of Food Composition in Japan (Fifth revised edition). A mobile personal computer (Casio Computer Co., Ltd., Japan, 990g, W197×L223×H21.2mm) was used to install the software and was loaned to the subjects. Patients were trained to input the data

Table1 Clinical backgrounds of the three diabetes mellitus.

Subject	Age	Type of diabetes mellitus	Body mass index (kg/m^2)	Insulin dose (Units)
a	41	Type 2	25.2	Breakfast:20 Dinner:26
b	19	Type 1	17.5	Breakfast:R6 Lunch:R4 Dinner:R4 Pre-sleep:N11
c	30	Type 1	22.4	Breakfast:L8 Lunch:L6 Dinner:L8 Pre-sleep:N8
d	42	Type 1	20.3	Breakfast:L8 Lunch:L12 Dinner:L4 Pre-sleep:N26

male: a, c, female: b, d
L:lyspro-insulin, R:regular insulin, N:neutral protamine hagedorn (NPH) insulin

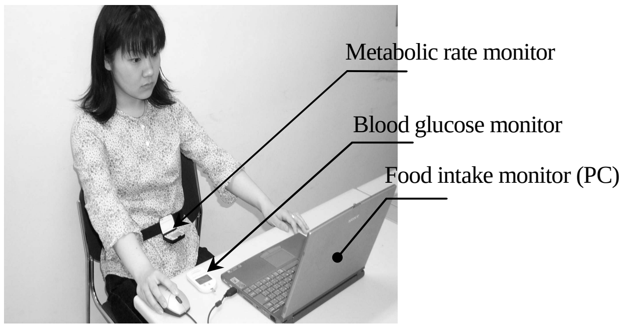

Fig.1 The data collection of blood glucose level, metabolic rate, and food intake using three portable monitors (PC: personal computer).

themselves.

Using these portable devices, the FBG (BG_m, mg/dL), total metabolic rate (Q_{out}, cal), and food intake (cal) of the subjects were measured every day for 3 months. The total metabolic rate (Q_{out}) was calculated according to the equation from basic metabolic rate (B_m, cal), quantity of motion (E_x, cal) and quantity of micro-motion (E_0, cal).

$$Q_{out} = 1.1\,(B_m + E_x + E_0) \qquad \text{(cal)} \qquad (1)$$

Moreover total food intake Q_{in} (cal), breakfast food intake $Q_{in\text{-}m}$ (cal), and supper food intake $Q_{in\text{-}d}$ (cal) were used as the food intake variables.

