Electronic Disease Surveillance System Based on Inputs from People with Diabetes: An Early Outbreak Detection Mechanism

Ashenafi Zebene Woldaregay[a], Klaske van Vuurden[a], Eirik Årsand[b, c], Taxiarchis Botsis[a], Gunnar Hartvigsen[a, c]

[a]Department of Computer Science, University of Tromsø – The Arctic University of Norway, Tromsø, Norway
[b]Department of Clinical Medicine, University of Tromsø – The Arctic University of Norway, Tromsø, Norway
[c]Norwegian Centre for eHealth Research, University Hospital of North Norway, Tromsø

Abstract

Pandemics or epidemics are serious concerns for any public health authority and mandate for proper monitoring and early detection strategies. In this study, we focus on people with diabetes and propose the use of continuous blood glucose, insulin, and dietary data, to develop an algorithm for the early detection of infections during the incubation period (i.e. before the onset of the first symptoms).

We present a system that consists of three modules: the blood glucose prediction, the outbreak detection, and the information dissemination and reporting module. The novel approach incorporated in the system is an interval prediction mechanism that is based on a set of autoregressive models and predicts the blood glucose values for an individual with diabetes. The actual blood glucose value is compared against the predicted interval, which is generated using autoregressive (AR) and Autoregressive moving average (ARMA) methods. The system was trained and validated based on continuous blood glucose measurements (CGM) from two individuals with type 1 diabetes. The single step point prediction was found to be accurate with a Root Mean Square Error (RMSE) of 0.2121 mmol/l. Moreover, we accurately monitored the blood glucose fluctuations for an individual with a significance level of $\alpha = 0.01$. The model was also tested against an artificially simulated dataset, which resembles blood glucose evolution of an infected individual with diabetes, and successfully detected statistically significant deviations from the normal blood glucose values. Our prototype system is still under development and has not been fully tested yet. Our initial findings though are promising and we plan to further test and validate our approach.

***Keywords**:*

Diabetes Mellitus, Continuous blood glucose measurement, Self-management system, Blood glucose prediction, Outbreak detection, Electronic disease surveillance.

Introduction

Most of the existing self-management applications for people with diabetes include modules for continuous monitoring of the blood glucose measurements (CGM) to assist individuals in better controlling their blood glucose (BG) levels. Mobile devices and smart phones offer considerable advantages towards the development of sophisticated apps [2, 9, 11, 14]. Recently, mobile self-management applications for people with diabetes have been integrated with Electronic Health Records [4, 13, 15]. If this integration is coupled with timely CGM data from people having diabetes, it can further enhance the establishment of efficient and effective disease surveillance systems.

Previous findings indicated that BG levels are elevated due to any exposure to pathogens [10]. Årsand et al. demonstrated an elevation in BG levels for both type 1 and type 2 diabetes individuals after the infection by Influenza, Cholera, Plague, Ebola, Anthrax, or SARS viruses [3]. Botsis et al. also described the positive correlation between BG elevation and infections in people with type 1 diabetes [7]. These findings suggest the potential use of the BG parameter for the early detection of disease outbreaks in the general population [3, 7]. Other parameters (such as body temperature, white blood cell count and blood pressure) are directly associated with the presence of infections in the body [6, 12]. Multiple incidents with abnormal values for the above parameters in the population may indicate the presence of an outbreak [1, 10]. We therefore argue that the incorporation of all these parameters into advanced modeling solutions can potentially support the early detection of outbreaks. The objective of our research is the development of a reliable electronic disease surveillance system for the analysis of diabetes data at both the individual and the population level. In this paper, we describe our initial exploration and our first-hand results.

Materials and Methods

Datasets

This research was conducted using data from two individuals with type 1 diabetes. The Dexcom CGM and the diabetes diary[1] that have been developed by Norwegian Center for E-health Research (previously known as NST) were used for the data collection. These modules are part of a mobile applica-

tion designed for diabetes management. The collected data included continuous BG measurements from the Dexcom CGM (in 5 minutes intervals) for one month and BG, insulin, diet and physical activity data from the diabetes diary for one year[1]. We used these datasets to train and validate the developed system for its goodness of fit to the BG dynamics of the two subjects in their non-infection status. We subsequently tested our system with a simulated dataset that included consecutive patterns of high BG values; this resembled the CGM during the infection period. Various increments per minutes ($\frac{\Delta BG}{minutes(t)}$) and various time intervals of elevated BG were considered.

