LEARNING THE SENSING DELAY FOR PERSONALIZED CONTINUOUS DIABETES MONITORING

Subhamoy Biswas
University of Waterloo
s2biswas@uwaterloo.ca

Pouyan Keshavarz Motamed
University of Waterloo
pouyan.keshavarz@uwaterloo.ca

Peyman GhavamiNejad
University of Waterloo
pghavaminejad@uwaterloo.ca

Irfani Rahmi Ausri
University of Waterloo
irfani.ausri@uwaterloo.ca

Ali Etemad
Queen's University
ali.etemad@queensu.ca

Mahla Poudineh
University of Waterloo
mahla.poudineh@uwaterloo.ca

ABSTRACT

Interstitial fluid (ISF) serves as a rich source of biomarkers, enabling minimally invasive continuous health monitoring through ISF sensors. Despite their potential advantages, ISF sensors face a major challenge related to the delay in the transfer of target analytes from blood to ISF, compared to blood-based measurements. Particularly, this delay can vary significantly from subject to subject. While machine learning algorithms have been developed for continuous glucose measurement within ISF, these algorithms have not considered this delay. In this paper, we quantify the delay between ISF and blood-based detection of glucose and ketone bodies in diabetic rats using decision-tree based algorithms. Accounting for this delay can eventually improve the accuracy of ISF sensors for continuous health monitoring of individual patients.

1 INTRODUCTION

Interstitial fluid (ISF), the fluid underneath the skin, between blood vessels and body cells, is a rich source of health biomarkers. Wearable sensors for analyzing ISF and measuring these biomarkers offer a minimally invasive approach to continuous health monitoring (Saifullah & Rad, 2023). These wearable sensors which are increasingly becoming prevalent, can potentially replace invasive blood-based measurements (Madden et al., 2020). For example, continuous glucose monitoring (CGM) devices, a common ISF sensor, have transformed diabetes care by facilitating real-time glucose tracking (Teo et al., 2022). However, a major limitation of wearable ISF sensors is the presence of a delay between the actual concentration of biomarkers in blood and the concentration recorded by the ISF sensors (Rossetti et al., 2010). The delay is mainly caused by variations in transport efficiencies between the ISF and the blood in circulation (Keenan et al., 2009). Studies have shown a 5 to 20-minute delay in detecting changes in blood glucose compared to the measurements performed by CGM (Basu et al., 2015). Yet, the delay for other crucial biomarkers such as ketone bodies and insulin remains unknown. Furthermore, the delay can vary among individuals, emphasizing the need to determine personalized delays. To successfully incorporate continuous wearable ISF monitoring into clinical practice, it is essential to be able to understand the delay between ISF and blood levels for clinically important analytes in a precise and personalized manner.

Over the past years, several machine learning-guided platforms were developed to analyze a patient's ISF glucose information obtained from CGM devices (Vettoretti et al., 2020). These ISF-based glucose models explored the classification of glycemic control profiles for alerting patients of the risk of events like post-prandial hypoglycemia (glucose < 72mg/dl) (Cappon et al., 2019; Reddy et al., 2019; Vehí et al., 2020) and introduced time series models that predict glucose levels for the next 15 minutes up to 1 hour. These models used both classical statistical methods like autoregressive models (Cobelli, 2007) as well as more advanced artificial neural network frameworks (Alfian et al., 2020), long short-term memory networks (Martinsson et al., 2020), Gaussian processes (Georga et al., 2015), support vector machines (Georga et al., 2013; Xie1 & Wang, 2018) and transformer-inspired encoder-decoder models (Armandpour et al., 2021; Sergazinov et al., 2023) to name a few.

However, these approaches have not accounted for the delays between the blood and ISF measurements and are limited to glucose monitoring. Recently, we developed a new wearable platform for continuous monitoring of glucose and ketone bodies (called CGM-CKM device) in ISF. Specifically, patients with type 1 diabetes need to track their ketone levels in addition to their glucose to avoid the risk of diabetes ketoacidosis (DKA). A personalized knowledge of the delay between the levels of glucose and ketones in blood and ISF for individual patients improves the performance of this real-time monitoring device, leading to a better quality of life for people living with diabetes.

In this paper, we used decision-tree based algorithms to quantify the delay between the ISF-based detection from the CGM-CKM device and blood-based measurement of glucose and ketone levels in different diabetic rats. We achieved this by applying three different boosted decision-tree algorithms that estimated an ISF-level time series corresponding to a reference blood-based time series (Figure 1). We cross-correlated the predicted ISF response from the device against the blood levels and estimated a delay specific to each tested animal. Determining the delays between ISF and blood will enable auto-adjustment of the CGM-CKM device for future human studies.

We aim to share our data and results with the workshop participants and collectively take steps towards improving personalized health monitoring using wearable devices.