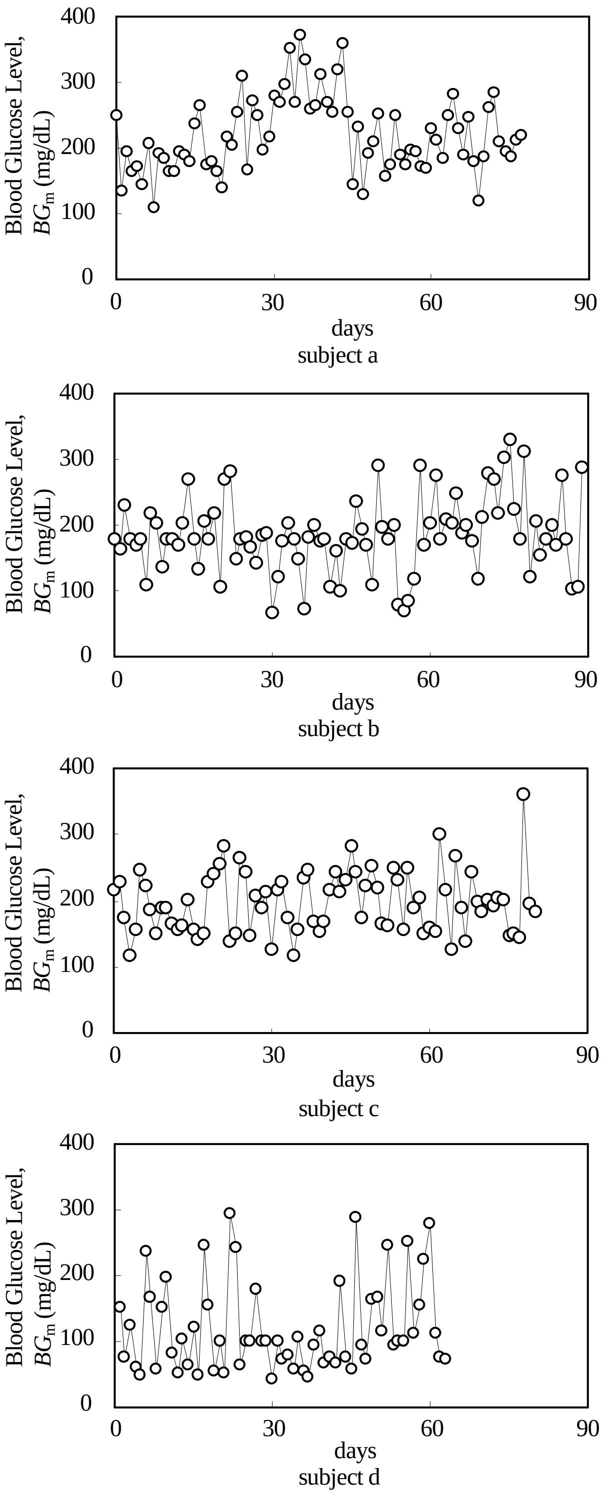

Fig.2 Time-course changes of the fasting blood glucose levels (FBG)of the 4 subjects (1mmol/L =18mg/dL).

