MEDINFO 2017: Precision Healthcare through Informatics
A.V. Gundlapalli et al. (Eds.)

doi:10.3233/978-1-61499-830-3-619

An Early Infectious Disease Outbreak Detection Mechanism Based on Self-Recorded Data from People with Diabetes

Ashenafi Zebene Woldaregay[a], Eirik Årsand[b], Taxiarchis Botsis[a], Gunnar Hartvigsen[a, b]

[a] Department of Computer Science, University of Tromsø – The Arctic University of Norway, Tromsø, Norway,
[b] Norwegian Centre for E-health Research, University Hospital of North Norway, Tromsø, Norway

Abstract

People with diabetes experience elevated blood glucose (BG) levels at the time of an infection. We propose to utilize patient-gathered information in an Electronic Disease Surveillance Monitoring Network (EDMON), which may support the identification of a cluster of infected people with elevated BG levels on a spatiotemporal basis. The system incorporates data gathered from diabetes apps, continuous glucose monitoring (CGM) devices, and other appropriate physiological indicators from people with type 1 diabetes. This paper presents a novel approach towards modeling of the individual's BG dynamics, a mechanism to track and detect deviations of elevated BG readings. The models were developed and validated using self-recorded data in the non-infection status using Dexcom CGM devices, from two type 1 diabetes individuals over a 1-month period. The models were also tested using simulated datasets, which resemble the individual's BG evolution during infections. The models accurately simulated the individual's normal BG fluctuations and further detected statistically significant BG elevations.

Keywords:

Blood Glucose; Disease Outbreaks; Diabetes Mellitus, Type 1

Introduction

Diabetes mellitus is a chronic disease that causes blood glucose (BG) metabolic disorder [1], either due to failure of the pancreas beta cells to produce insulin (type 1) or failure of the body to react to insulin in the proper way (type 2) [2]. Many people with diabetes strive to control their BG levels as close to the normal range as possible to avoid medical complications. In this regard, many diabetes self-management applications and devices have been developed to support this patient group, of which almost all of them take the ubiquitous nature of mobile devices as an advantage to base their development [3-6]. Moreover, diabetes self-management applications and devices have shown feasible to integrate with the patient's Electronic Health Record system [7-9]. The advent of information technology and availability of biosensors and point of care technologies have also paved the way for quantified self and an easy near patient testing [10; 11]. These advancements have further enhanced the opportunity of using the individual's diabetes data and other physiological indicators for secondary purposes.

People with diabetes experience elevated BG levels when in infectious state [12; 13]. A positive correlation between high BG levels and infections has been demonstrated in the cases of Influenza, Cholera, Plague, Ebola, Anthrax and SARS [14; 15]. Use of BG levels for an early outbreak detection has been suggested in works of literature [13; 14; 16-19]. For example, Granberg et al. [19] introduced an automatic infection detection system based on the individual's BG levels. Årsand et al. [14] described the system architecture, model and requirements of disease surveillance based on patient observable parameter, i.e. blood glucose. Botsis et al. [20] assessed the development of electronic disease surveillance systems for detecting infections at the early stages, i.e. during the incubation period. Aside from these potential methods, no practical way of facilitating this has been identified.

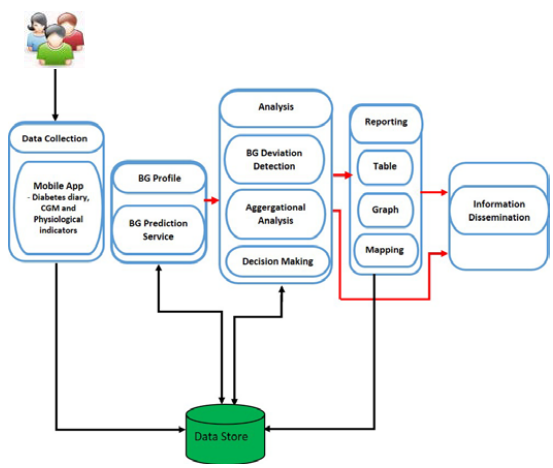

Figure 1 - Proposed EDMON System Architecture.