Methods

The system can predict the BG values with a confidence interval and assess this prediction against the actual BG values. It can further analyze the measured and predicted BG values for the presence of any aberrant pattern. If there is a detection of any abnormality, the system will generate and send a notification signal to the concerned bodies or authorities and support the investigation by displaying this on the map of the interest. The system consists of a BG prediction module, an outbreak detection module, and an information dissemination and reporting module.

Blood glucose prediction module

This module includes a personalized health model that monitors the BG fluctuations of the individual with diabetes. It predicts the single step BG value using the previous BG, insulin, diet and physical activity records. This module also calculates the confidence interval of the predicted values based on the recent empirical distribution of errors between the actual value and the predicted value. The prediction module utilizes a black box approach using an autoregressive model that incorporates Autoregressive (AR), Autoregressive with Exogenous input (ARX), Autoregressive Moving Average (ARMA), and Autoregressive Moving Average with Exogenous input (ARMAX) methods. Autoregressive models were selected because they rely on the most recent information to forecast the future values. In our approach, it is very important to follow the persons' cyclical habit on a weekly or longer-period basis. The model simplicity and reproducibility were the factors that were considered in our selection. The well-defined procedure for calculating the intervals of the forecasts is definitely important as well. We evaluated and compared the performance of these models using the Root Mean Square Error (RMSE) function.

Outbreak detection module

The outbreak detection module is necessary for comparing the actual BG values with the predicted intervals. This module is built on mathematical models that can compare and detect any statistically significant deviations between the measured and the predicted BG values. This outbreak detection mechanism evaluates whether the actual BG values are outside of the predicted interval for the individual. Moreover, moving window z-score are used for better detection accuracy. The purpose of this moving window z-score is the detection of any significant deviations (anomalies in the data) based on the moving mean and standard deviation. Given a window size w, the mean and standard deviations are used to check the agreement of the actual BG measurement with the previous trend in w. This module also performs an aggregation analysis, which counts the maximum number of events on a spatio-temporal basis. In other words, it detects a disease outbreak in both space and time using a specified threshold that is defined based on the region it covers (space) and occurrence of statistically deviated BG values (time) for a number of individuals. If the number of people in the cluster exceeds the threshold, an alarm will be sent to public health authorities or hospitals. The performance of this module is evaluated based on the accuracy of detecting the cluster in a timely manner. A Receiver Operating Characteristic (ROC) curve is used to determine the best operating threshold of the system.

Information dissemination and reporting module

A principal function of the disease surveillance system is the generation of reports containing information about the detected disease outbreak. The related information is presented in tables, graphs and maps. The corresponding module submits the reports to the authorities and other interested parties via SMS and Email. Initially, an SMS is sent followed by an email to the responsible persons with the adequate information regarding the outbreak. The email contains information about the spatial and temporal distribution of the disease outbreak on a map of the region, the degree of severity and other critical data.

Design and Implementation

Prediction Model and Interval Prediction

The prediction of the BG values is based on an Autoregressive model including autoregressive with Yule Walker Algorithm, Autoregression using ratio of consecutive data points and Autoregressive Moving Average using Yule Walker Algorithm.

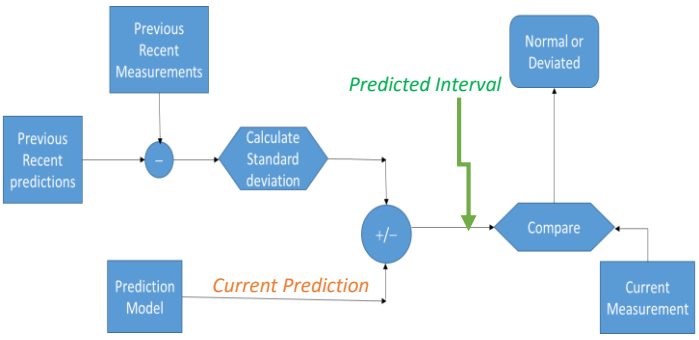

Figure 1: The Proposed Algorithm

Based on the point prediction and the empirical error distribution between the measured and the predicted value, a prediction interval is calculated with a certain confidence interval (1-α)*100%, where α is the level of significance [8].

[1] www.diabetesdagboka.no

As shown in Figure 1, the proposed algorithm computes the predicted intervals based on the previous recent predictions and measurements along with the current point predictions. The empirical distributions of errors between the previous predictions and measurements are the basis for the current interval prediction. This is clearly shown in Figure 1, where the predicted intervals are compared with the current measurements.

The system was developed in MATLAB version R2015b. A system identification toolbox along with the partial autocorrelation function (PACF) was used to identify the optimal model order. The autoregressive (AR) and autoregressive moving average (ARMA) were developed based on the CGMs that are shown in the Figure 2.