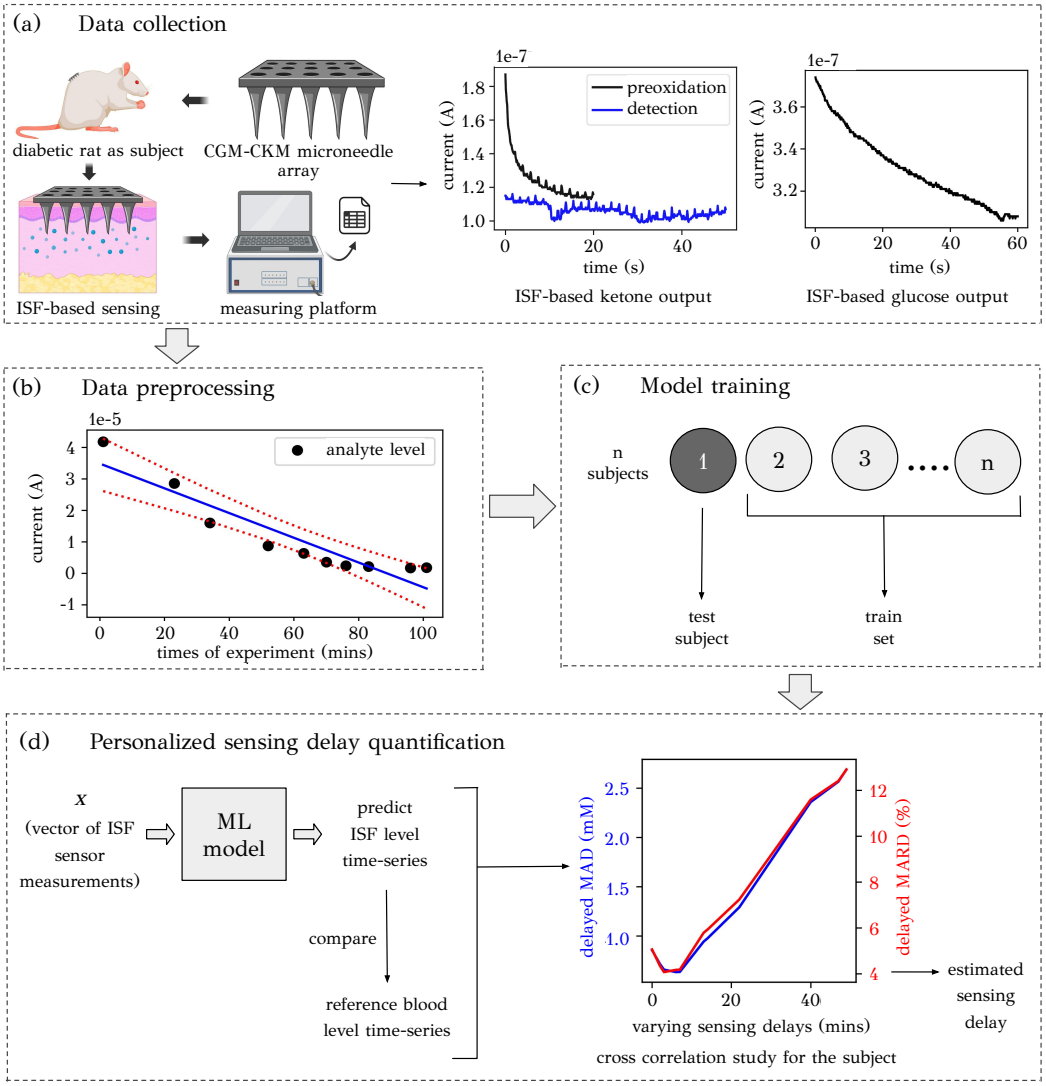


Figure 1: Overview of the protocol: (a) ISF-based data collection from CGM-CKM device, (b) data pre-processing to remove outliers, (c) leave-one-subject-out model training scheme, and (d) evaluating personalized sensing delays for each rat using blood-based measurements as reference.

2 METHODOLOGY

2.1 PROBLEM FORMULATION

Let $n \in \mathbb{N}$ be the number of tested subjects in the dataset for glucose or ketone measurements. We defined $[n] := \{1, 2, \ldots, n\}$. Let k indicate any subjects chosen randomly from the set $[n]$ and m_k be the number of times that subject k was tested. A dataset with examples of the form $\mathcal{D} = \{(\mathbf{x}_{i,k}, y_{i,k})\}_{i\in[m_k],k\in[n]}$ was defined. $\mathbf{x}_{i,k} \in \mathbb{R}^s$ was the vector of all input features extracted from the CKM or CGM sensor platform, $s \in \mathbb{N}$ was the number of features in $\mathbf{x}_{i,k}$, and $y_{i,k} \in \mathbb{R}$ indicated the corresponding output analyte concentration. Here, the primary objectives were: (a) to construct a mapping (or function) $F_k : \mathbb{R}^s \to \mathbb{R}$ for every individual k that learns from the vector of input features from CKM or CGM sensor $\mathbf{x}_{i,k}$ to make an estimation for the ketone or glucose levels, respectively, in blood $y_{i,k}$, and (b) to find an optimal delay τ_k by which the subject k's predicted ISF levels $F_k(\mathbf{x}_{i,k}; \theta_k)$ should be delayed to match with its corresponding blood concentration $y_{i,k}$ as close as possible. θ_k was the set of learned parameters of the model F_k for subject k.

2.2 QUANTIFYING THE DELAY BETWEEN BLOOD AND ISF

To learn the personalized ISF prediction models F_k and consequently find the delay τ_k, we considered a leave-one-subject-out scheme, where, at a time, we fixed only 1 subject (i.e., k). For this k, let k^* represent any other subject in the dataset $\mathcal{D}$ other than k, that is, $k^* \in [n]\backslash\{k\}$. Next, we divided $\mathcal{D}$ into two partitions - a training dataset $\mathcal{D}^k_{train}$ and a test dataset $\mathcal{D}^k_{test}$. We considered $\mathcal{D}^k_{train} = \mathcal{D}^{\backslash k} = \{(\mathbf{x}_{i,k^*}, y_{i,k^*})\}_{i\in[m_{k^*}],k^*\in[n]\backslash\{k\}}$ and $\mathcal{D}^k_{test} = \mathcal{D}^k = \{(\mathbf{x}_{i,k}, y_{i,k})\}_{i\in[m_k]}$. It should be noted that $\mathcal{D}^k_{train}$ and $\mathcal{D}^k_{test}$ were mutually exhaustive and disjoint. We then used all the examples in $\mathcal{D}^k_{train}$ to train F_k using a learner algorithm. It should be noted that F_k varies for every k based on the number of trees and loss function, emphasizing on personalized predictions.

