

Hypoglycemia and hyperglycemia detection using ECG: A multi-threshold based personalized fusion model

Darpit Dave^a, Kathan Vyas^b, Gerard L. Cote^{c,d,e}, Madhav Erraguntla^{a,e,*}

^a Wm Michael Barnes '64 Department of Industrial and Systems Engineering, Texas A&M University, Emerging Technology Building, College Station, TX 77843, USA

^b Department of Computer Science and Engineering, Texas A&M University, L.F. Peterson Building, College Station, TX 77843, USA

^c Department of Biomedical Engineering, Texas A&M University, Emerging Technology Building, College Station, TX 77843, USA

^d Department of Electrical and Computer Engineering, Wisenbaker Engineering Building, Texas A&M University, College Station, TX 77843, USA

^e Center for Remote Health Technologies and Systems, Texas A&M Engineering Experiment Station, Texas A&M University, College Station, TX 77843, USA

ARTICLE INFO

Keywords:

noninvasive glucose monitoring
ECG
hypoglycemia
hyperglycemia
precision medicine
ECG-beat morphology
heart rate variability (HRV)
noninvasive sensors
machine learning
fusion model

ABSTRACT

Monitoring glucose levels is critical for effective diabetes management. Continuous glucose monitoring devices estimate interstitial glucose levels and provide alerts for glycemic excursions. However, they are expensive and invasive. Therefore, low-cost, noninvasive alternatives are useful for patients with diabetes. In this article, we explore electrocardiogram signals as a potential alternative to detecting glycemic excursions by extracting intra-beat (beat-morphology) and inter-beat (heart rate variability) information. Unlike prior methods that focused only on the standard clinical excursion thresholds (70 mg/dL for hypoglycemia, 180 mg/dL for hyperglycemia), our proposed approach trains independent machine learning models at various excursion thresholds, aggregating their outputs for a final prediction. This allows learning morphological patterns in the neighborhood of the standard excursion thresholds. Our personalized fusion models achieve an AUC of 75 % for hypoglycemia and 78 % for hyperglycemia detection across patients, resulting in an average improvement of 4 % compared to the baseline models (trained using only standard clinical thresholds) for detecting glycemic excursions. We also find that combining morphology and HRV information outperforms using them individually (5 % for hypoglycemia and 6 % for hyperglycemia). The data used in this article was collected from 12 patients with type-1 diabetes, each monitored over a 14-day period at Texas Children's Hospital, Houston. The results indicate that a combination of morphological and HRV features is essential for noninvasive detection of glycemic excursions. Also, morphological changes can happen at varying glucose levels for different patients and capturing these changes provide valuable information that leads to improved prediction performance for detecting glycemic excursions.

1. Introduction

Diabetes mellitus commonly known as diabetes, is the body's inability to balance blood glucose (BG) levels. It results from defects in either insulin secretion, predominant in patients with type-1 diabetes, where the body fails to produce sufficient insulin for the cells; or insulin inaction, as seen in patients with type-2 diabetes, where the cells get resistant to insulin. Diabetes is a serious global health threat with an estimated 537 million people affected by 2021 and projections indicating a rise to 1.31 billion people by 2050 [1]. Many studies have shown high correlations between poor glycemic control and various health conditions [2–7]. Diabetes and its related complications, if not well managed, can lead to serious short-term and long-term

consequences. Sustained hyperglycemia (high glucose) is often a catalyst for heart diseases, kidney diseases, strokes, blindness, and amputations [8]. On the other hand, hypoglycemia (low glucose) leads to short-term complications such as loss of consciousness, palpitations, seizures and in some cases, coma and death [9]. Diabetes is incurable and can only be managed through proper insulin treatment based on frequent monitoring of glucose levels [2,10].

Traditionally, glucose was measured through the finger-stick method, where patients draw a small blood sample by pricking tip of the finger and then run it through a glucometer [11]. The finger-stick method is inexpensive and accurate, but it is painful and provides only a snapshot of the glucose levels. CGM devices overcame these drawbacks by providing frequent and automated glucose measurements.

* Corresponding author.

E-mail addresses: darpidave@tamu.edu (D. Dave), kathan@tamu.edu (K. Vyas), gcote@tamu.edu (G.L. Cote), merraguntla@tamu.edu (M. Erraguntla).

<https://doi.org/10.1016/j.bspc.2024.106569>

Received 2 October 2023; Received in revised form 27 May 2024; Accepted 13 June 2024

Available online 1 July 2024

1746-8094/© 2024 Elsevier Ltd. All rights are reserved, including those for text and data mining, AI training, and similar technologies.

CGMs also implement alert systems that can notify patients about low and high glucose levels, allowing patients to take intervention measures. Multiple studies have shown the effectiveness of CGMs in diabetes management [12,13]. However, CGMs, despite their success, are invasive, expensive, and subject to strict regulations which can be a major deterrent to their use in low-income populations, patients with type-2 diabetes, and the pre-diabetic population [1]. Therefore, noninvasive alternatives for estimating glucose levels can be very beneficial.

Recent advances in noninvasive technologies, have led researchers to explore several physiological signals collected noninvasively such as electrocardiography (ECG), photoplethysmography (PPG), skin conductivity, near-infrared (NIR) spectroscopy, electrodermal activity to detect glycemic excursions [14–17]. These physiological features have shown correlation with glycemic changes [14,18]. Among them, ECG signals have emerged to be the most popular because changes in BG levels stimulate the autonomic nervous system and lead to variations in the heart functions [19]. ECG signals can capture these variations and estimate the corresponding glycemic changes accurately.

