Detection of glycemic excursions using morphological and time-domain ECG features

Kathan Vyas[1], Carolina Villegas[4], Elizabeth Kubota-Mishra[4], Darpit Dave[2], Madhav Erraguntla[2], Gerard Coté[3], Daniel J. DeSalvo[4], Siripoom McKay[4] and Ricardo Gutierrez-Osuna[1]

[1]*Department of Computer Science and Engineering, Texas A&M University*
[2]*Department of Industrial and Systems Engineering, Texas A&M University*
[3]*Department of Biomedical Engineering, Texas A&M University*
[4]*Pediatrics, Diabetes and Endocrinology, Baylor College of Medicine*

Abstract*—Managing diabetes often involves monitoring blood glucose in real time to detect excursions (e.g., hypoglycemia and hyperglycemia). Continuous glucose monitors (CGMs) are generally used for this purpose, but CGMs are both expensive and invasive (they require inserting a flexible needle under the skin). To address this issue, we examine whether non-invasive devices, such as electrocardiograms (ECG), can be used to predict glucose excursions. In particular, we consider two types of cardiac information: (1) heartbeat morphology, which generally requires ECG recordings, and (2) heartbeat timing, which can be obtained from inexpensive wrist-worn devices, such as fitness trackers. We use convolutional networks to analyze beat morphology, and recurrent networks and feature engineering to analyze the inter-beat interval (IBI) time series. Then, we validate individual models and their combinations on an experimental dataset containing ECG and CGM recordings for then young adults with type 1 diabetes. We find that beat morphology outperforms beat timing in hypoglycemia prediction, but the reverse happens for hyperglycemia prediction. In both prediction problems, combining morphology and time-domain information outperforms using each source of information independently.

Keywords—hypoglycemia, ECG, CGM, deep learning

I. Introduction

Complications from diabetes are significant causes of morbidity and mortality worldwide [1]. Sustained hyperglycemia (HG; high glucose) can lead to serious long-term complications, including heart disease, kidney disease, stroke, blindness, and amputations [2, 3]. In contrast, hypoglycemia (hg; low glucose) is more problematic in the short-term, as it can lead to confusion, irritability, palpitations, or even result in severe loss of attention, coma, or death [3]. Therefore, monitoring glucose is a critical component in diabetes prevention and management.

The most common method to monitor glucose requires patients to prick themselves with a lancet, draw a drop of blood, place it onto a disposable test strip, and insert into a glucometer [4]. This technique is inexpensive and accurate, but drawing blood is painful and only provides a one-time measurement that can potentially miss critical events. Continuous glucose monitors (CGMs) overcome this limitation, as they can measure glucose continuously and in real time. However, CGMs are invasive devices, as they require inserting a microneedle through the patient's skin, and keeping the sensor in place for 7-14 days. Further, CGMs are expensive, and generally only prescribed to patients with poorly-controlled diabetes. This is particularly problematic since 75% of patients with diabetes live in low- or middle-income countries [5].

As an alternative, several physiological signals that can be measured noninvasively, such as electrocardiography (ECG) and skin conductivity, contain information that correlates with glucose excursions [6]. Among them, the ECG signal, has achieved state-of-the-art performance when combined with deep learning (DL) [7]. However, ECGs are expensive and not designed for long-term use. Thus, recent interests have shifted towards photo-plethysmography (PPG) at the wrist, as it is widely available in consumer products (e.g., smartwatches, fitness trackers).

This study examines cost-accuracy tradeoffs between ECG (accurate but pricey) and PPG (affordable but less accurate) when used to detect glycemic excursions. Specifically, we compare the two main sources of information in these signals: (1) beat morphology, which generally requires an ECG, and (2) inter-beat intervals (IBI), which can be obtained from PPG. Experimental data for this work is part of a larger study where patients with type-1 diabetes wore a number of physiological sensors for up to two weeks, along with their prescribed CGM. In particular, we use ECG recordings to extract the two types of information; this avoids confounding information in PPG (the IBI sequence) with its signal quality (motion artifacts). Then, we use convolutional neural networks (CNN) and long-short-term memories (LSTMs) to predict glycemic excursions (HG, hg) from beat morphology and beat timing, respectively, and in various combinations.

II. Related work

When considering the problem of predicting glucose excursions, the vast majority of studies have focused on *forecasting*: use glucose readings from the past few hours to predict glucose levels –or risk of hg—at a future time [8].

Early work in this area used traditional techniques (e.g., autoregressive models), but current approaches are based on recurrent DL models, such as LSTMs [9, 10]. Note that glucose forecasting requires the patient to wear a CGM, so the forecasting model is an early warning system.