We propose the development of an Electronic Disease Surveillance Monitoring Network (EDMON), which may support the detection of infections before the onset of the first symptoms. The system incorporates data gathered from diabetes apps, CGM devices, and other appropriate physiological indicators from people with type 1 diabetes. It consists of five different modules, as shown in Figure 1, with different functionalities; data collection, BG profile, analysis, reporting, and information dissemination modules. The task of the data collection module is to gather the individual's diabetes data and other physiological indicators from diabetes apps, wearables, and sensors, whereas BG profile module will make a prediction of individuals' upcoming BG levels based on past status. The analysis module will analyze the individual's physiological indicators and compare the predicted and actual BG levels for

any statistically significant deviations. It also aggregates and determines the presence of any aberrant pattern within a cluster of people based on a spatiotemporal basis. If outbreak is detected, the reporting module will prepare the information in a way suitable for the end users, whereas the information dissemination module will handle the delivery of the information for the end users. Generally, the objective of the EDMON project is to design and develop an electronic disease monitoring network based on inputs from people with diabetes, which can track real-time BG levels of each individual independently and detect a cluster of infected people with statistically significant BG deviations on a spatiotemporal basis. This paper presents the first step of the EDMON development with modeling of the individual's BG dynamics. We consider the development of the personalized model as the core part of the EDMON system and describe the details of our approach in the next sections.

Materials and Methods

Materials

The models were developed using one-month BG data from two type 1 diabetes people, as shown in Figure 2, sampled using Dexcom CGM devices. The actual self-recoded data were used to develop and validate the model fit with individuals' BG dynamics in the non-infection state. We also generated simulated datasets that resemble the individual's BG fluctuations during infections, as shown in Figure 3. The simulated datasets included unexpected elevated BG levels for a certain period of time with some random and steady increments per hours ($\Delta BG/hour$). The unexpected BG variations were defined as 1) any unstructured variations outside of the individual's modeled structured variations and, 2) inherent stochastics phenomena of the BG dynamics that cannot be quantified with the developed models using the key diabetes parameters (e.g. BG, insulin intake, physical activities, and dietary habits). The simulated datasets were used to test the model's performance in capturing the assumed infection-related, unexpected, elevated BG readings. The algorithm was developed using Matlab version R2015b.

Methods

Our models combined a novel approach for BG monitoring and outlier detection, which was based on a set of autoregressive models and predicts the individuals' expected BG values on an interval basis. The actual BG value was compared with the predicted intervals, which was generated using autoregressive (AR) model [21], Eq. (1), and autoregressive moving average (ARMA) model [21; 22], Eq. (2).

$$y(n) = \sum_{i=1}^{q} \mathcal{A}(i) y(n-i) + e(n) \quad (1)$$

Where, $\mathcal{A}(q)$ are autoregressive coefficients, and $y(n)$ is nth BG value.

$$\sum_{i=0}^{q} \mathcal{A}(i) y(n-i) = \sum_{i=0}^{q} \mathcal{C}(i) e(n-q) \quad (2)$$

Where, $\mathcal{A}(q)$ are autoregressive coefficients, $\mathcal{C}(q)$ are moving average coefficients, and $y(n)$ is the nth BG value, and $e(n)$ are Gaussian noise with zero mean and constant variance σ_e^2.

The prediction intervals [23-25] were computed based on the empirical distribution of errors between the predicted and actual BG values for the prediction horizon under consideration, using Eq. (3).

$$\tilde{y}(n) = y(n) \pm z_{\alpha/2} \sqrt{Var[e_k(n)]} \quad (3)$$

Where, $\tilde{y}(n)$ is the predicted BG intervals, $y(n)$ is the model's point BG prediction, $z_{\alpha/2}$ is the assumed errors distribution, α is level of significance, and $Var[e_k(n)]$ is variance of the errors for a specific window size, w. The prediction interval was computed and compared to various values of window size and level of significance. The optimal prediction interval was reported with a window size (w) and level of significance (α).

MATLAB system identification toolbox along with partial autocorrelation function (PACF) were used to identify the optimal model order. The empirical distribution of errors between the actual and predicted values were assumed to follow a normal distribution.