To test our model F_k and find the delay for subject k, we defined two time series. Let $\mathbf{y}^{blood}_k := \{y_{i,k}\}_{i\in[m^k]}$ be the series of blood-based levels of k, to be used as ground truth for comparison with ISF levels. Let $\mathbf{y}^{ISF}_k := \{F_k(\mathbf{x}_{i,k}; \theta_k)\}_{i\in[m^k]}$, be the series of the ISF-level predictions for k. Let $\mathbf{t}_k = \{t_{i,k}\}_{\forall i\in[m^k]}$ be the set of all time instants at which an ISF experiment was performed for k. For positive scalar $\tau \in \mathbb{R}^+$, we defined an operation, $\mathbf{t}_k + \tau := \{t_{i,k} + \tau\}_{i\in[m^k]}$ (element-wise addition). By definition of ISF delays, $\mathbf{y}^{ISF}_k$ ideally should be a delayed version of $\mathbf{y}^{blood}_k$. To find this delay, first, we linearly interpolated $\mathbf{y}^{ISF}_k$ with respect to $\mathbf{t}_k$. From the interpolated curve for ISF response, we re-sampled it, but instead, at times $\mathbf{t}_k + \tau$, such that $\tau \in [\tau_{max}] \cup \{0\}$, τ_{max} being the physiologically acceptable limit in minutes for ISF delays. By re-sampling at the modified time instants, the original ISF response $\mathbf{y}^{ISF}_k$ was now delayed by τ minutes to give us a delayed response - we labeled it as $\mathbf{y}^{ISF}_{k,\tau}$. We chose τ_{max}, the maximum limit of any delay to be 50 min. Finally, we compared $\mathbf{y}^{ISF}_{k,\tau}$ with $\mathbf{y}^{blood}_k$ and simultaneously, varied τ in $[\tau_{max}] \cup \{0\}$ to find the optimal delay $\tau = \tau_k$ for subject k that minimizes the error metrics (see Section 3.2 for more information). We propose that this τ_k is the delay during the transmission of bioalaytes from blood to ISF in subject k. If $\tau_k = 0$, it means transport of the analyte from blood to ISF takes place in negligible time, whereas if $\tau_k > 0$, it means transmission in k requires a significantly higher duration.

3 EXPERIMENTS

3.1 DATA COLLECTION

Glucose and ketone sensing using CGM-CKM device: For both ketone and glucose monitoring, the number of diabetic rats tested was $n = 4$. In case of CKM, the device output for a certain subject k at every time of experiment $t_{i,k}$ consisted of two time series - a pre-oxidation current $\mathbf{p}_{i,k}$ representing the amount of ketone recognition element present at the start of ketone detection, and a detection current $\mathbf{d}_{i,k}$, indicating the amount of recognition element left after detection. We annotated the 4 subjects for ketone sensing as - Rat 1, Rat 2, Rat 3 and Rat 4 as in Table 1. Figure 3(a) (see Appendix) gives a sample pair of $\mathbf{p}_{i,k}$ and $\mathbf{d}_{i,k}$ profiles for Rat 2 ($k = 2$ for ketone sensor) at time $t_{i,2} = 93$ min. The profiles were recorded every $0.1s$. $\mathbf{p}_{i,k}$ was measured up to $20s$ and $\mathbf{d}_{i,k}$ up to $50s$ for every k.

For CGM, the device output for each subject k had two steps. First, a time-varying current $\mathbf{g}_{i,k}$ indicating the glucose levels in k at time $t_{i,k}$ recorded every $0.1s$ for up to $60s$, and second, a

Table 1: a summary of the best possible individual delays τ_k and their corresponding metrics

Target	Rat	MAD(mM) / MARD(%)			τ_k (mins)		
		Grad. Boost	CatBoost	XGBoost	Grad. Boost	CatBoost	XGBoost
Ketone	Rat 1	0.100 / 7.714	0.164 / 10.982	0.192 / 12.803	31	23	18
	Rat 2	0.144 / 8.500	0.184 / 9.739	0.114 / 10.254	9	11	1
	Rat 3	0.352 / 20.953	0.376 / 24.542	0.541 / 35.701	15	7	11
	Rat 4	0.246 / 6.379	0.328 / 7.109	0.254 / 6.162	29	44	31
Glucose	Rat 5	4.147 / 18.849	3.643 / 20.966	4.000 / 23.428	0	0	0
	Rat 6	1.208 / 13.502	3.104 / 32.851	4.384 / 45.599	13	9	9
	Rat 7	1.143 / 5.378	0.799 / 3.457	2.497 / 12.374	22	19	10
	Rat 8	0.661 / 4.075	0.682 / 4.477	0.655 / 4.070	3	10	10

Table 2: the following table compares the overall $MARD$ and MAD for all the algorithms

Target	Algorithms used	Performance metrics	
		Overall MAD (mM)	Overall MARD (%)
Ketone	Gradient Boosting	0.2261	10.8887
	XGBoost	0.2915	16.3618
	CatBoost	0.2849	13.4851
Glucose	Gradient Boosting	1.5272	8.3838
	XGBoost	2.4439	16.1028
	CatBoost	1.5588	10.6467

current-voltage loop for every k to interpret the scale of sensor response in k and thereby calibrate $\mathbf{g}_{i,k}$. We annotated the 4 animal subjects on which the glucose sensing was performed as - Rat 5, Rat 6, Rat 7 and Rat 8 as in Table 1. Figure 3(c) and (d) (see Appendix) shows a sample current-voltage loop and its corresponding $\mathbf{g}_{i,k}$ profile. Because the $\mathbf{g}_{i,k}$ current ranges varied from one in vivo rat to another, they were brought to the same level by defining an adjustment factor. For this, from the current-voltage loop for every k, we recorded the average current value corresponding to the maximum voltage level $0.6V$ for every rat. Let this sampled current be $G_{0.6,k}$. Choosing $G_{0.6,1}$ for the Rat 5 ($k = 1$ for glucose) as the reference, we divided the $G_{0.6,k}$ values for all the other rats 6, 7 and 8 ($k = 2, 3, 4$ for glucose respectively) by $G_{0.6,1}$ to find their individual scaling, or adjustment factors. We later used these factors to re-scale their individual $\mathbf{g}_{i,k}$ profiles to a comparable range.