Alternatively, noninvasive wearable sensors *may* be used to estimate glucose [6]. Early work focused on skin temperature/conductivity, which chance during hg [11]. Several devices were marketed in the 1980s [12, 13], but they had multiple issues, such as false alarms due to perspiration unrelated to hg (a reported 3:1 false alarm to true alarm ratio [14]), and missed alarms in patients who do not experience these symptoms due hg unawareness [15]. More recent work has focused on ECG, as several changes in cardiac output have been associated with hg, most notably a lengthened QT interval [16] and a reduction in HRV [17]. HRV measurements are particularly appealing since they only require detection of the R peaks, which are prominent in ECG and also in PPG, whereas extracting the QT interval and other morphological features requires somewhat clean ECG recordings. In a recent study, Porumb et al. [7] used DL techniques to learn changes in ECG heartbeat morphology that occur during hg without requiring extraction of ECG fiduciary points. On an experimental dataset with non-diabetes adults, the authors report 76% accuracy in predicting nocturnal hg when using individual beats and 81% when using 200-beat segments.

III. Methods

We adopt similar signal-processing steps and DL architectures as those in [7] to compare beat morphology and IBI series by their ability to predict glycemic excursions.

A. Preprocessing

In a first step, we identify ECG segments of good quality. For this purpose, we rely on two signal-quality measures that the BioHarness reports every minute: heart-rate confidence (HRC) and ECG sensor noise (ECG-N), and only consider segments with HRC=100 and ECG-N<0.001. Next, we extract R peaks from valid ECG segments using Neurokit2 [18]. To extract ECG beats, we then place an analysis window around each R peak. Porumb et al. [7] used a fixed-length (160 sa.) window, but this can be problematic because, as heart rate increases (RR interval decreases), a fixed-length window can extend into the neighboring heart beats. Instead, we used a variable window whose length is a percentage of the RR interval (33% back, 66% forward), and zero-pad to ensure beats are of the same length. Finally, we label each ECG according to the next (closest) CGM reading: hg: <70 mg/dl, HG >180 mg/dl, normal otherwise.

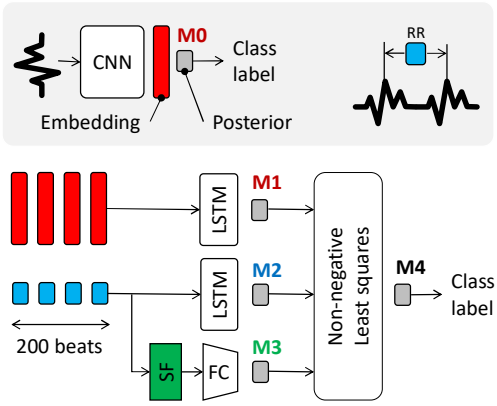

FIG. 1. (LEFT) EXTRACTING MORPHOLOGY EMBEDDINGS AND RR TIME SERIES. (RIGHT) COMBINING MULTIPLE SOURCES OF INFORMATION INTO A SINGLE PREDICTION.

B. Morphology information

To extract information from beat morphology, we use a CNN that consumes individual ECG beats (a 200-dim vector). To optimize hyper-parameters, we evaluate CNNs with (5, 7, 10) layers and (50, 100, 200, 300) kernels. As network size increases, so does training time and memory size, and also accuracy. However, past an "optimal" size, larger models only show modest improvements in accuracy but at a significant cost in training time and memory requirements. Based on these results (not shown), the final model has 10 layers with 50 kernels, each kernel of size 5, with batch normalization. We flatten the output of the final CNN layer and feed it to a fully connected (FC) network with three layers (size: 256, 30, 1). It is this final FC layer that performs binary classification[1]; the previous dense layer (size: 30) serves as an embedding for each ECG beat in subsequent models

C. Timing information

We use two distinct approaches to extract information from IBIs. Following [7], our first approach consists of feeding 200-beat IBI sequences to an LSTM and training it to predict glucose excursions. The LSTM contains a single layer with 400 units, and three fully connected (FC) layers (size: 256,30,1). Our second approach is more conventional, computes various statistical measures of heart rate and HRV [18] from the 200-beat IBI sequence, and feeds them to a FC model (size: 256,128,30,1). We extract 16 statistical features (SFs) using the Neurokit2 toolbox [18]; see footnote[2].

D. Model comparisons

Illustrated in Fig. 1, we evaluate five different models that predict glycemic excursions from beat morphology (CNN), IBI series (LSTM and SFs), and combinations of

[1] We use 'glorot_uniform' as the kernel initializer, 'zeros' as the bias initializer, and the Adam optimizer with a learning rate of '0.0001' for model development.