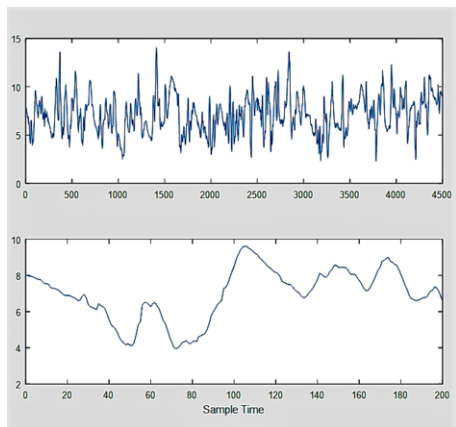

Figure 2 - Plot of the CGM data. [x and y-axes represent CGM's sampling time and BG level in mmol/l, respectively].

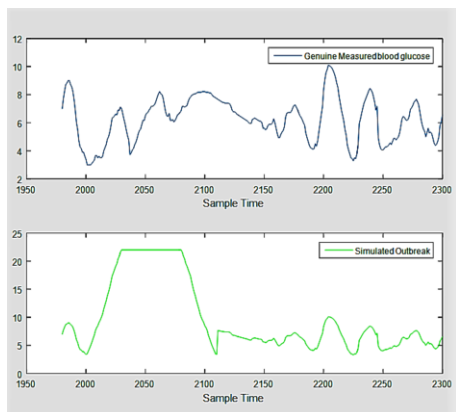

Figure 3 - Simulated BG values in response to infections (Top -non-infectious BG levels, Bottom -simulated infection related BG levels). [x and y-axes represent CGM's sampling time and BG level in mmol/l, respectively].

Results

Autoregressive (AR) Model

For both the first and the second subject, an autoregressive (AR) model of order five (p = 5) was found to be optimal for the point BG prediction and fitted best with a root mean square error (RMSE) of 0.2159 and 0.3068 respectively. For the first subject, the predicted interval was found to be optimal with a window size of $w = 100$ and a statistically significant level of $\alpha = 0.01$, see Figure 4. The predicted interval for the second subject was also found to be optimal with a window size of $w = 200$ and a statistically significant level of $\alpha = 0.01$, see Figure 5.

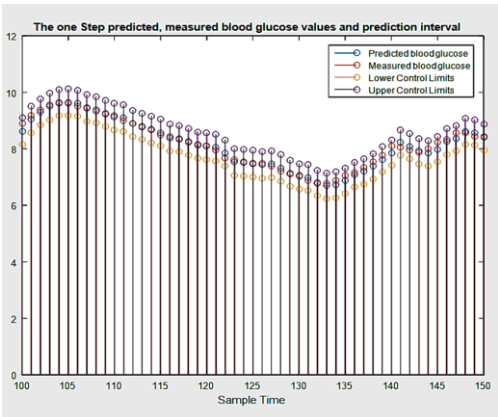

Figure 4 - Subject one- autoregressive (AR) model. [x and y-axes represent CGM's sampling time and BG level in mmol/l, respectively].

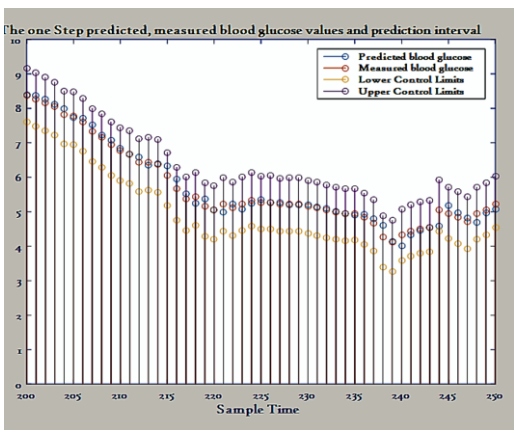

Figure 5 - Subject two - autoregressive (AR) model. [x and y-axes represent CGM's sampling time and BG level in mmol/l, respectively].

Autoregressive Moving Average (ARMA) Model

For both subjects, the optimal autoregressive moving average (ARMA) model order was found to be an autoregressive order of 6, and a moving average order of 2. The point BG prediction for the first subject resulted in a root mean square error (RMSE) of 0.2114. The predicted interval was found to be effective with a window size of $w = 100$ and a statistically significant level of $\alpha = 0.01$, see Figure 6. The point BG prediction for the second subject also resulted in a RMSE of 0.2915. The predicted interval was found to be effective with a reasonable window size of $w = 200$ and a statistically significant level of $\alpha = 0.01$, see Figure 7.