1.1. ECG: A prominent signal for glucose monitoring

Features extracted from the ECG signal for prediction, can broadly be categorized into (i) Morphology features (e.g.: QT-interval (QT), R amplitude, etc.), and (ii) HRV features (e.g.: heart rate, standard deviation of NN-intervals (SDNN), etc.). Early works in this area focused on the use of features such as heart rate (HR) and heart rate variability (HRV) measures, morphological patterns like QT-interval (QT), corrected QT-interval (QTc) and related changes [20–23]. These features were used with an extreme machine learning (ELM) model to detect nocturnal hypoglycemia episodes in type-1 diabetes population and reported 78 % sensitivity and 60 % specificity [20]. A popular study in the literature examined corrected QT-interval (QTc) prolongation during hypoglycemia among ten adults with type-1 diabetes [24]. The study found QTc derived through both, Bazett's and Fridericia's formulas, to be elongated during hypoglycemia compared to the baseline. In another study comprising 22 subjects (9 healthy, 6 T1DM but otherwise healthy and 7 T1DM with disease complications) cardiac repolarization features viz. QTc and RT-amplitude ratio were used for hypoglycemia detection [22]. More recently, a convolutional neural networks (CNNs) based model was proposed as an alternative to manual feature engineering for extracting morphology features from the raw ECG signal [25]. This was used to identify individual beats as hypoglycemia or normal. The authors also propose a Long-short term memory (LSTM) architecture that combines a sequence of 200-beats and classifies them as hypoglycemia or normal. In addition, the authors propose a majority-voting scheme over a 10-minute period for better annotation of the results. However, the analysis in this paper is limited to the detection of nocturnal hypoglycemia episodes only. A similar approach of using CNN layers to extract features from the raw ECG-signal was proposed in another study [26]. The authors adopt a majority-voting scheme over 10-beats in the terminal of the CNN layer for better classification and interpretability. In this work, the authors consider a multi-class classification problem by extending prediction to hypoglycemia, hyperglycemia, and normal ranges. In our previous study [27], we proposed the use of ECG signal and accelerometry data to detect hypoglycemia and hyperglycemia independently. A total of nine time-domain HRV features computed over 5 consecutive windows of 1-minute each were used to predict hypoglycemia and hyperglycemia. A recent work also found HRV features and HR to be prominent for detecting hypoglycemia events over 1-minute non-overlapping windows [28]. The authors in the study use a combination of HRV features (computed with Garmin Vivoactive 4 s-based PPG data), along with motion and EDA-based features (from Empatica E4). The study included 22 individuals (16 males, 6 females) and reported a prediction performance of $AUC = 0.76$ for detecting hypoglycemia.

Multiple glycemic prediction studies based on ECG and PPG signals

have relied on experiments conducted in controlled settings to acquire data and perform the analysis. Also, except for two recent studies [26,27], the literature has primarily focused on the detection of hypoglycemia only. The goal of diabetes management is to attain euglycemia and improve the time-in-range (TIR) of the target glucose range. This requires accurately detecting hypoglycemia and hyperglycemia [2,26,29]. The majority-voting scheme is a popular approach in the machine learning literature and has been adopted by previous works to improve the interpretability of results when aggregating beat-level predictions over an interval. However, this approach may lead to poor performance when individual beat-level predictions are inconsistent within an interval [30]. A better approach is required to aggregate these individual beat-level predictions and improve performance.

1.2. Glycemic excursion thresholds

A general consensus-based recommendation for hypoglycemia is any CGM value < 70 mg/dL and for hyperglycemia is any CGM value > 180 mg/dL [2]. However, a recent study by the conglomerate HYPOResolve [31,32], emphasized the need to define an optimal threshold of sensor glucose readings that is consistent with the actual hypoglycemia events occurring in people. The study indicates the need to understand conventionally defined CGM-based hypoglycemia episodes to actual patient reported hypoglycemia (PRH).

In a separate independent study [33], the authors noticed that expected morphological changes are not consistently detected or visible for all hypoglycemia events (considering 70 mg/dL as the hypoglycemia threshold). The study examines three HRV features: the standard deviation of the NN intervals (SDNN) and the square root of the mean standard differences of successive NN intervals (RMSSD) as time-domain features, and the ratio between low and high frequency (LF:HF) as a frequency-domain feature. All three features exhibited statistically significant changes during hypoglycemia readings, indicating the potential for detecting hypoglycemia events through HRV features. However, the observed changes in these features were inconsistent across patients who experienced hypoglycemia events. The authors identified factors such as duration of diabetes, physical activity, and rate of declining glucose values as factors associated with prominent changes in the HRV features during hypoglycemia events. The study was based on a patient cohort of 23 patients with type-1 diabetes where hypoglycemia was defined as glucose ≤ 70 mg/dL.

A recent study [34], explored changes in cardiac repolarization features like corrected QT-interval, T-wave and HRV features in type-1 diabetes patients during hypoglycemia. A key finding from the study was that changes in ECG-based features extended beyond the hypoglycemia event (defined at 70 mg/dL), indicating morphological changes in the neighborhood of the standard clinical threshold of hypoglycemia (70 mg/dL).

Another recent study using noninvasive smartwatches to monitor interstitial glucose highlights the need to personalize glycemic excursion thresholds especially when studying associated morphological changes [35]. The authors redefined and created personalized 'high' and 'low' glucose thresholds for each participant in the study using the previous 24-hours of their individual CGM readings on a rolling basis. Readings one standard deviation below (or above for hyperglycemia) the mean for the last 24-hours of CGM readings are defined as 'PersLow' or hypoglycemia.