Data preprocessing and input features: The varying physiological conditions inside each diabetic rat incorporated noise into the $\mathbf{p}_{i,k}$ and $\mathbf{d}_{i,k}$ profiles during ketone sensing. As such, we used a moving average filter to smooth both the current profiles. The $\mathbf{g}_{i,k}$ profiles for glucose however had negligible noise and did not require any such low-pass filters. To define the input features in $\mathbf{x}_{i,k}$, we take samples from the output current profiles from the CGM-CKM platform. For ketone, we sampled $\mathbf{p}_{i,k}$ and $\mathbf{d}_{i,k}$ profiles at their end points. Let these samples be $P_{i,k}$ and $D_{i,k}$ respectively. For glucose, we sampled $\mathbf{g}_{i,k}$ at $10s$ and $40s$. Let these samples be $g_{i,k,10}$ and $g_{i,k,40}$. For CKM, the ratio $\frac{D_{i,k}}{P_{i,k}}$ indicates the percentage of recognition element left after detecting ketone bodies - directly proportional to ketone concentration in subject k. Administration of insulin doses as a remedial action against DKA during our ISF experiments ensured a drop in ketone levels with time and hence, $\frac{D_{i,k}}{P_{i,k}}$ ideally should decrease monotonically with time $t_{i,k}$. Similarly for CGM, $g_{i,k,10}+g_{i,k,40}$ being directly proportional to glucose levels, should also decrease with time. Presence of outliers often disrupts these trends, and so, we used a robust framework to remove such outliers (see Appendix and Figure 4 for more details). Finally, for ketone, we considered $\mathbf{x}_{i,k} = [P_{i,k} \quad D_{i,k} \quad t_{i,k}]$ and for glucose, $\mathbf{x}_{i,k} = [g_{i,k,10} + g_{i,k,40} \quad t_{i,k}]$.

3.2 Performance metrics

To test the performance of F_k for subject k, we applied it on an exclusively new set of test examples $\mathcal{D}^k_{test}$. To compare $\mathbf{y}^{ISF}_k$ with $\mathbf{y}^{blood}_k$, we considered two metrics, mean absolute difference (absolute error), MAD_k and mean absolute relative difference (percentage error), $MARD_k$. Both of these metrics are widely-accepted to measure the accuracy of continuous monitoring biosensors, especially for glucose (Heinemann et al., 2019; Zueger et al., 2012). While comparing delayed ISF $\mathbf{y}^{ISF}_{k,\tau}$ with $\mathbf{y}^{blood}_k$, we annotated the metrics as $MAD_{k,\tau}$ and $MARD_{k,\tau}$ respectively. For every k, we varied τ to find the optimal delay $\tau = \tau_k$ that minimizes $MARD_{k,\tau}$ and $MAD_{k,\tau}$ between $\mathbf{y}^{ISF}_{k,\tau}$ and ground truth $\mathbf{y}^{blood}_k$. For CKM, we were also interested if the model can differentiate between normal ketosis ($F_k(\mathbf{x}_{i,k}; \theta_k) \leq 1.5mM$) and hyperketonemia ($F_k(\mathbf{x}_{i,k}; \theta_k) > 1.5mM$). Hence, for ketone, we considered a third metric, $mistakes_{k,\tau}$, that counts the number of misclassified examples in the binary classification between normal ketosis and hyperketonemia (see Appendix).

3.3 RESULTS AND DISCUSSION

Figure 1 represents an overall schematic of our approach. To learn F_k for every k, we applied 3 different decision-tree based algorithms - gradient boosting (GBA) (Friedman, 2001), CatBoost (Prokhorenkova et al., 2019) and XGBoost (Chen & Guestrin, 2016). We computed the performance metrics, MAD and $MARD$, for different values of τ with 1min interval for both CKM and CGM and chose the optimum delay that minimized MAD (see Figure 5 in Appendix for sample plots). Table 1 lists the estimated delays for each rat for different learning algorithms. For individual rats, varying delays ranging from 9 – 44 mins for CKM device and from 0 - 22 min for CGM device, were found. The differences between the delays in individual rats clearly established the variability in ketone and glucose kinetics, even under controlled conditions with genetically similar animals. This has important implications for human patients, who are genetically diverse and have different environmental conditions, intensifying this inter-individual variability. At the same time, the delays τ_k obtained for every rat k for both CKM and CGM for all the 3 algorithms were within the same range, indicating the consistency of our frameworks in quantifying delays for the CGM-CKM device. For both CGM and CKM, we combined the delayed ISF predictions and the reference blood values for all subjects and determined the overall MAD and $MARD$ for each algorithm as in Table 2, out of which GBA out-performed other algorithms for both analytes. We further correlated the delayed ISF predictions obtained from GBA with the blood references for both CKM and CGM (Figure 2). Correlation values of 0.941 and 0.790 were obtained for CKM and CGM devices, respectively, when compared with blood-based measurements. Comparing the calculated MAD and MARD for CGM (Table 3, Appendix) and CKM (Table 4, Appendix) devices with the reported glucose and ketone studies showed similar performance.

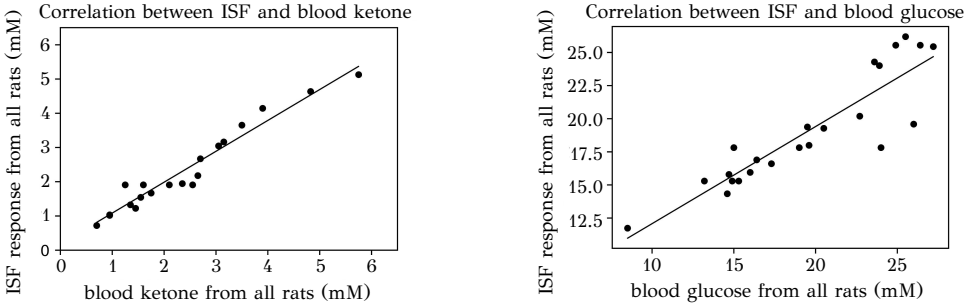


Figure 2: Overall correlation for CKM (left) and CGM (right).