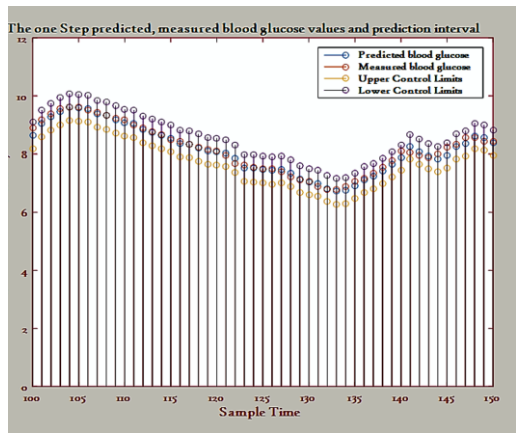

Figure 6 - Subject one - autoregressive moving average (ARMA) model. [x and y-axes represent CGM's sampling time and BG level in mmol/l, respectively].

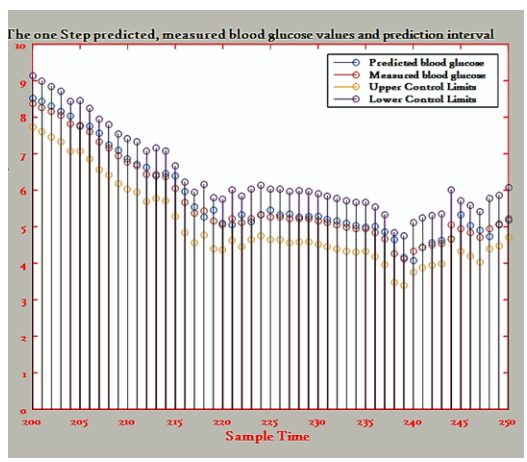

Figure 7 - Subject two- autoregressive moving average (ARMA) model. [x and y-axes represent CGM's sampling time and BG level in mmol/l, respectively].

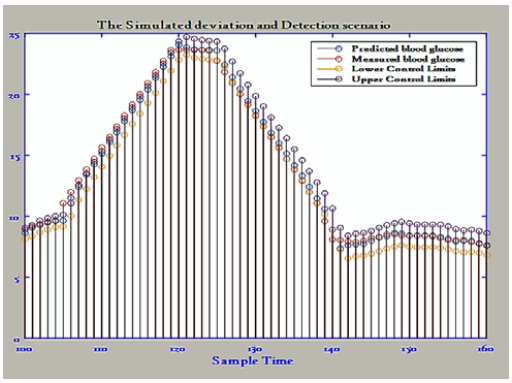

Figure 8 - Detection capabilities of the developed algorithm on testing data. [x and y-axes represent CGM's sampling time and BG level in mmol/l, respectively].

Deviation Detection/Surveillance

The algorithm was tested with simulated datasets, for its capability of detecting unexpected BG variations that may occur during infections, as shown in Figure 3. The developed algorithm successfully detected statistically significant elevated BG readings, as shown in Figure 8. As clearly shown in the Figure, the algorithm was highly sensitive to the slope, and clearly captured the rise and fall of the individual's BG readings.

Discussion

The recent advancement, syndromic surveillance uses health-related data that precede diagnosis and laboratory verification to produce signals with sufficient probability of outbreaks to warrant further actions [26-30]. The development of strategies for early detection of outbreaks is worthy, given the limitations of the existing disease surveillance systems. We presented a modelling approach for the early detection of infections in people with diabetes. This set of models were developed as part of the EDMON system that will rely on real-time data collection from people with diabetes. Our approach was capable of detecting statistically significant and unexplained BG elevations of various size and duration.

One of the limitations of our study is the sample size. We also have used only BG as the input variable. Furthermore, the assumption of a normal error distribution could be a limitation, which needs further exploration. To alleviate these limitations, we plan to explore other more robust approaches and involve real infection related BG data, more input variables (insulin intake, physical activity, and diet) along with a larger sample size.