[2] Mean of RR intervals; median of absolute values of successive diffs. b/w RR intervals; standard deviation (STD) of RR intervals; median absolute deviation of RR intervals; STD of average RR intervals; median absolute deviation of RR intervals divided by median of absolute diffs. of their successive differences; mean of STDs of RR intervals; interquartile range of RR intervals; square root of the mean of the sum of successive diffs. b/w adjacent RR intervals; proportion of RR intervals <50ms; STD of successive diffs. b/w RR intervals; (12) proportion of RR intervals <20ms; STD of RR intervals divided by mean of RR intervals; HRV triangular index, measuring the total number of RR intervals divided by the height of the RR intervals histogram; root mean square of the sum of successive diffs. divided by the mean of the RR intervals; geometrical parameter of HRV.

them in an ensemble learning fashion.

- **Model M0 (CNN)**: The model makes an individual prediction for each ECG beat, as described in Section III.B. Thus, this model relies exclusively on beat morphology.
- **Model M1 (CNN→LSTM)**: This model feeds a sequence of 200 consecutive CNN embeddings to an LSTM. Thus, this model relies on temporal changes in beat morphology.
- **Model M2 (RR→LSTM)**: This model feeds a sequence of 200 consecutive RR intervals to an LSTM, as described in Section III.C. Thus, this model relies exclusively on beat timing.
- **Model M3 (SF):** This model takes sixteen SFs [18], obtained over 200-beat IBI sequence, and feeds them to a FC network. Thus, this model also relies exclusively on beat timing.
- **Model M4 (Combination):** This model combines predictions from M1-M3 using non-negative least squares.

For the 200-beat sequences that are consumed by the LSTMs, we determine missing beats from the IBI series and then introduce synthetic beats via linear interpolation. We use the same procedure to generate synthetic CNN embeddings for missing beats.

IV. Results

We evaluated the five models on two binary classification problems: hg vs. normal, and HG vs. normal. For each patient, we performed stratified three-fold cross validation, with 2/3rds of the data used to train the model and the remaining 1/3rd for testing. Splitting was done at the level of CGM readings, rather than ECG beats. Thus, all the heart beats associated with a given CGM reading are used either for training or for testing.

In a first analysis, we validated our choice of using a variable length-window to extract individual heart beats vs. using the fixed-length window of Porumb et al. [7]. For each participant/fold combination, we trained two separate CNNs, one that consumed fixed-length ECG beats and one that consumed zero-padded variable-length ECG beats. As shown in Fig. 2, at high heart rates (low RR intervals), the fixed-length window can cause problems as it extends into the previous heartbeat[3], whereas a variable-length window does not. Across patients, the average AUC for hg using fixed-length windows is 52.2%, and increases to 56.1% when using variable-length windows, a difference that is statistically significant ($p = 0.008$; one-tailed). For this reason, we use a variable-length window for all future analyses.

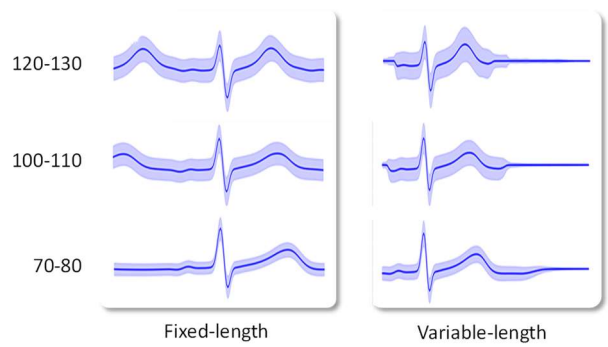

FIG. 2. FIXED- VS. VARIABLE-LENGTH WINDOWS. AT HIGH HEART RATES (TOP-LEFT), THE FIXED-LENGTH WINDOW CAPTURES THE T WAVE OF THE PREVIOUS HEART BEAT

Performance of the five models on hg prediction is summarized in Fig. 3; each bar represents the average AUC across patients. One-way ANOVA shows that there is statistically significant difference across models ($p \ll 0.001$). Post hoc-comparisons shows no statistical differences between M3 and M2 ($p = 0.08$) or between M2 and M0 ($p = 0.06$), indicating that beat timing and beat morphology provide similar amount of information about hg. However, we find statistically significant differences between M0 and M1 ($p = 0.001$), and between M1 and M4 ($p = 0.01$), indicating that combining both sources of information achieves significantly higher performance than using either source of information alone.