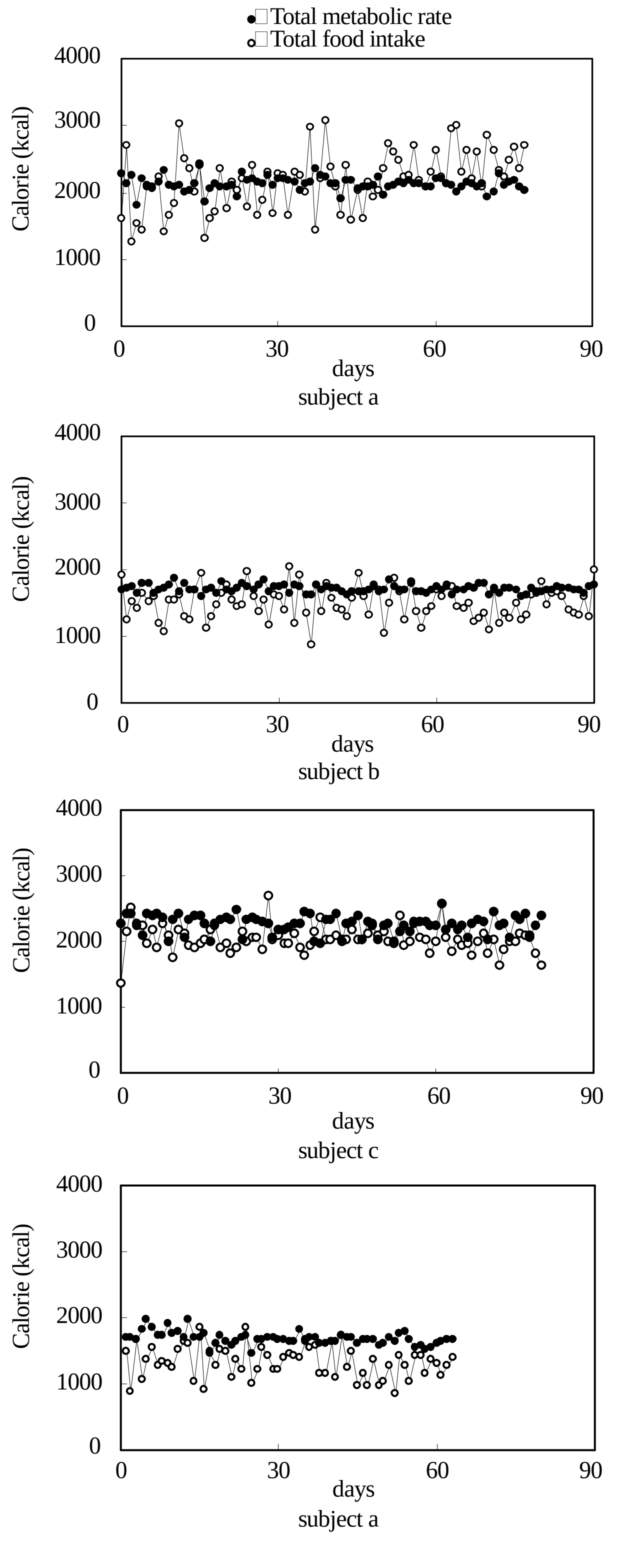

Fig.3 Time-course changes of the total metabolic rate (Q_{out}) and the total food intake (Q_{in}) of the 4subjects.

With regard to the analytical method, the data mining method used in this study is not qualitative data mining, such as when a new variable (input variable) is found from the data and that is generally used in the field of economics. Rather, it is a quantitative data mining method such as those based on knowledge of a cause and result relationship where an estimated model is

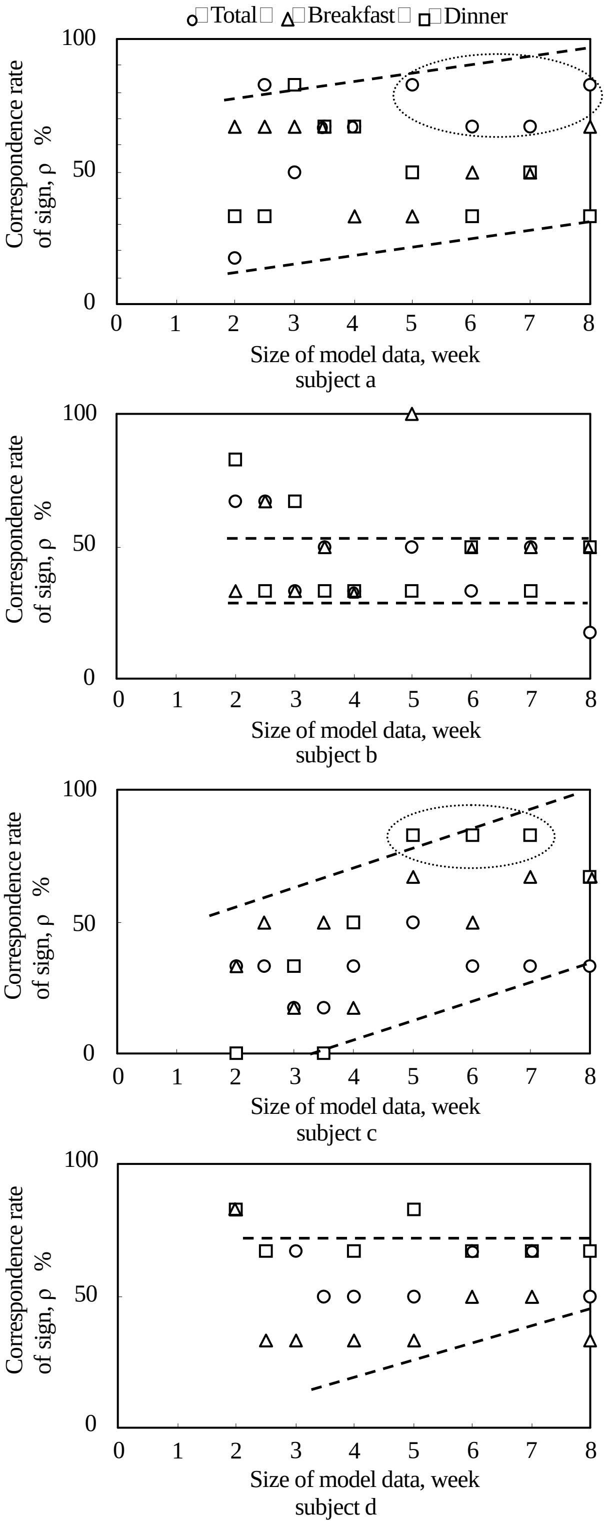

Fig.4 Correlation between correspondence rate, ρ, of sign and size of model data.