Most recently [36], another study using smartwatches for noninvasive hypoglycemia detection during cognitive and psychomotor stress found that although HRV measures constitute relevant features for hypoglycemia detection. However, the accuracy of ML decision-making varies across different levels of hypoglycemia. The authors emphasize the need to study physiological changes corresponding to glycemic changes in the presence confounding factors like stress and, glycemic sequence (euglycemia followed by hypoglycemia).

The above studies show that ECG morphological changes associated

with low glucose do not necessarily occur at a specific glucose level. They can happen at different thresholds for different subjects. Even for the same subject, these can occur at different thresholds at different times. This indicates the need for a prediction model that incorporates changes occurring within a range of glucose values in addition to specific thresholds of 70 mg/dL and 180 mg/dL for detecting hypoglycemia and hyperglycemia respectively.

There are multiple factors that directly impact blood glucose levels, like food intake [37], physical activities [38], stress [39]. Additionally, indirect factors including body temperature [40], autonomic functions [41] can affect glucose and related fluctuations within the person [42]. However, these factors are extremely personalized, meaning each of these factors impact glucose dynamics in varying degrees to different individuals [43]. Investigating the role of these multiple factors simultaneously or individually for excursion detection can be valuable, but the scope of this article is limited to identifying ECG related information most useful for accurate glycemic excursion detection.

More specifically, the major contributions of our work are:

1. A combination of ECG morphology features and HRV features is useful for hypoglycemia and hyperglycemia detection as opposed to their independent use.
2. An ML approach that leverages the patterns in a sequence of beat-level glycemic predictions for improved interval predictions.
3. A fusion model approach to leverage ECG morphological patterns at varying thresholds for improved level glycemic predictions at standard clinical thresholds.

2. Methods and materials

This section describes the dataset used in this study, the data processing steps and the experimental setup. We also describe the prediction model used and the different configurations explored in this study. To provide a summary of all the relevant information in this article and improve the transparency of reporting machine learning modeling results in healthcare, we report the necessary details as per the MI-CLAIM checklist in Appendix T1 [44].

2.1. Clinical datasets

To evaluate the noninvasive detection of glycemic excursions, 12 participants were recruited at Texas Children's Hospital, Houston-TX, USA. These 12 participants were adults with type-1 diabetes. Each participant was enrolled in the study for 14 days. The participants were asked to wear 3 noninvasive devices (Zephyr Bioharness, Empatica E4, and Oura ring) during the study period. In addition to a CGM device and an insulin pump, these three devices are used by the patients as part of their diabetes management. Data from the Zephyr Bioharness and CGM (considered as the ground truth) were used in this paper and detailed in Table 1:

Throughout the study duration, the subjects were asked to wear the devices continuously, including both day and night periods, except for the time allocated for charging. However, variability in the available ECG data across the 24-hour window for each participant indicates that patients might have removed their device(s) while taking a shower or

Table 1
Devices used in the study and data acquired.

Device	Worn as	Data Collected and Frequency
Zephyr Bioharness	Chest strap	ECG at 250 Hz, Accelerometer at 100 Hz (3-axis), Heart Rate (HR) at 1 Hz, HR Confidence (HRC) at 1 Hz
Dexcom CGM	Over the arm	Subcutaneous glucose readings at 5 min intervals

some other activities. The entire study cohort comprises patients with type-1 diabetes and falls in the age group of 20–40 years. More details on these patients' demographic details and glycemic profiles during the study period can be found in the Appendix. (Appendix T2, Appendix T3, and Appendix T4).

Of the 12 enrolled patients, 3 patients (Patient ID: 2, 4 and 12) were excluded from the analysis for hypoglycemia detection and 1 patient (Patient ID: 5) was excluded for hyperglycemia detection. This exclusion was based on an extremely low number of available glycemic excursion events (hypoglycemia/ hyperglycemia) to develop and validate the proposed approach. Detailed information regarding the glycemic profiles of these patients is provided in the Appendix (Appendix T3 and T4).

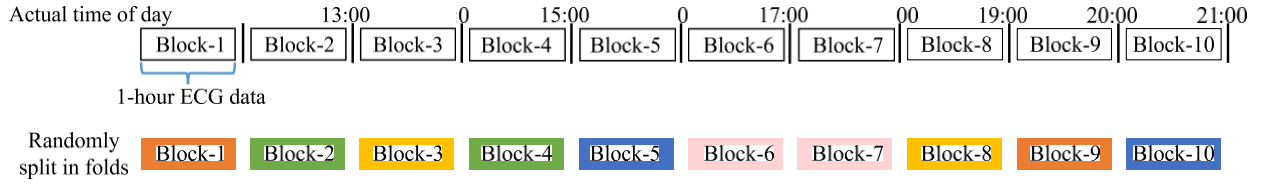
2.2. Preprocessing

ECG signals are vulnerable to distortions caused by motion artifacts [45–48], necessitating the preprocessing of the ECG signal to extract useful information. We conduct a thorough ECG processing to extract the maximum amount of high-quality data for our subsequent analysis. We detect R-peaks from the raw ECG signal in the first step using NeuroKit2 [49]. In the next step, we identify clean ECG beats based on the signal-quality measure 'HRCconfidence' (HRC) provided by Zephyr Bioharness at a frequency of 1 Hz. HRC is a score, expressed as a percentage of the confidence that the module is picking up heart rate during activity based on wear detection and quality of ECG. This score ranges from 0 to 100, with 100 indicating the best quality and 0 representing the worst. We exclusively consider beats (detected R-peaks) with an HRC score greater than 90 for the selection of clean ECG beats for our analysis. This decision is based on the availability of beats for different HRC cut-offs (Appendix F2) and a thorough visual inspection of the detected beats at different scores. For the selected beats with the R-peak as the anchor point, we identify each beat's remaining peaks (P, Q, S, T). Beats for which NeuroKit2 failed to detect either of the P, Q, S, and T peaks were discarded from the analysis. Based on the detection of beats and the fiducial points, we compute the beat-level morphology features and heart-rate variability (HRV) features next.