Model performance in HG is also shown in Fig. 3. As before, one-way ANOVA indicates there are statistically significant differences across models ($p \ll 0.001$). In this case, however, beat morphology (M0) performs worse than beat timing, either using the raw RR series (M3; $p = 0.008$) or the SFs (M2: $p = 0.002$). As with hg, the best performing model is the one that combines predictions from beat morphology and beat timing.

V. Discussion

This study examined whether non-invasive sensors, ECG in particular, may be used to predict glycemic excursions in a type-1 diabetes population. In particular, we compared two types of information: beat morphology and beat timing. This comparison is significant since beat morphology generally requires ECG, which is expensive and not designed for long-term use, whereas beat timing may be extracted from PPG, which are more affordable and available as consumer products.

Results from our study show two different trends. For hypoglycemia, beat morphology outperforms beat timing,

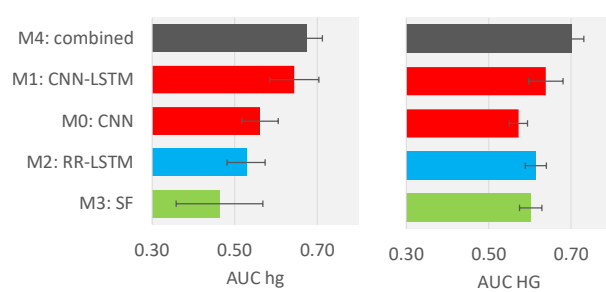

FIG. 3. MODEL PERFORMANCE ON HYPOGLYCEMIA (LEFT) AND HYPERGLYCEMIA (RIGHT) PREDICTION

[3] As heart rate increases, notice how the T wave is closer to the R peak; this is the reason why hypoglycemia prediction studies often use the corrected QT period relative to the RR interval [19].

suggesting that ECG devices are needed. For hyperglycemia, however, beat timing outperforms beat morphology, suggesting that this problem could be solved using PPG devices. In both problems, however, combining both sources of information provides a noticeable boost in performance.

When comparing the two binary tasks, we find that the AUC is systematically higher for hyperglycemia than for hypoglycemia. This result may be due to the fact that, in our dataset, hyperglycemia is three times as prevalent as hypoglycemia. Since both are the minority class, the one with more examples will have better class balance, which generally leads to better model performance.

VI. Future work

At present, we extract a single 200-beat sequence for each CGM reading. However, it would be possible to extract multiple partially-overlapping sequences per CGM reading, and treat each of them as a separate example. This strategy can be used to generate more examples for the minority class and reduce class imbalance. Likewise, instead of making an independent prediction for each heart beat (i.e., model M0), combining predictions from all the beats associated with a single CGM reading (e.g., using a majority vote among all beat predictions) can improve accuracy as long as the M0 predictors operate above chance level.

We used non-negative linear regression to combine predictions from multiple classifiers, as this made the model easier to interpret. In fact, examining model coefficients indicates that, on average, the CNN-LSTM classifier receives 65% of the weight, compared to 25% for RR-LSTM and 10% for the AF classifier. However, if interpretability is not as critical as accuracy, better ensemble predictions could be generated by using more complex functions (e.g., a small FC network). More complex functions could also be obtained using the “early fusion” approach of combining models at the embedding level (second-to-last layer), rather than our “late-fusion” approach that combines models at the probability level (last layer).

This study focused on ECG signals. However, as part of our larger study at Baylor College of Medicine, patients wear three physiological devices: (1) the BioHarness chest strap, which measures ECG, respiration, and 3D accelerometry, (2) an Empatica E4 watch, which measures PPG, EDA, 3D accelerometry, and skin temperature, and (3) an Oura ring, which measures PPG and skin temperature. Thus, our future work includes examining these additional sensing modalities in isolation and in combination, and understand how usability considerations may impact model accuracy. To put an example, wearing a ring is less intrusive than wearing a watch, but a ring device does not generally provide as much information as a watch device. Combining information from multiple devices may also help eliminate potential confounders. For example, elevated heart rate could be due to glucose excursions or to physical activity; to rule out the latter, the classifier could use accelerometer data to verify physical activity was not the cause.

VII. Acknowledgments

This work was supported by NSF award 2037383.

VIII. References

[1] E. Standl, K. Khunti, T. B. Hansen, and O. Schnell, "The global epidemics of diabetes in the 21st century: Current situation and perspectives," *European J. Preventive Cardiology,* vol. 26, no. 2_suppl, pp. 7-14, 2019.