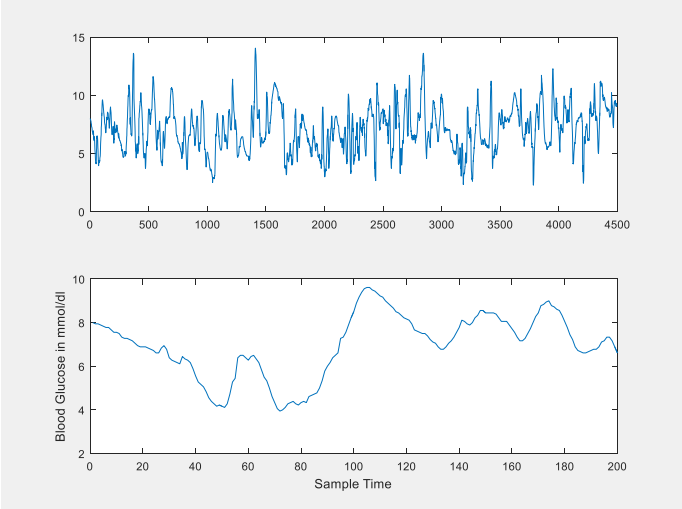

Figure 2: Plot of the entire sets and the first 200 data elements of the continuous blood glucose data.

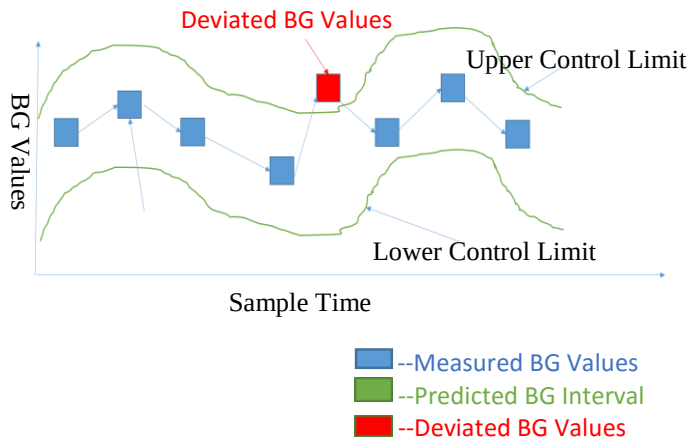

Figure 3: The proposed solution for the detection of the blood glucose deviation for an individual patient.

Outbreak Detection/Surveillance

The proposed solution is similar to a control chart/statistical process control algorithm, where the controls are determined by the intervals predicted from the individual blood concentration profiles defined by the AR models. As shown in Figure 3, the next BG value can be effectively controlled by the predicted upper and lower control limits with a reasonable accuracy. Moreover, as described in the above section (see outbreak detection module), the output results from the moving window z-score and the output results from the predicted intervals mechanism are augmented for better accuracy.

Results

We used the autoregressive models to predict the BG values using CGMs in 5-minute intervals. Autoregressive model using Yule-Walker algorithm, autoregressive model using ratio of the consecutive data points and autoregressive moving average with Yule-Walker algorithm were implemented and tested for 8495 data points. The RMSEs were calculated for 4495 testing data points. The first model, a fifth order autoregressive (AR), efficiently predicted the single step BG values with a RMSE equal to 0.9727 *mmol/l*. The second model, a fifth order autoregressive, is also capable of predicting the single step BG values. The prediction produced interesting results with a RMSE equal to 0.3413 *mmol/l*. Furthermore, the third model, an autoregressive moving average with a third order autoregressive terms and a second order moving average terms, is also capable of predicting the single step BG values. The prediction generated promising results with a RMSE equal to 0.2121 *mmol/l*. The prediction interval calculated from these models was constructed with a significance level of $\alpha = 0.01$, which means that one is 99% confident that the future values fall within the predicted intervals. Both the first and the second models produced intervals with reasonable sizes. However, the third model had a shortcoming in producing a good prediction interval, which is too narrow.