2.3. Validation mechanism

In this study, we implement a five-fold cross-validation scheme with temporal data splitting. Here, we first order the data based on time-stamps and segment it into one-hour blocks. Next, we randomly distribute these one-hour blocks into five equal partitions, ensuring each partition contains a mix of positive class (hypoglycemia or hyperglycemia) and negative class (non-hypoglycemia or non-hyperglycemia). It is important that while each partition is guaranteed to have positive class data samples, the partitions may not be stratified i.e., each partition can have varying proportions of positive and negative classes. This deliberate design mimics a real-world scenario where excursion events are not identically distributed over time. The main goal behind this validation mechanism is to minimize the temporal correlation between the data in the training and test sets. This ensures that the trained model does not get any undue advantage for making predictions on the test data, preventing overly optimistic prediction results [12,50]. A visual representation of our validation mechanism can be found in Fig. 1.

The decision to consider hourly blocks of data is justified by (a) prior literature indicating that autocorrelation between continuous CGM values diminishes beyond a period of 30–60 min [51,52] and (b) considerable evidence supporting a similar validation approach in the domain of glycemic predictions [25]. The temporal-splitting validation approach is significantly more rigorous than the simple random splitting (at the beat-level or CGM-level). This is evidenced by an empirical comparison of assessing model performance across the two validation settings (refer to Table T5 in Appendix).



Iteration/ Folds	Training Data					Test Data
1						
2						
3						
4						
5						
Final predicted data						

Fig. 1. Temporal validation: Validation scheme for splitting hourly blocks of data.

2.4. Feature extraction

2.4.1. Morphology features

We compute morphology features based on the fiducial points (P, Q, R, S, and T peaks) using the processed data described above. A total of 35

morphology features are extracted, comprising 9 Euclidean-based distances between individual peaks, 10 interval-based distances, 5 amplitudes of individual peaks and 9 slopes computed between the individual peaks. Additionally, we compute ‘RR’ as the interval distance between the current beat and the next detected beat, and ‘HR’ is taken as