4 CONCLUSION

In summary, our work enables improving the accuracy of wearable sensors for continuous glucose and ketone monitoring via quantifying physiological delays between ISF and blood compartments. We are currently expanding our data collection efforts with the CGM-CKM device to obtain longer time-series sequences from a larger number of rats. These collected measurements will be used to enhance the precision of our algorithms for assessing the delay and develop new representation learning algorithms for time-series data obtained from wearable devices. Subsequently, we plan to assess the CGM-CKM device in human patients and utilize the developed predictive model to measure delay in human subjects.

5 LIMITATIONS

Our work comes with a few limitations. First, the evaluation was limited to only 4 diabetic rats per analyte and our aim is to increase the number in future. Secondly, we quantified the personalized sensing delays using only decision-tree based algorithms. We aim at expanding this to other state-of-the-art machine learning algorithms to improve the generalizability of the work.

ACKNOWLEDGEMENT

This research was supported by the Juvenile Diabetes Research Foundation (JDRF-2-SRA-2022-1167-M-B).

REFERENCES

Felix Aberer, Martin Hajnsek, Markus Rumpler, Sabine Zenz, Petra M. Baumann, Hesham Elsayed, Adelheid Puffing, Gerlies Treiber, Thomas R. Pieber, Harald Sourij, and Julia K. Mader. Evaluation of subcutaneous glucose monitoring systems under routine environmental conditions in patients with type 1 diabetes. *Diabetes, Obesity and Metabolism*, 19(7):1051–1055, 2017.

Ganjar Alfian, Muhammad Syafrudin, Muhammad Anshari, Filip Benes, Fransiskus Tatas Dwi Atmaji, Imam Fahrurrozi, Ahmad Fathan Hidayatullah, and Jongtae Rhee. Blood glucose prediction model for type 1 diabetes based on artificial neural network with time-domain features. *Biocybernetics and Biomedical Engineering*, 40(4):1586–1599, 2020.

Shridhara Alva, Kristin Castorino, Hyun Cho, and Junli Ou. Feasibility of continuous ketone monitoring in subcutaneous tissue using a ketone sensor. *Journal of Diabetes Science and Technology*, 15(4):768–774, 2021.

Mohammadreza Armandpour, Brian Kidd, Yu Du, and Jianhua Z. Huang. Deep personalized glucose level forecasting using attention-based recurrent neural networks. *IEEE IJCNN*, 2021.

Ananda Basu, Simmi Dube, Sona Veettil, Michael Slama, Yogish C. Kudva, Thomas Peyser, Rickey E. Carter, Claudio Cobelli, and Rita Basu. Time lag of glucose from intravascular to interstitial compartment in type 1 diabetes. *Journal of Diabetes Science and Technology*, 9(1): 63–68, 2015.

Giacomo Cappon, Andrea Facchinetti, Giovanni Sparacino, Pantelis Georgiou, and Pau Herrero. Classification of postprandial glycemic status with application to insulin dosing in type 1 diabetes—an in silico proof-of-concept. *Sensors*, 19, 2019.

Tianqi Chen and Carlos Guestrin. Xgboost: A scalable tree boosting system. *KDD '16: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 785–794, 2016.

Mark P. Christiansen, Satish K. Garg, Ronald Brazg, Bruce W. Bode, Timothy S. Bailey, Robert H. Slover, Ashley Sullivan, Suiying Huang, John Shin, Scott W. Lee, and Francine R. Kaufman. Accuracy of a fourth-generation subcutaneous continuous glucose sensor. *Diabetes Technology and Therapeutics*, 19(8):446–456, 2017.

Mark P. Christiansen, Leslie J. Klaff, Ronald Brazg, Anna R. Chang, Carol J. Levy, David Lam, Douglas S. Denham, George Atiee, Bruce W. Bode, Steven J. Walters, Lynne Kelley, and Timothy S. Bailey. A prospective multicenter evaluation of the accuracy of a novel implanted continuous glucose sensor: Precise ii. *Diabetes Technology and Therapeutics*, 20(3):197–206, 2018.

Giovanni Sparacino; Francesca Zanderigo; Stefano Corazza; Alberto Maran; Andrea Facchinetti; Claudio Cobelli. Glucose concentration can be predicted ahead in time from continuous glucose monitoring sensor time-series. *IEEE Transactions on Biomedical Engineering*, 54(5):931–937, 2007.

Edward R. Damiano, Firas H. El-Khatib, Hui Zheng, David M. Nathan, and Steven J. Russell. A comparative effectiveness analysis of three continuous glucose monitors. *Diabetes Care*, 36(2): 251–259, 2013.

Ketan K. Dhatariya, Nicole S. Glaser, Ethel Codner, and Guillermo E. Umpierrez. Diabetic ketoacidosis. *Nature Reviews Disease Primers*, 6, 2020.

Jerome H. Friedman. Greedy function approximation: a gradient boosting machine. *Annals of statistics*, pp. 1189–1232, 2001.

Satish K. Garg, Mark Kipnes, Kristin Castorino, Timothy S. Bailey, Halis Kaan Akturk, John B. Welsh, Mark P. Christiansen, Andrew K. Balo, Sue A. Brown, Jennifer L. Reid, and Stayce E. Beck. Accuracy and safety of dexcom g7 continuous glucose monitoring in adults with diabetes. *Diabetes Technology and Therapeutics*, 24(6):373–380, 2022.