Conclusion

EDMON is an early outbreak detection system that relies on self-gathered health-related data from people with type 1 diabetes. In this paper, we presented a novel approach that can track BG levels and detect statistically significant BG elevations. The testing and validation of this approach on a large-scale basis could support the development of an outbreak detection system based on real-time data collection from people with diabetes. We believe such efforts may lay the foundations for the next generation disease surveillance systems and provoke further thoughts in this valuable field.

References

[1] W.H. Organization, Diabetes Fact Sheet, in, 2016.
[2] R.L. Richesson, S.A. Rusincovitch, D. Wixted, B.C. Batch, M.N. Feinglos, M.L. Miranda, W.E. Hammond, R.M. Califf, and S.E. Spratt, A comparison of phenotype definitions for diabetes mellitus, *J Am Med Inform Assoc* **20** (2013), e319-326.
[3] E. Årsand, N. Tatara, G. Østengen, and G. Hartvigsen, Mobile Phone-Based Self-Management Tools for Type 2 Diabetes: The Few Touch Application, *Journal of Diabetes Science and Technology* **4** (2010), 328-336.
[4] D.-Z. Issom, A.Z. Woldaregay, T. Chomutare, M. Bradway, E. Årsand, and G. Hartvigsen, Mobile applications for people with diabetes published between 2010 and 2015, *Diabetes Management* **5** (2015), 539-550.
[5] C.C. Quinn, S.S. Clough, J.M. Minor, D. Lender, M.C. Okafor, and A. Gruber-Baldini, WellDoc mobile diabetes management randomized controlled trial: change in clinical and behavioral outcomes and patient and physician satisfaction, *Diabetes Technol Ther* **10** (2008), 160-168.
[6] K. Waki, H. Fujita, Y. Uchimura, K. Omae, E. Aramaki, S. Kato, H. Lee, H. Kobayashi, T. Kadowaki, and K. Ohe, DialBetics: A Novel Smartphone-based Self-management Support System for Type 2 Diabetes Patients, *J Diabetes Sci Technol* **8** (2014), 209-215.
[7] P.Y. Benhamou, Improving diabetes management with electronic health records and patients' health records, *Diabetes & Metabolism* **37** (2011), S53-S56.
[8] T.C. Veinot, K. Zheng, J.C. Lowery, M. Souden, and R. Keith, Using Electronic Health Record Systems in Diabetes Care: Emerging Practices, *IHI ... : proceedings of the ... ACM SIGHIT International Health Informatics Symposium. ACM SIGHIT International Health Informatics Symposium* **2010** (2010), 240-249.
[9] O.A. Walseth, E. Årsand, T. Sund, and E. Skipenes, Wireless Transfer of Sensor Data into Electronic Health Records, in: *Connecting Medical Informatics and Bio-Informatics*, IOS Press, 2005.
[10] T. Botsis, S. Walderhaug, A. Dias, K. van Vuurden, J.G. Bellika, and G. Hartvigsen, Point-of-care devices for healthy consumers - a feasibility study, *J Telemed Telecare* **15** (2009), 419-420.
[11] B.C. Delaney, C.J. Hyde, R.J. McManus, S. Wilson, D.A. Fitzmaurice, S. Jowett, R. Tobias, G.H. Thorpe, and F.D. Hobbs, Systematic review of near patient test evaluations in primary care, *BMJ* **319** (1999), 824-827.
[12] S. Knapp, Diabetes and infection: is there a link?--A mini-review, *Gerontology* **59** (2013), 99-104.
[13] J.N. Lauritzen, E. Arsand, K. Van Vuurden, J.G. Bellika, O.K. Hejlesen, and G. Hartvig-sen, Towards a mobile solution for predicting illness in Type 1 Diabetes Mellitus: Development of a prediction model for detecting risk of illness in Type 1 Diabetes prior to symptom onset, (2011), 1-5.
[14] E. Arsand, O.A. Walseth, N. Andersson, R. Fernando, O. Granberg, J.G. Bellika, and G. Hartvigsen, Using Blood Glucose Data as an Indicator for Epidemic Disease Outbreaks, *Connecting Medical Informatics and Bio-Informatics* **116** (2005), 217-222.
[15] T. Botsis, O. Hejlesen, J.G. Bellika, and G. Hartvigsen, Electronic disease surveillance for sensitive population groups - the diabetics case study, *Stud Health Technol Inform* **136** (2008), 365-370.
[16] T.Botsis, O.Hejlesen, J.G.Bellika, and G.Hartvigsen, Blood glucose levels as an indicator for the early detection of infections in type-1 diabetics, in: *Advances in Disease Surveillance*, 2007, p. 147.
[17] T. Botsis and G. Hartvigsen, Exploring new directions in disease surveillance for people with diabetes: lessons learned and future plans, *Stud Health Technol Inform* **160** (2010), 466-470.
[18] A.Z. Woldaregay, K. van Vuurden, E. Årsand, T. Botsis, and G. Hartvigsen, Electronic Disease Surveillance System Based on Input from People with Diabetes: An Early Outbreak Detection Mechanism, in: *Proceedings from The 14th Scandinavian Conference on Health Informatics 2016, Gothenburg, Sweden, April 6-7 2016*, Linköping University Electronic Press, 2016, pp. 23-27.
[19] O. Granberg, J.G. Bellika, E. Arsand, and G. Hartvigsen, Automatic infection detection system, *Stud Health Technol Inform* **129** (2007), 566-570.
[20] T. Botsis, J.G. Bellika, and G. Hartvigsen, New Directions in Electronic Disease Surveillance: Detection of Infectious Diseases during the Incubation Period, *International Conference on Ehealth, Telemedicine, and Social Medicine: Etelemed 2009, Proceedings* (2009), 176-183.
[21] M. Eren-Oruklu, A. Cinar, L. Quinn, and D. Smith, Estimation of Future Glucose Concentrations with Subject-Specific Recursive Linear Models, *Diabetes Technology & Therapeutics* **11** (2009), 243-253.
[22] V. Marmarelis and G. Mitsis, *Data-driven Modeling for Diabetes: Diagnosis and Treatment*, Springer Science & Business, 2014.
[23] C. Chatfield, Calculating Interval Forecasts, *Journal of Business & Economic Statistics* **11** (1993), 121-135.
[24] B.E. Hansen, Interval forecasts and parameter uncertainty, *Journal of Econometrics* **135** (2006), 377-398.
[25] Y.S. Lee and S. Scholtes, Empirical prediction intervals revisited, *International Journal of Forecasting* **30** (2014), 217-234.
[26] T.C. Chan, C.C. King, M.Y. Yen, P.H. Chiang, C.S. Huang, and C.K. Hsiao, Probabilistic daily ILI syndromic surveillance with a spatio-temporal Bayesian hierarchical model, *PLoS One* **5** (2010), e11626.