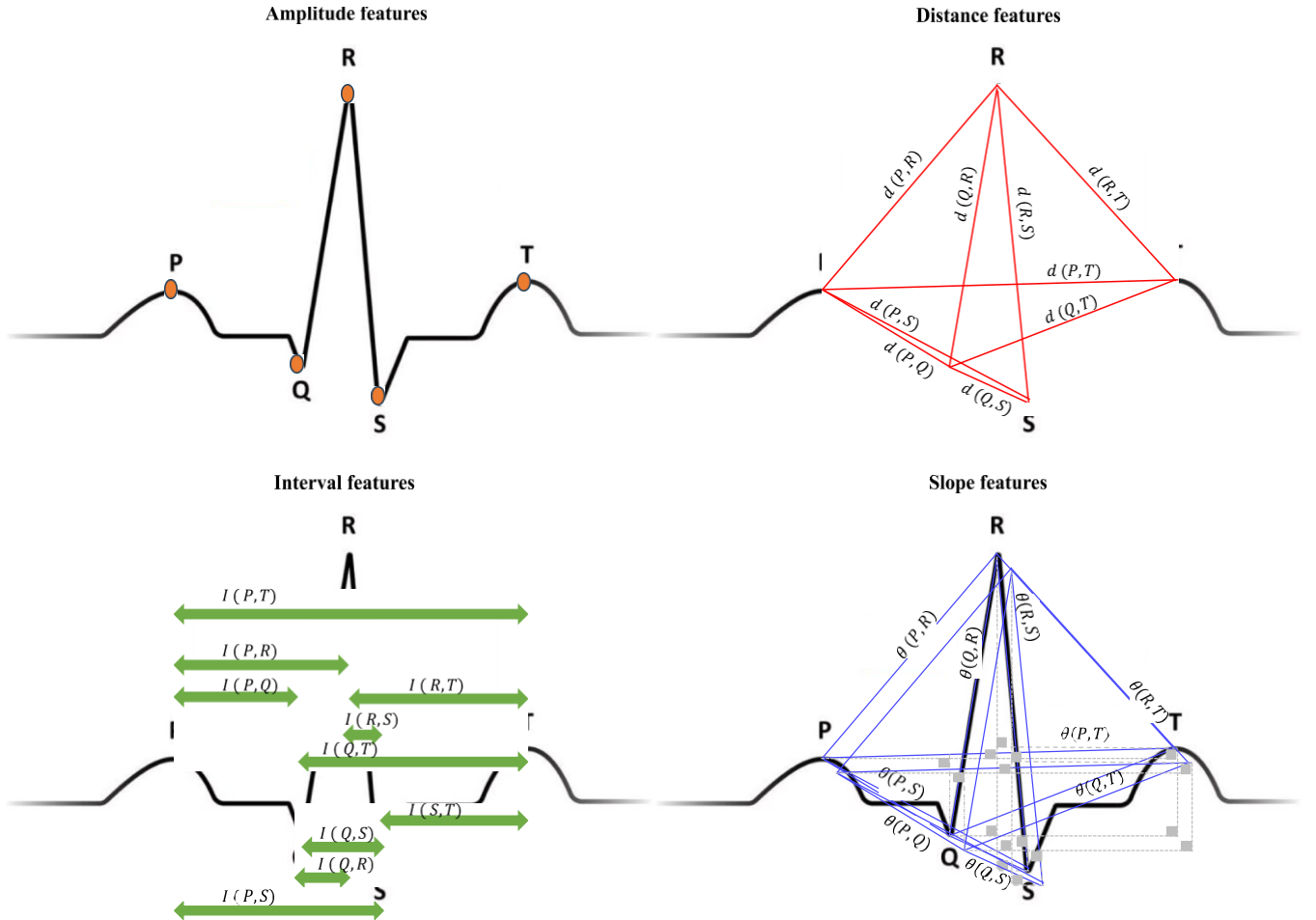


Fig. 2. Extraction of Amplitude, Distance-based, Interval-based, Slope-based features for beat-level prediction.

provided by Zephyr, including them as features. Fig. 2 provides a visual representation of the features.

2.4.2. HRV features

A total of 18 HRV time-domain features are computed using the NeuroKit2 package. A detailed description of the HRV features extracted can be found in Appendix T5.

2.4.3. Interval-level features

The beat-level ML model makes glycemic predictions at the individual beat-level. We aggregate beat-level predictions within a 1-minute interval for computing features at the interval-level. The extracted features are detailed in Table 2. A probability threshold of 0.5 was used to convert probabilities to classes. We also include the hour of the day encoded as a cyclical feature [53,54] (see Table 2).

2.5. Key definitions and evaluation metrics

1. **Classification task:** This article explores excursion detection as two separate classification tasks: (a) hypoglycemia detection and (b) hyperglycemia detection using ECG data as the input. CGM readings are used only as a reference point to categorize the output label dichotomously. The nearest CGM reading in the forward direction is employed for the ECG data point to mark and assign the associated output label. Except for the beat-level model (which predicts for individual beats), all models in this article are evaluated for a 1-minute interval prediction. This means that, for every 1-minute interval, if the associated CGM reading is within the excursion region, it is classified as a positive class/event. This 1-minute interval window aligns with established practices in the literature for similar applications [28,36]. Importantly, these 1-minute intervals are non-overlapping i.e., each data point belongs to precisely one 1-minute interval. For the beat-level model, predictions are made for each individually detected beat. Standard clinical thresholds of $< 70\text{mg/dL}$ for hypoglycemia and $> 180\text{mg/dL}$ for hyperglycemia are used as excursion thresholds [2,55]. This means that all the beats or 1-minute intervals associated with a CGM value will be marked as positive/negative class based on that single CGM reading. In training multiple independent beat-level models within the fusion model, where excursion thresholds other than standard clinical thresholds are considered have been explicitly specified. However, the final prediction of the fusion model is still being evaluated at the widely accepted standard clinical thresholds.
2. **Evaluation metrics:** To provide a comprehensive and robust evaluation of the binary classification models developed in this study, we report model performance using the Area under the curve (AUC) metric. The area under the receiver operating characteristic (ROC) curve characterizes the trade-off between the true positive rate and the true negative rate at various decision threshold settings. We provide model performance details on other relevant metrics like sensitivity (recall/ true positive rate): the ability of the classifier to detect true incidents correctly. Higher sensitivity indicates a lower

type II error rate, specificity: the ability of the classifier to correctly detect negative incidents. Higher specificity has a lower type I error rate, precision (positive predictive value): ratio of true predictions over total predictions, and f1-score: harmonic mean of precision and recall (Appendix T7). The above metrics were carefully chosen based on the ability to provide the reader with a fair estimate of the model performance as well as standard metrics used for evaluation in the literature for similar applications [25,28,36,56].

3. **Modeling approach:** We have adopted a personalized modeling approach for developing all models in this article. This means the development of separate independent models for each patient. This approach is supported by the previous literature [25] and is also based on our findings indicating a significant inter-subject variability in ECG features among patients. We conducted a multi-way Kruskal-Wallis H-test for all the twelve patients' ECG features across hypoglycemia, hyperglycemia and normal glucose ranges. A significant interaction across the 12 subjects indicated a significant ($p\text{value} < 0.001$) difference in ECG features. Further investigation through pairwise differences between subjects, using a post-hoc comparison with a two-way Kruskal-Wallis H-test and Dunn's test, revealed inter-subject variability across each paired patient combination for almost all the features. A $p\text{value} < 0.05$ was used as evidence of statistical significance.

2.6. Beat-level and Interval-Level models

One ML model for beat-level predictions and three for interval-level predictions are developed. All the models are based on the Random Forest (RF) algorithm. RF was chosen based on high performance and lower variance based on our previous works [9,12,27]:

1. **M_{Beat} (Beat – level model) :** This model makes predictions for individual beats. The model output (predicted probabilities) is aggregated for computing features over 1-minute intervals and used in interval-level models. The input to this model is the morphology features described in Fig. 3.
2. **M_{MV} (Majority – voting model) :** This model makes interval-level predictions based on a simple majority among the associated beat-level predictions. For a fair comparison, performance metrics are derived in 2-steps: (i) the beat-level predicted probabilities are converted into class labels based on an optimal threshold chosen for each split. (ii) for predicting the interval-level class, the majority threshold is varied (against the default 50 % as a majority) to have the best possible performance for each split.
3. **M_{Morph} (Morphology – features model) :** This model uses features aggregated from posterior probabilities extracted from the output of M_{Beat} (Table 2). These features are input to a secondary model to get interval-level predictions.
4. **M_{HRV} (HRV – features model) :** This model only relies on HRV features to make interval-level predictions. The HRV features (as in Appendix T5) are computed for a 1-minute interval.
5. **$M_{Morph+HRV}$ (Morphology – aggregation + HRV features) :** In this model, we combine the input features of M_{Morph} and M_{HRV} to make interval-level predictions. This evaluates the combined effect of intra-beat (morphology features) information and inter-beat (HRV features) information.