Eleni I. Georga, Vasilios C. Protopappas, Diego Ardigo, Michela Marina, Ivana Zavaroni, Demosthenes Polyzos, and Dimitrios I. Fotiadis. Multivariate prediction of subcutaneous glucose concentration in type 1 diabetes patients based on support vector regression. *Journal of Biomedical and Health Informatics*, 17(1), 2013.

Eleni I. Georga, Vasilios C. Protopappas, Demosthenes Polyzos, and Dimitrios I. Fotiadis. Evaluation of short-term predictors of glucose concentration in type 1 diabetes combining feature ranking with regression models. *Medical and Biological Engineering and Computing*, 53:1305–1318, 2015.

Lutz Heinemann, Michael Schoemaker, Günther Schmelzeisen-Redecker, Rolf Hinzmann, Adham Kassab, Guido Freckmann, Florian Reiterer, and Luigi Del Re. Benefits and limitations of mard as a performance parameter for continuous glucose monitoring in the interstitial space. *Journal of Diabetes Science and Technology*, 14(1):135–150, 2019.

D. Barry Keenan, John J. Mastrototaro, Gayane Voskanyan, and Garry M. Steil. Delays in minimally invasive continuous glucose monitoring devices: A review of current technology. *Journal of Diabetes Science and Technology*, 3(5):1207–1214, 2009.

Julia Madden, Conor O'Mahony, Michael Thompson, Alan O'Riordan, and Paul Gavin. Biosensing in dermal interstitial fluid using microneedle based electrochemical devices. *Sensing and Bio-Sensing Research*, 29, 2020.

John Martinsson, Alexander Schliep, Björn Eliasson, and Olof Mogren. Blood glucose prediction with variance estimation using recurrent neural networks. *Journal of Healthcare Informatics Research*, 4:1–18, 2020.

Stefan Pleus, Michael Schoemaker, Karin Morgenstern, Günther Schmelzeisen-Redeker, Cornelia Haug, Manuela Link, Eva Zschornack, and Guido Freckmann. Rate-of-change dependence of the performance of two cgm systems during induced glucose swings. *Journal of Diabetes Science and Technology*, 9(4):801–807, 2015.

Liudmila Prokhorenkova, Gleb Gusev, Aleksandr Vorobev, Anna Veronika Dorogush, and Andrey Gulin. Catboost: unbiased boosting with categorical features. 2019.

Ravi Reddy, Navid Resalat, Leah M. Wilson, Jessica R. Castle, Joseph El Youssef, and Peter G. Jacobs. Prediction of hypoglycemia during aerobic exercise in adults with type 1 diabetes. *Journal of Diabetes Science and Technology*, 13(5):919–927, 2019.

Paolo Rossetti, Jorge Bondia, Josep Vehí, and Carmine G. Fanelli. Estimating plasma glucose from interstitial glucose: The issue of calibration algorithms in commercial continuous glucose monitoring devices. *Sensors*, 10(12), 2010.

Khaled Mohammed Saifullah and Zahra Faraji Rad. Sampling dermal interstitial fluid using microneedles: A review of recent developments in sampling methods and microneedle-based biosensors. *Advanced Materials Interfaces*, 10(10), 2023.

Renat Sergazinov, Mohammadreza Armandpour, and Irina Gaynanova. Gluformer: Transformer-based personalized glucose forecasting with uncertainty quantification. *IEEE International Conference on Acoustics, Speech, and Signal Processing*, 2023.

Evelyn Teo, Norasyikin Hassan, Wilson Tam, and Serena Koh. Effectiveness of continuous glucose monitoring in maintaining glycaemic control among people with type 1 diabetes mellitus: a systematic review of randomised controlled trials and meta-analysis. *Diabetologia*, 65:604–619, 2022.

Josep Vehí, Iván Contreras, Silvia Oviedo, Lyvia Biagi, and Arthur Bertachi. Prediction and prevention of hypoglycaemic events in type-1 diabetic patients using machine learning. *Health Informatics Journal*, 26(1):703–718, 2020.

Martina Vettoretti, Giacomo Cappon, Andrea Facchinetti, and Giovanni Sparacino. Advanced diabetes management using artificial intelligence and continuous glucose monitoring sensors. *Sensors*, 20, 2020.

R. Paul Wadwa, Lori M. Laffel, Viral N. Shah, and Satish K. Garg. Accuracy of a factory-calibrated, real-time continuous glucose monitoring system during 10 days of use in youth and adults with diabetes. *Diabetes Technology and Therapeutics*, 20(6):395–402, 2018.

John B. Welsh, Francine R. Kaufman, and Scott W. Lee. Accuracy of the sof-sensor glucose sensor with the ipro calibration algorithm. *Journal of Diabetes Science and Technology*, 6(2), 2012.

Jinyu Xie1 and Qian Wang. Benchmark machine learning approaches with classical time series approaches on the blood glucose level prediction challenge. *CEUR-WS*, 2148, 2018.

Thomas Zueger, Peter Diem, Stavroula Mougiakakou, and Christoph Stettler. Influence of time point of calibration on accuracy of continuous glucose monitoring in individuals with type 1 diabetes. *Diabetes Technology and Therapeutics*, 14(7), 2012.

APPENDIX

In this article, we explore the challenge of evaluating biosensor delays using the case study of a subject suffering from diabetic ketoacidosis (DKA). DKA is a potentially is a life-threatening complication of diabetes characterized by insulin deficiency, subsequently leading to hyperglycemia, increased levels of ketosis, and an overall metabolic acidosis, resulting in symptoms like electrolyte imbalance and potential organ failure in human bodies (Dhatariya et al., 2020). For this problem, we primarily focus on the continuous monitoring of two bioanalytes - ketone bodies and glucose.