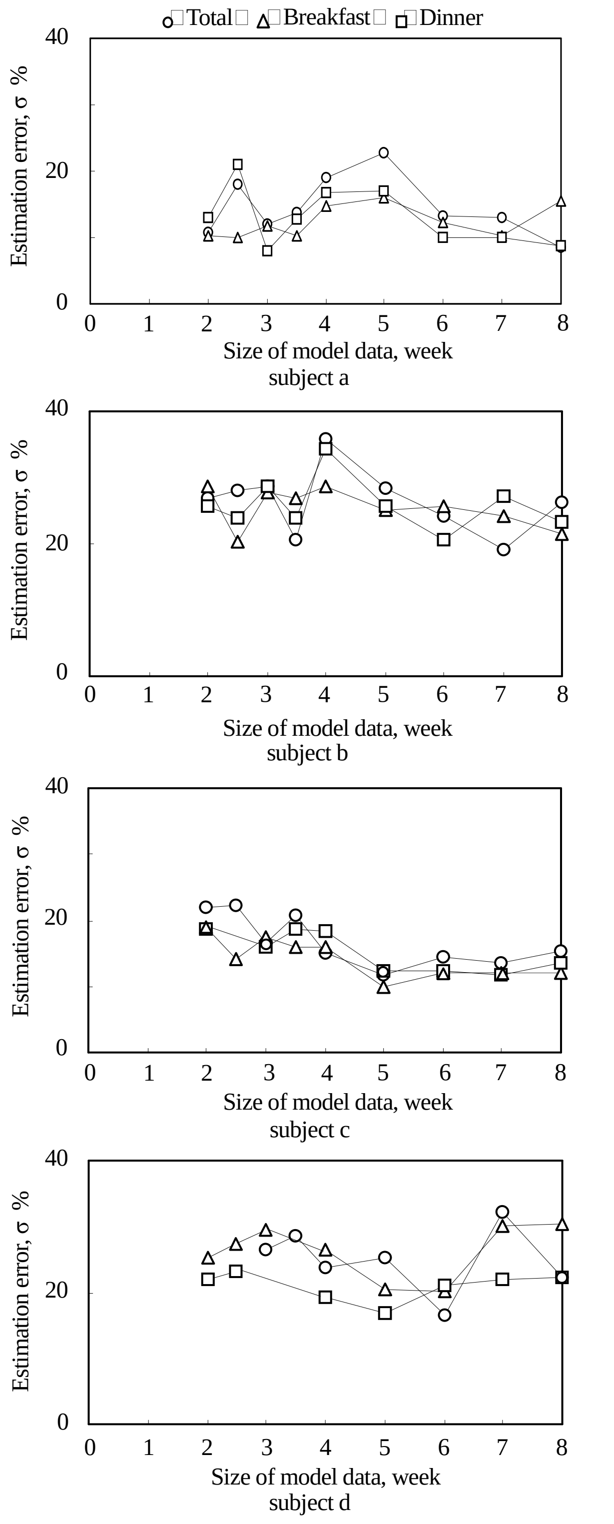

Fig.5 Correlation between estimation error, σ, and size of model data.

created and which are commonly used in the industrial field.