2.7. Fusion Model

The fusion model (**MF**) extends the interval-level baseline model $M_{Morph+HRV}$ as depicted in Fig. 4. Multiple beat-level models (M_{Beat}) are trained at different glycemic excursion thresholds for hypoglycemia and hyperglycemia. For hypoglycemia, individual classifier models (M_{Beat}) are trained at thresholds of 55, 60, 65, 70, 75, 80, 85 and 90. Similarly for hyperglycemia, individual classifier models (M_{Beat}) are trained at

Table 2
Interval-level features extracted by aggregating beat-level predictions.

Features	Description
% of hypo beats	% of beats classified as hypoglycemia (threshold = 0.5)
Longest hypo-sequence	The longest sequence of predicted hypoglycemia beats (threshold = 0.5)
Mean	Mean predicted probability
Group 1	% of predicted probabilities in the interval (0,0.2]
Group 2	% of predicted probabilities in the interval (0.2,0.4]
Group 3	% of predicted probabilities in the interval (0.4,0.6]
Group 4	% of predicted probabilities in the interval (0.6,0.8]
Group 5	% of predicted probabilities in the interval (0.8,1]
Hour	Hour of the day

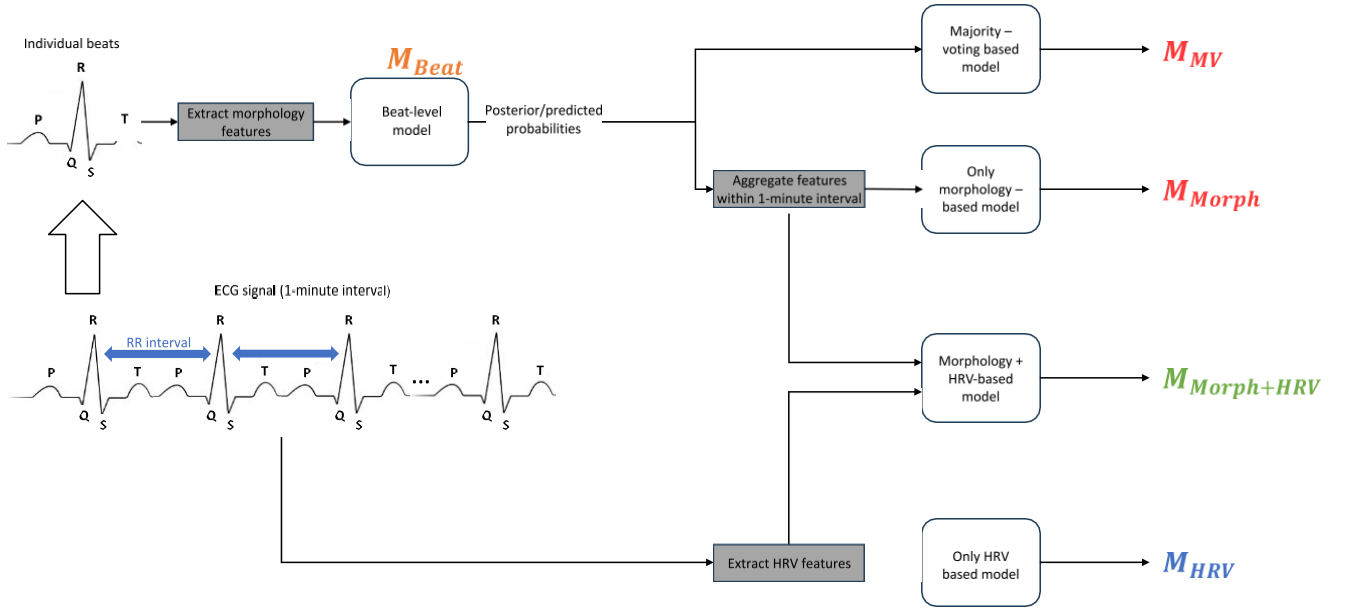


Fig. 3. Workflow of different interval-level modeling approaches in this study and their comparative analysis.