[27] L. Chen, K.S.M.T. Hossain, P. Butler, N. Ramakrishnan, and B.A. Prakash, Flu Gone Viral: Syndromic Surveillance of Flu on Twitter Using Temporal Topic Models, in: *2014 IEEE International Conference on Data Mining*, 2014, pp. 755-760.
[28] C. Chung-Kuo, J.M. Overhage, and J. Huang, An application of sensor networks for syndromic surveillance, in: *Proceedings. 2005 IEEE Networking, Sensing and Control, 2005.*, 2005, pp. 191-196.
[29] T.C. Weng, T.C. Chan, H.T. Lin, C.K. Chang, W.W. Wang, Z.R. Li, H.Y. Cheng, Y.R. Chu, A.W. Chiu, M.Y. Yen, and C.C. King, Early detection for cases of enterovirus- and influenza-like illness through a newly established school-based syndromic surveillance system in Taipei, January 2010 ~ August 2011, *PLoS One* **10** (2015), e0122865.
[30] W. Yan, L. Palm, X. Lu, S. Nie, B. Xu, Q. Zhao, T. Tao, L. Cheng, L. Tan, H. Dong, and V.K. Diwan, ISS--an electronic syndromic surveillance system for infectious disease in rural China, *PLoS One* **8** (2013), e62749.

Address for correspondence

Ashenafi Zebene Woldaregay, Department of Computer Science, University of Tromsø - The Arctic University of Norway, 9037 Tromsø, Norway, Email: ashenafi.z.woldaregay@uit.no