[2] M. Qaraqe, M. Erraguntla, and D. Dave, "AI and Machine Learning in Diabetes Management: Opportunity, Status, and Challenges," in *Multiple Perspectives on Artificial Intelligence in Healthcare*: Springer, 2021, pp. 129-141.

[3] DCCT, "Epidemiology of Severe hypoglycemia in the diabetes control and complications trial. The DCCT Research Group," *Am. J. Medicine,* 2004.

[4] S. H. Atkin, A. Dasmahapatra, M. A. Jaker, M. I. Chorost, and S. Reddy, "Fingerstick glucose determination in shock," *Annals of Internal Medicine,* vol. 114, no. 12, pp. 1020-1024, 1991.

[5] IDF, "International Diabetes Federation: IDF diabetes atlas," Brussels, Belgium: International Diabetest Federation, 2021, pp. 36-39.

[6] D. Howsmon and B. W. Bequette, "Hypo- and Hyperglycemic Alarms: Devices and Algorithms," (in eng), *J Diabetes Sci Technol,* vol. 9, no. 5, pp. 1126-37, Apr 30 2015.

[7] M. Porumb, S. Stranges, A. Pescapè, and L. Pecchia, "Precision medicine and artificial intelligence: a pilot study on deep learning for hypoglycemic events detection based on ECG," *Scientific Reports,* vol. 10, no. 1, pp. 1-16, 2020.

[8] S. Oviedo, J. Vehí, R. Calm, and J. Armengol, "A review of personalized blood glucose prediction strategies for T1DM patients," (in eng), *Int J Numer Method Biomed Eng,* vol. 33, no. 6, Jun 2017.

[9] J. Martinsson, A. Schliep, B. Eliasson, C. Meijner, S. Persson, and O. Mogren, "Automatic blood glucose prediction with confidence using recurrent neural networks," in *Proc. KHD@ IJCAI*, 2018.

[10] S. Mirshekarian, R. Bunescu, C. Marling, and F. Schwartz, "Using LSTMs to learn physiological models of blood glucose behavior," in *Proc. IEEE EMBC Annual Intl. Conf.*, 2017, pp. 2887-2891.

[11] R. E. Bolinger, R. Stephens, B. Lukert, and D. Diederich, "Galvanic Skin Reflex and Plasma Free Fatty Acids During Insulin Reactions," *Diabetes,* vol. 13, no. 6, pp. 600-605, 1964.

[12] J. C. Pickup, "Preliminary evaluation of a skin conductance meter for detecting hypoglycemia in diabetic patients," (in eng), *Diabetes Care,* vol. 5, no. 3, pp. 326-9, May-Jun 1982.

[13] L. Levandoski, N. White, D. Popp, and J. Santiago, "Teledyne Sleep Sentry®—a possible aid for the detection of symptomatic nocturnal hypoglycemia in insulin-dependent diabetics," *Artificial Systems for Insulin Delivery,* pp. 353-6, 1983.

[14] W. L. Clarke, W. R. Carter, M. Moll, D. J. Cox, L. A. Gonder-Frederick, and P. E. Cryer, "Metabolic and Cutaneous Events Associated With Hypoglycemia Detected by Sleep Sentry," *Diabetes Care,* vol. 11, no. 8, pp. 630-635, 1988.

[15] P. E. Cryer, "Diverse causes of hypoglycemia-associated autonomic failure in diabetes," (in eng), *N Engl J Med,* vol. 350, no. 22, pp. 2272-9, May 27 2004.

[16] B. Eckert and C. D. Agardh, "Hypoglycaemia leads to an increased QT interval in normal men," (in eng), *Clin Physiol,* vol. 18, no. 6, pp. 570-5, Nov 1998.

[17] J. P. Singh, M. G. Larson, C. J. O'Donnell, P. F. Wilson, H. Tsuji, D. M. Lloyd-Jones, and D. Levy, "Association of hyperglycemia with reduced heart rate variability (The Framingham Heart Study)," *The American Journal of Cardiology,* vol. 86, no. 3, pp. 309-312, 2000/08/01/ 2000.

[18] D. Makowski, T. Pham, Z. J. Lau, J. C. Brammer, F. Lespinasse, H. Pham, C. Schölzel, and S. Chen, "NeuroKit2: A Python toolbox for neurophysiological signal processing," *Behavior research methods,* vol. 53, no. 4, pp. 1689-1696, 2021.

[19] M. Desai, L. Li, Z. Desta, M. Malik, and D. Flockhart, "Variability of heart rate correction methods for the QT interval," (in eng), *Br J Clin Pharmacol,* vol. 55, no. 6, pp. 511-7, Jun 2003.