In this quantitative data mining method, after the estimated objective has been set (output variable of the model), variables (candidates for input variable) are selected from basic knowledge and a suitable variable is determined, taking the delay time into consideration. Through such a process an input/output model will be created. In other words, it can be considered that by using the data mining analytical information, an inverse problem is solved based on a fundamental cause-result relationship. In short, if solving a forward problem is defined as finding a result (output) from a cause (input), it could be considered that solving an inverse problem would be finding a cause (input) from a result (output). If we use an analytical method that automatically creates a model from the given data, it would be difficult to evaluate the results obtained (model). Therefore, we created a model and verified its appropriateness by taking the following steps (Topological Case-Based Modeling, TCBM, Yamatake Co., Japan) [7]. The output from the model was set as the FBG (BG_c) of the every next morning, and preparation and verification of the model was conducted using the following procedure:

1. Determination of the input variables; the input variables were the SMBG (BG_m), the total metabolic rate (Q_{out}), and the three food intake (Q_{in}, $Q_{in□m}$, $Q_{in□d}$). During the model period if a value was missing the mean value was used for completion.
2. Extension of the input variables; using a Time Delay Analysis method the delayed data of the FBG, total metabolic rate and food intake were generated to add to the functions. As a result the biorhythm was also investigated.
3. Narrowing down the input variables; using both stepwise method and cluster analysis. From this analysis the optimal combination of input variables was automatically determined. The input variables for modeling were set such that the FBG, total metabolic rate and food intake were always included. By trial and error we reduced the number of variables to 5 or less in the manual operation.
4. Modeling and verification; modeling was done using the narrowed down variables and the FBG was predicted. Comparing the predicted values with the data obtained in the verification period the model was verified.

The verification period, which remained unchanged, was the last week of the time series data. The model period varied over a range from two to eight weeks following the verification data.

Correspondence rate (ρ) and error in the estimation (σ) were used as indexes to evaluate the estimation accuracy for the trend in blood glucose level to fluctuations. The agreement ratio of the sign between measured values and the estimated values was defined as ρ. The sign was obtained from the difference between measured FBG and estimated FBG for the next day. On the other hand, σ was the mean value of largest error everyday during the verification period.

3. RESULTS

The mean values of the FBG were 219.5, 184.9, 197.4 and 119.8 mg/dL for subject a, b, c, and d respectively. The differences in maximum value ($BG_{m□max}$) and minimum value ($BG_{m□min}$) of FBG were 262, 263, 244, and 254 mg/dL for subject a, b, c, and d respectively. Large difference could not be observed in these results between the subjects (Fig.2).

The mean values of the total metabolic rate (Q_{out}) were 2132, 1715, 2261 and 1680 kcal for subject a, b, c, and d respectively. On the other hand, the mean values of the total food intake (Q_{in}) were 2189, 1500, 2038 and 1322 kcal respectively. The total food intakes were about 10% lower than the total metabolic rate, except for subject a (Fig.3). The change of weight between before and after the measurement were 3.8% increase, 4.3% decrease, 1.4% increase and no change for subject a, b, c, and d respectively.

No improvement was found in the relationship between correspondence rate, ρ, and model period for data collected over longer periods (Fig.4). When the input variable of the food intake was intentionally changed a favorable correspondence rate was found for subject "a" when the total food intake was chosen. On the other hand, for subject "c" a favorable correspondence rate was observed when the supper food intake was chosen. Correspondence rates of up to 90 % were obtained, although the correlation varied for each subject.

The error in estimation decreased as the model period increased (Fig.5). After the 8th week of the model period the errors were 8.9, 26.1, 13.5 and 22.2 % for subject a, b, c, and d respectively. The error in the estimation was independent of food intake.

4. DISCUSSION

Since no significant changes were observed in the time-course change of the blood glucose levels for the four subjects, it was considered that their glycemic control was well-balanced. In the other, the total food intake decreased about 10 % compared with the total metabolic rate in three of the subjects although the change of the weight were under 4 % in all of the subjects, namely, the energy balance did not agree perfectly. It was thought that the main variable of it was a lack of the input operation about the food intake depends on the subject. The main reason using the data mining method was to find out regularities with good correlation between the metabolic rate, the food intake and blood glucose level, and to obtain these absolute value was not the purpose of it. That is, it was considered that the difference in the metabolic rate and the food intake would not become an essential defect. One of the feature of this research was to collect all the data such as the FBG, the metabolic rate and the food intake from diabetic patients only using the portable devices, since the researches which report these *in vivo* information of diabetic patients throughout several months were not so many yet.