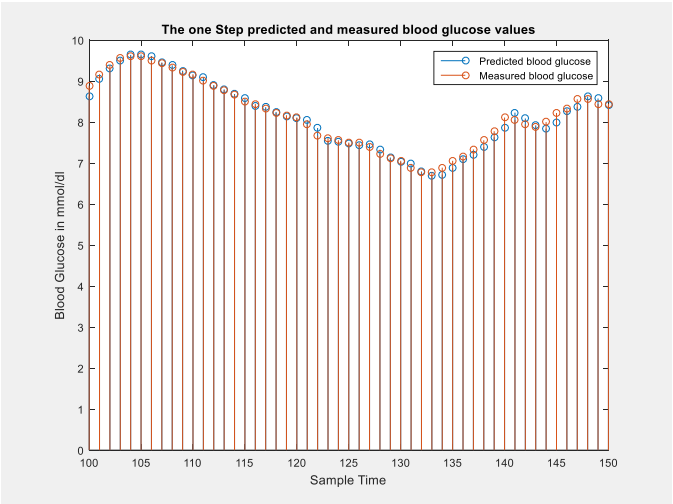

Figure 4: The predicted and measured blood glucose.

The point prediction and its interval prediction for the single diabetes subject are given in the Figure 4 and 5. These results were generated from the first model, the Autoregressive model using Yule-Walker algorithm.

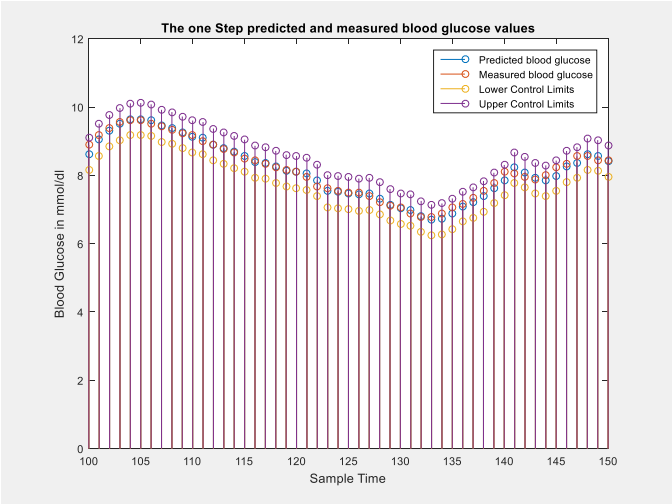

Figure 5: The predicted interval, predicted and measured blood glucose.

The moving window z-score process is also capable of detecting high BG values based on trends for various periods and rates of growth. For example, as shown in Figure 6, it can detect BG values over a long period of time. Therefore, an outlier can be detected by setting threshold values of three and more standard deviations from the mean value.

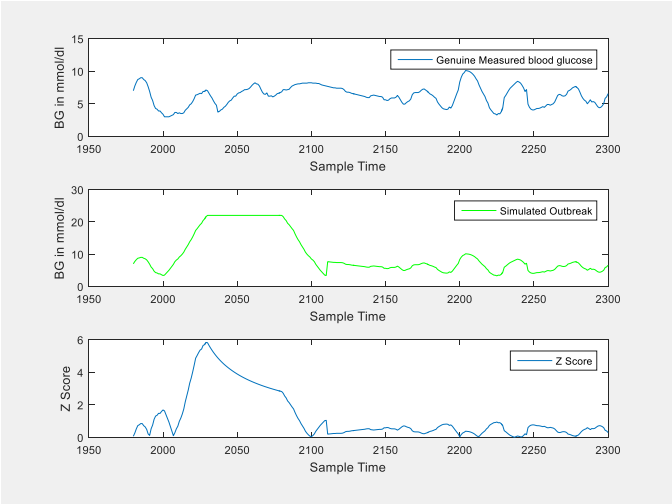

Figure 6: Measured blood glucose, simulated outbreak and the moving window z-score.

Assumptions, Biases and Limitations

The major limitation of this project is the sample size. We based our experiments and simulation on two individuals with type 1 diabetes, and more data is needed to further validate our approach. Moreover, the "holiday effect" has not been considered in this study. The "*Holiday effect*" is the bad eating style of people with diabetes in the holiday season [10] and usually leads to high BG values. In such cases our system may generate false alarms, especially given the absence of frequent measurements for other supporting parameters, such as the white blood cell counts and temperature readings from these individuals [5, 6].

Conclusion

With the advent of information technology, the transition from paper- into electronic-based reporting has revolutionized the disease surveillance systems. Our system should be grouped under the syndromic surveillance systems that also use certain data (absenteeism, Internet search volume, over the counter pharmacy sells and so forth) prior to the confirmation of infections through diagnosis. However, this information is generated after the onset of the first symptoms and syndromic surveillance systems that focus on the incubation period have not been developed yet. This is the novel and unique characteristic of our work. Our system incorporates a BG prediction mechanism that can both predict the BG values for an individual and efficiently detect an infection during the incubation period. Even though we have not fully tested and evaluated our approach, we believe that our initial findings are very promising to support our next steps. The systematic evaluation and validation of our system is among our future plans. We also hope to pave the way for the next generation disease surveillance systems.