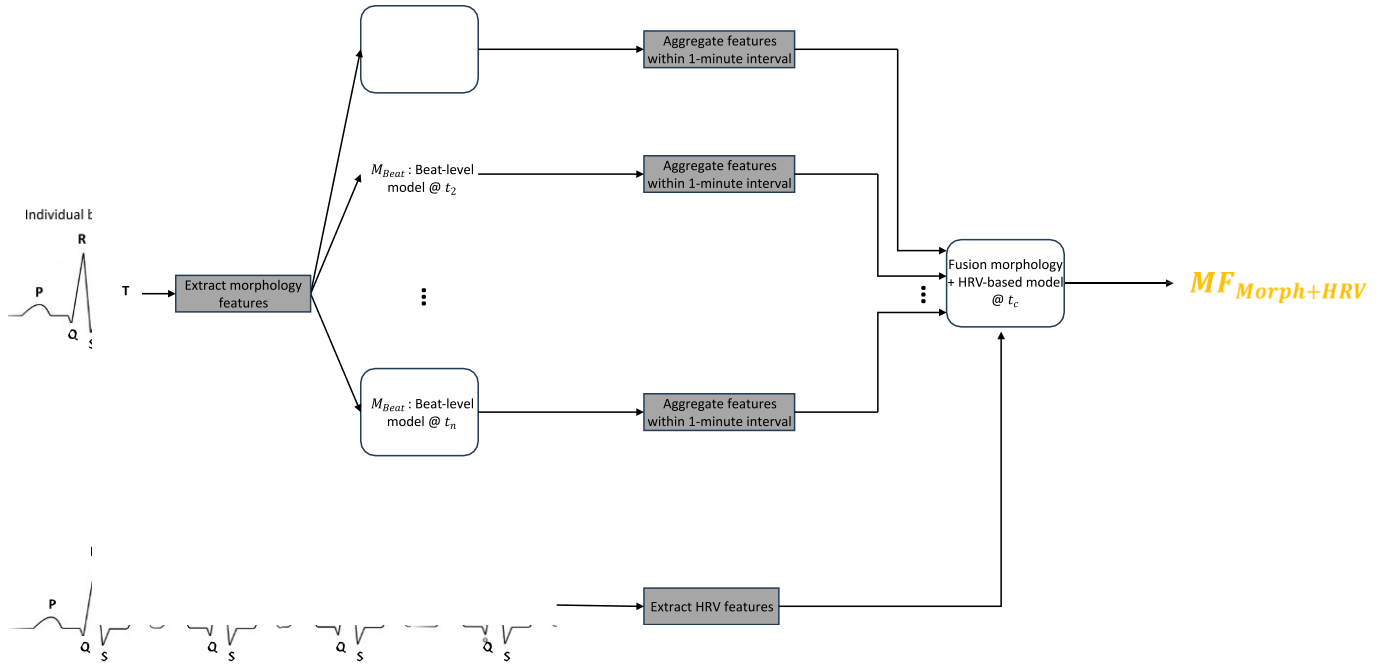


Fig. 4. Workflow of the fusion-based approach used in this study.

thresholds of 150, 165, 180, 200, 225 and 250. We hypothesize that morphological changes may not happen at a specific threshold but across a range of glycemic values, even for an individual subject and at different times. To leverage these varying patterns at different glycemic levels, the interval-level features, derived from the posterior probabilities of multiple beat-level (M_{Beat}) models, trained at different thresholds are concatenated and used as input at the final step of the fusion model to make predictions. The fusion model is denoted by $MF_{Morph+HRV}$. While the fusion models consist of features derived from beat-level models trained at different thresholds, performance evaluation (at the interval-level) is conducted using the conventional thresholds of 70 mg/dL for hypoglycemia and 180 mg/dL for hyperglycemia.

3. Results

Aggregating individual beat-level predictions within an interval is essential for interpretability and making clinical decisions. We evaluate and compare four different aggregation approaches for two binary classification tasks: (a) hypoglycemia detection and (b) hyperglycemia detection. We also propose a fusion-based approach (using morphological changes at varying glycemic excursion thresholds) that improves performance over the baseline model (morphological changes at a single standard clinical threshold). For the following results, M_{Beat} represents the beat-level model. M_{MV} and M_{Morph} are models using aggregated features within an interval based on only beat-level predictions (intra-beat information), M_{HRV} are models based on only HRV features (inter-beat information), whereas $M_{Morph+HRV}$ and $MF_{Morph+HRV}$ are models

using aggregated features for an interval based on beat-level predictions and HRV features (intra-beat and inter-beat information).

3.1. Performance Comparison: Interval-level aggregation

3.1.1. Hypoglycemia

Performance across the four aggregation approaches for hypoglycemia detection are summarized in Fig. 5 (left). The height of individual bars shows the mean AUC across the patients whereas the error bars represent the standard deviation. One-way ANOVA shows there is a statistically significant difference across models ($p < 0.01$). Fig. 5 (right) displays pairwise statistical tests between different aggregation methods as well as compares the reference beat-level M_{Beat} model. Since we are interested in finding the group (feature-set) with the best performance, we perform one-tailed pairwise t-tests ($H_0: \mu_1 = \mu_2, H_a: \mu_1 < \mu_2$) as post-hoc analysis.

Pairwise comparisons show statistical significance with M_{Morph} over M_{HRV} ($p = 0.02$, one-tailed). $M_{Morph+HRV}$ has statistically significant improvement over M_{Morph} ($p = 0.01$, one-tailed) and M_{HRV} ($p < 0.01$, one-tailed). This indicates that models using morphology features combined with HRV features perform better compared to using either set of features independently. When comparing model performance for individual patients, $M_{Morph+HRV}$ improved performance for 7 patients (out of 9) over M_{Morph} and 9 patients (out of 9) over M_{HRV} . Also, $M_{Morph+HRV}$ significantly ($p = 0.03$, one-tailed) improves performance over the reference model, M_{Beat} . The performance of the majority-voting model M_{MV} is significantly lower compared to all other aggregation approaches.

3.1.2. Hyperglycemia

A performance comparison is summarized in Fig. 6 (left) for hyperglycemia detection. One-way ANOVA shows a statistically significant difference across different models ($p < 0.01$). Post-hoc pairwise comparisons ($H_0: \mu_1 = \mu_2, H_a: \mu_1 < \mu_2$) (Fig. 6 – right) show that, similar to hypoglycemia, M_{Morph} is (statistically) significantly better than M_{HRV} ($p < 0.01$, one-tailed). Also, combining both sets of features $M_{Morph+HRV}$ offers a significant advantage over M_{Morph} ($p < 0.01$, onetailed) and M_{HRV} ($p < 0.01$, onetailed) for hyperglycemia detection. The improvement is consistently observed across all 11 patients for hyperglycemia detection. This shows that interval-level features (derived from beat-level predictions) and HRV features complement each other in enhancing prediction performance. Like hypoglycemia detection, $M_{Morph+HRV}$ significantly improves performance over the base model, M_{Beat} ($p < 0.01$, one-tailed).