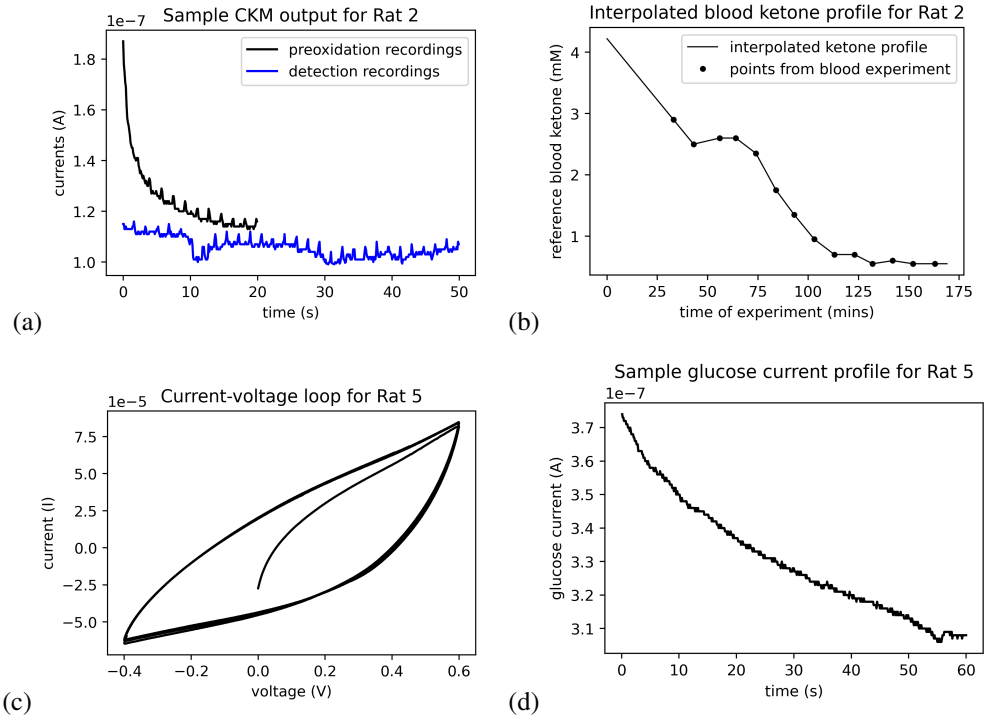


Figure 3: (a): a sample CKM sensor output - the black and blue curves indicate a sample preoxidation and detection current profiles respectively; (b) the corresponding blood ketone interpolation graph; (c): a sample CGM current-voltage loop to interpret the adjustment factors for each rat; (d): the corresponding glucose current profile

Figure 3 shows a sample overview of the CGM-CKM device output. In case of CKM ((a) and (b)), due to experimental restrictions, the blood-based measurements were performed at a different set of time of experiments for each subject k in comparison to ISF experiments. Let these time instants for blood ketone recordings be $\hat{\mathbf{t}}_k = \{\hat{t}_{j,k}\}_{j\in[m_k^{blood}]}$ (in minutes), m_k^{blood} being the number of times blood based ketone measurements taken for subject k. In other words, cardinality of $\mathbf{t}_k \cup \hat{\mathbf{t}}_k > m_k$ and $> m_k^{blood}$, for every subject k. To solve this disparity, the blood ketone levels, at first, were linearly interpolated with respect to $\hat{\mathbf{t}}_k$, as in Figure 3(b). From this curve, the blood ketone values were then re-sampled, but instead, at times $\mathbf{t}_k$, for comparison with our ISF-based ketone predictions. Unlike the ketone measurements, in case of CGM ((c) and (d)), the reference blood glucose levels were recorded at the same instants as the times of ISF experiments, that is, for our glucose experiments, $\mathbf{t}_k = \hat{\mathbf{t}}_k$ and $m_k = m_k^{blood}$.

Having interpreted the CGM-CKM device outputs, we then proceeded to remove any outliers from our sampled data points for both the analytes. To remove outliers for CKM, we first fixed any rat k and and performed a linear fit for $\left\{\frac{D_{i,k}}{P_{i,k}}\right\}_{i\in[m_k]}$ versus $\mathbf{t}_k$ as in Figure 4 and rejected those ratios and their time instants which lied outside the 95% confidence interval (shown in red) of the linear fit. For CGM, we performed a similar task using a linear fit for $\{g_{i,k,10} + g_{i,k,40}\}_{i\in[m_k]}$ versus $\mathbf{t}_k$

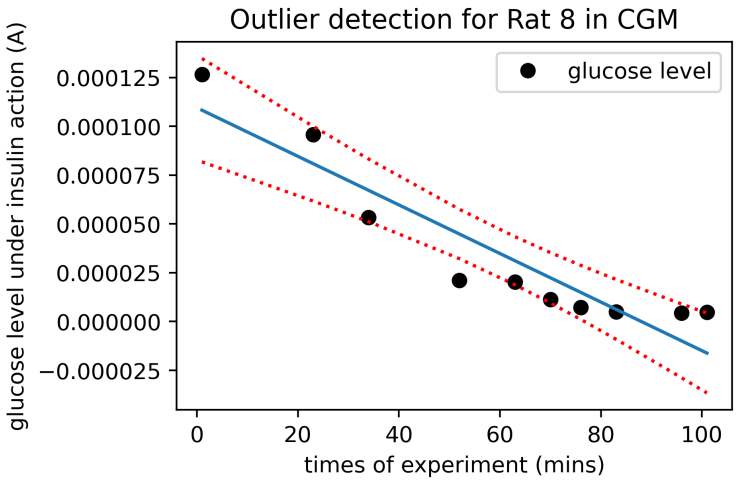


Figure 4: A sample plot for outlier detection.