To estimate the trend in blood glucose level fluctuations we analyzed the correspondence rate. The result revealed that increasing the data acquisition period had no effect on improving the correspondence rate. Furthermore, the input variables for food intake that would lead to a better correspondence rate varied from subject to subject, which indicated that the food intake habits of each subject were different. In other words, the accuracy of estimating the fluctuations in blood glucose level relies on recognising the subject's life style or choosing the appropriate input variables rather than increasing the data acquisition period.

Although the error in estimation showed better results with the longer model period, it was distributed over a range from 8 to 26 %. This range was considered too large to estimate the absolute value.

Furthermore, the subjects sometimes had no meal or forgot to measure their blood glucose level, all of which leads to a missed value. For TCBM, since the collection of continuous data over time proved indispensable for modeling, it had become apparent that another method for obtaining the missing values was required.

5. CONCLUSION

In this report we presented three types of *in-vivo* data, blood glucose level, metabolic rate and food intake. Measurements of such data required development of a portable and patient operated device. Only then data could be continuously collected over a period of time.

Based on the data recorded in this study we could predict blood glucose levels the next morning (FBG) by using data mining modeling. Although a large error rate was found for predicting the absolute value, conditions could be found that improved the accuracy of the predicting trends in blood glucose level fluctuations by 90%. However, in order to further improve the accuracy of the trend estimation it is necessary to obtain more details about the patient's life style and to optimise the input variables rather than collecting data over longer time periods.

Now after, it seem to be important not only to improve the prediction accuracy of the blood glucose level but also to find out any regularity by data mining which are useful for the improvement of the quality of life of the diabetic patients.

A part of this research was supported by the grant FY 2001 from Japan health promotion & fitness foundation in Japan (Research coordinator: Masaki Yamaguchi).

REFERENCES

[1] E.D. Lehmann, T. Deutsch, "Application of computers in diabetes care - a review", **Med Inform**, **20**-4, 1995, pp.281-329.

[2] H. Morikawa, K. Izumi, "Diabetes network systems for management and education of diabetic patients", **Second International Conference on Health Technology**, 1994, pp.121-126.

[3] M. Edmonds, M. Bauer, S. Osborn, H. Lutfiyya, J. Mahon, G. Doig, P. Grundy, C. Gittens, G. Molenkamp, D. Fenlon, "Using the vista 350 telephone to communicate the results of home monitoring of diabetes mellitus to a central database and to provide feedback", **International Journal of Medical Informatics**, **51**, 1998, pp.117-125.

[4] M. Kobayashi, K. Yamazaki, R. Hayashi, "Diabetes campaign in Toyama prefecture and development of computerized diabetes care", **Int. Forum for Diabetes Outcome Research**, 1999, pp.34-37.

[5] Ed. by U.M. Fayyad, G. Piatetsky-Shapiro, P. Smyth, and R. Uthurusamy, **Advances in Knowledge Discovery and Data Mining**, California: AAAI Press, 1996, pp.1-34.

[6] M. Yamaguchi, T. Makimura, Y. Fukushi, H. Tsutsui, C. Kaseda, K. Yamazaki, M. Kobayashi, "A Study of a Clinical Algorithm for Diabetes Care Based on Data Mining", **11th Korea-Japan Symposium on Diabetes Mellitus**, 2001, p.120.

[7] H. Tsutsui, A. Kurosaki and T. Sato, "Nonlinear Modeling Technique Using Historical Data for Case TCBM, Topological Case Based Modeling", **Journal of the Society of Instrument and Control Engineers**, **33**, 1997, pp.947-954.