Acknowledgements

The project is part of the Electronic Disease Surveillance Monitoring Network (EDMON), led by the University of Tromsø – The Arctic University of Norway and the Norwegian Centre for eHealth Research, University Hospital of North Norway.

References

[1] Adam, S., Bernadette, J., & David, B. (2007). Health Surveillance and Diagnosis for Mitigating a Bioterror Attack. *LINCOLN LABORATORY JOURNAL*, *17*(1), 101-113.

[2] Arsand, E., Tatara, N., Ostengen, G., & Hartvigsen, G. (2010). Mobile phone-based self-management tools for type 2 diabetes: the few touch application. *J Diabetes Sci Technol*, *4*(2), 328-336. doi:doi: 10.1177/193229681000400213

[3] Årsand, E., Walseth, O., Andersson, N., Fernando, R., Granberg, O., Bellika, J., & Hartvigsen, G. (2005). Using blood glucose data as an indicator for epidemic disease outbreaks. . *Studies in Health Technology and Informatics*, *116*, 217-222.

[4] Benhamou, P. Y. (2011). Improving diabetes management with electronic health records and patients' health records. *Diabetes & Metabolism*, *37*, S53-S56. doi:10.1016/s1262-3636(11)70966-1

[5] Botsis, T., Bellika, J. G., & Hartvigsen, G. (2009). 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, 176-183. doi:10.1109/eTELEMED.2009.9

[6] Botsis, T., & Hartvigsen, G. (2010). Exploring new directions in disease surveillance for people with diabetes: lessons learned and future plans. *Stud Health Technol Inform, 160*(Pt 1), 466-470. doi:10.3233/978-1-60750-588-4-466

[7] Botsis, T., Hejlesen, O., Bellika, J. G., & Hartvigsen, G. (2007). Blood glucose levels as an indicator for the early detection of infections in type-1 diabetics. *Advances in Disease Surveillance, 4*, 147.

[8] Chatfield, C. (1993). Calculating Interval Forecasts. *Journal of Business & Economic Statistics, 11*(2), 121-135. doi:10.2307/1391361

[9] Issom, D.-Z., Woldaregay, A. Z., Chomutare, T., Bradway, M., Årsand, E., & Hartvigsen, G. (2015). Mobile applications for people with diabetes published between 2010 and 2015. *Diabetes Management, 5*(6), 539-550. doi:10.2217/dmt.15.40

[10] Lauritzen, J. N., Arsand, E., Van Vuurden, K., Bellika, J. G., Hejlesen, O. K., & Hartvig-sen, G. (2011). 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. *IEEE*, 1-5. doi:10.1109/wirelessvitae.2011.5940877

[11] Quinn, C. C., Clough, S. S., Minor, J. M., Lender, D., Okafor, M. C., & Gruber-Baldini, A. (2008). WellDoc mobile diabetes management randomized controlled trial: change in clinical and behavioral outcomes and patient and physician satisfaction. *Diabetes Technol Ther, 10*(3), 160-168. doi:10.1089/dia.2008.0283

[12] Uzedhe , G. O., Okeke, O. N., Inyiama , H. C., & Idigo , V. E. (2014). Multi-Point Time-Averaging Data Acquisition of Health Indicators: A Reliable Process for Patient Medical Support. *International journal of Science and Technology, 3*(5), 264-272 Retrieved from http://www.journalofsciencestechnology.org/archive/2014/may_vol_3_no_5/96212138969453.pdf

[13] Veinot, T. C., Zheng, K., Lowery, J. C., Souden, M., & Keith, R. (2010). Using Electronic Health Record Systems in Diabetes Care: Emerging Practices. *IHI, 2010*, 240-249. doi:10.1145/1882992.1883026

[14] Waki, K., Fujita, H., Uchimura, Y., Omae, K., Aramaki, E., Kato, S., . . . Ohe, K. (2014). DialBetics: A Novel Smartphone-based Self-management Support System for Type 2 Diabetes Patients. *J Diabetes Sci Technol, 8*(2), 209-215. doi:10.1177/1932296814526495

[15] Walseth, O., Arsand, E., Sund, T., & Skipenes, E. (2005). Wireless transfer of sensor data into electronic health records. *Stud Health Technol Inform., 116*, 334-339.

Address for correspondence

Ashenafi Zebene Woldaregay,
Department of Computer Science, University of Tromsø – The Arctic University of Norway, Tromsø, Norway.
Email: awo012@post.uit.no
Tel: +4748682581