Table 3: the following table summarizes the MARD performance of some widely-used CGM systems implemented on human subjects

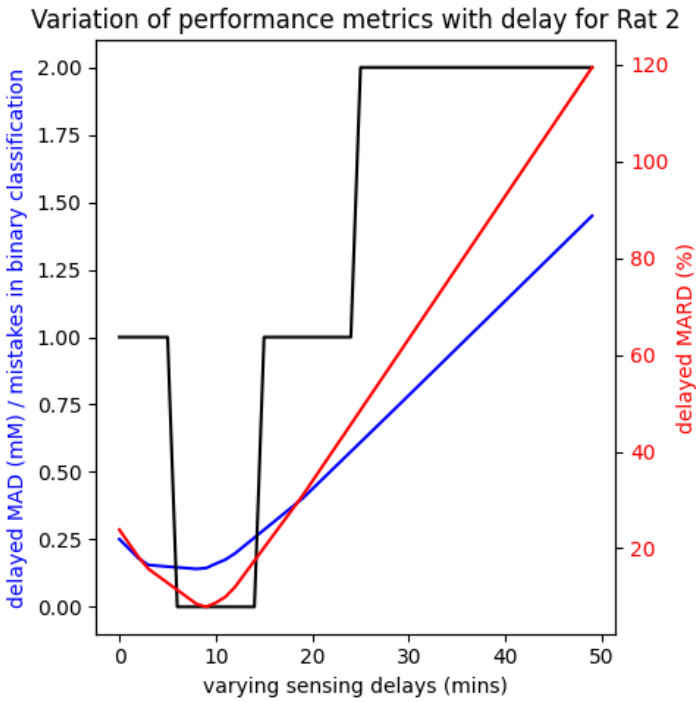
CGM system used	Reference	Number of subjects	Overall MARD
Medtronic SofSensor	Welsh et al. (2012)	Adults: 71, children: 61	Adults: 9.9%, children: 10%
Abbott Navigator I	Damiano et al. (2013)	6	11.8%
Dexcom G4 Platinum	Pleus et al. (2015)	10	10.9%
Abbott Freestyle Libre	Aberer et al. (2017)	12	13.2%
Medtronic Guardian Sensor 3	Christiansen et al. (2017)	88	9.6%
Senseonics Eversense	Christiansen et al. (2018)	90	8.8%
Dexcom G6	Wadwa et al. (2018)	262	10%
Dexcom G7	Garg et al. (2022)	318	8.2%

and defining similar confidence boundaries for each rat k to remove the outliers. Following this, we moved on to train F_k for every rat k for both CGM-CKM.

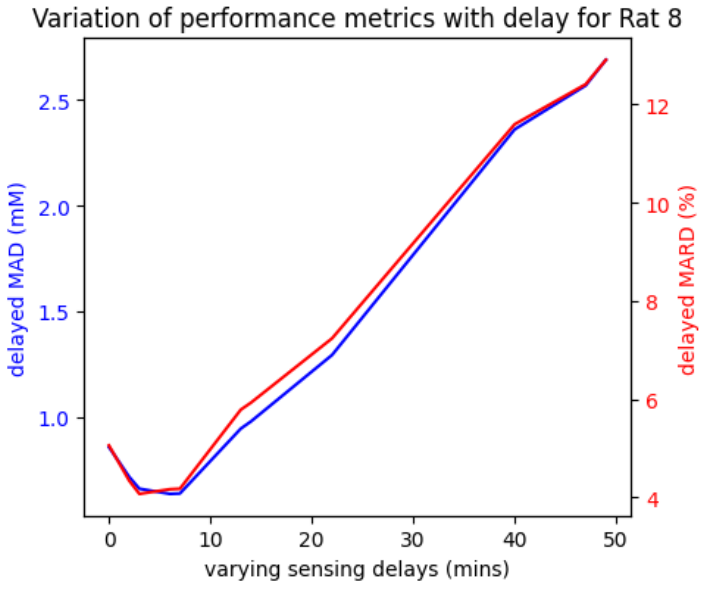
While testing the performance of F_k, we used metrics $MARD_k$ and MAD_k for both CGM and CKM and the metric $mistakes_k$ specifically for CKM. Here, $MARD_k = \frac{1}{m_k}\sum_{i=1}^{m_k}\left|\frac{y_{i,k}-F_k(\mathbf{x}_{i,k};\theta_k)}{y_{i,k}}\right|$ computes the mean absolute relative difference between ISF predictions $\mathbf{y}_k^{ISF}$ and blood reference $\mathbf{y}_k^{blood}$. We then varied τ to find the error $MARD_{k,\tau}$ between the delayed ISF $\mathbf{y}_{k,\tau}^{ISF}$ and $\mathbf{y}_k^{blood}$ and determine the optimal delay τ_k. This is shown as red curves in Figure 5. Similarly, we had $MAD_k = \frac{1}{m_k}\sum_{i=1}^{m_k}|y_{i,k} - F_k(\mathbf{x}_{i,k};\theta_k)|$ and for ketone, $mistakes_k = \sum_{i=1}^{m_k}\left(\mathbf{1}_{\mathrm{y_{i,k}>1.5}} \neq \mathbf{1}_{\mathrm{F_k(x_{i,k};\theta_k)>1.5}}\right) \vee \left(\mathbf{1}_{\mathrm{y_{i,k}\leq 1.5}} \neq \mathbf{1}_{\mathrm{F_k(x_{i,k};\theta_k)\leq 1.5}}\right)$, where $\vee$ indicates the logical "OR" operation. The variations of $MAD_{k,\tau}$ and $mistakes_{k,\tau}$ with τ are shown in blue and black curves respectively in Figure 5.

Table 4: the following table compares the overall MAD and $MARD$ obtained while implementing our ketone sensing framework with a state-of-the-art study (Alva et al., 2021)

Condition	Alva et al. (2021)	Our approach
overall MAD for ketone levels < 1.5mM	**0.129**	0.1799
overall $MARD$ for ketone levels ≥ 1.5mM	14.4%	**10.1083%**



(a)



(b)

Figure 5: Sample plots for variation of metrics with delays for (a) CKM and, (b) CGM