3.1.3. Performance comparison: fusion-model

Fusion model $MF_{Morph+HRV}$ significantly improves over the baseline model, $M_{Morph+HRV}$ (best-performing aggregation approach) for both hypoglycemia ($p = 0.03$, one-tailed) and hyperglycemia detection ($p = 0.02$, one-tailed) (Fig. 7). Fig. 8 and Fig. 9, give a patient-wise comparison of all the models. $MF_{Morph+HRV}$ improves performance over the baseline model, $M_{Morph+HRV}$ for 8 patients (out of 9) for hypoglycemia detection and 10 patients (out of 11) for hyperglycemia detection. Performance comparison of all the modeling approaches used for hypoglycemia and hyperglycemia detection can be found in Table 3.

3.2. Variable importance

We examined the importance of the features for making predictions at the beat-level (M_{Beat}) and interval-level (M_{MV} , M_{Morph} , M_{HRV} , $M_{Morph+HRV}$, $MF_{Morph+HRV}$) by using Shapley Additive Explanation (SHAP) plots [57] and RF-model-based variable importance plots (VIP). SHAP plots improve interpretability by illustrating the model's decision concerning predictions vs feature values. RF-based VIP evaluates the importance of each feature by measuring the reduction in model performance when excluding a predictor in the training process. We aim to understand the impact of individual features through SHAP plots and the relative importance of feature categories through VIP. To generate the VIP, we aggregate importance scores within the category (e.g.: HRV, interval-level features, etc.).

The SHAP plots, along with actual beat morphology plots, demonstrate the variable importance for hypoglycemia (Fig. 10) and hyperglycemia detection (in Fig. 11) in the M_{Base} model. For a sample patient in Fig. 10, high values of distances: RS, QR, PR, and amplitudes: R; and low values of RR, and distance: RS show an association with hypoglycemia. SHAP plots indicate that features associated with the fiducial point 'R' are pivotal in classifying hypoglycemia for this patient. This aligns with the adjacent beat morphology plots, highlighting distinct differences in the R amplitude region (Fig. 10 right panel). In the case of hyperglycemia detection for patient ID: 10 (Fig. 11), low values of intervals: QT, RT, PT, QR, ST and RR and high HR are driving the model towards predicting hyperglycemia, consistent with morphological changes around the 'T wave' (Fig. 11 right panel).

At the interval-level (Fig. 12), HRV features contribute approximately 42 % towards the prediction performance for both hypoglycemia, and hyperglycemia, while interval-level features derived from the output of M_{Base} model account for about 38 %. These findings corroborate the performance results, underscoring the significance of HRV features and interval-level features in enhancing predictions.

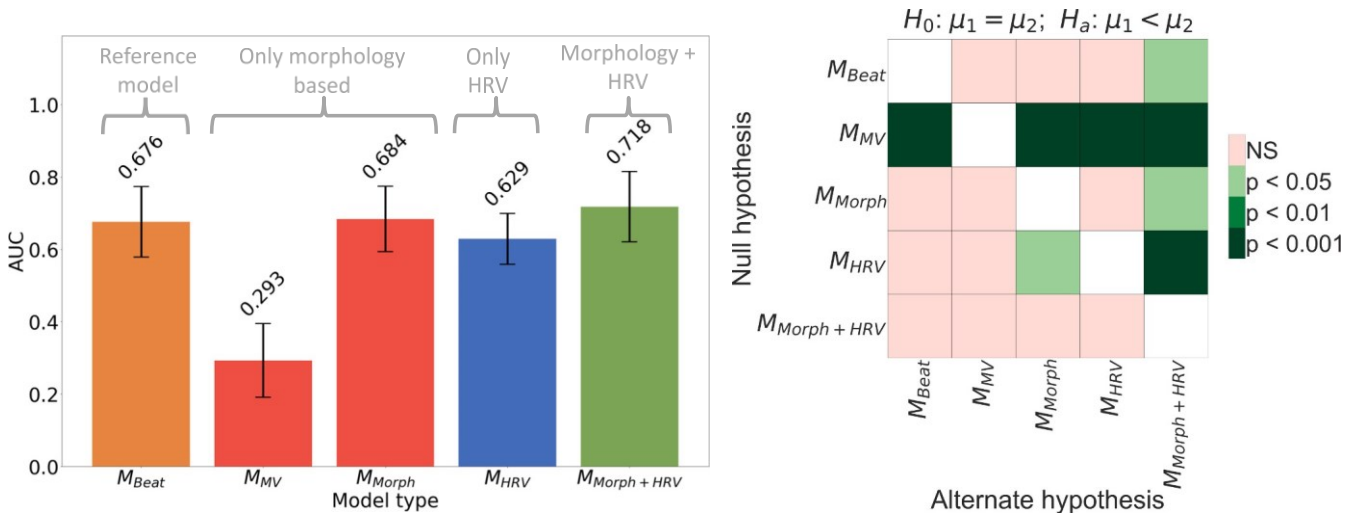


Fig. 5. Hypoglycemia detection: (left) model performance comparison across different aggregation approaches (right) pairwise student's t-test results.

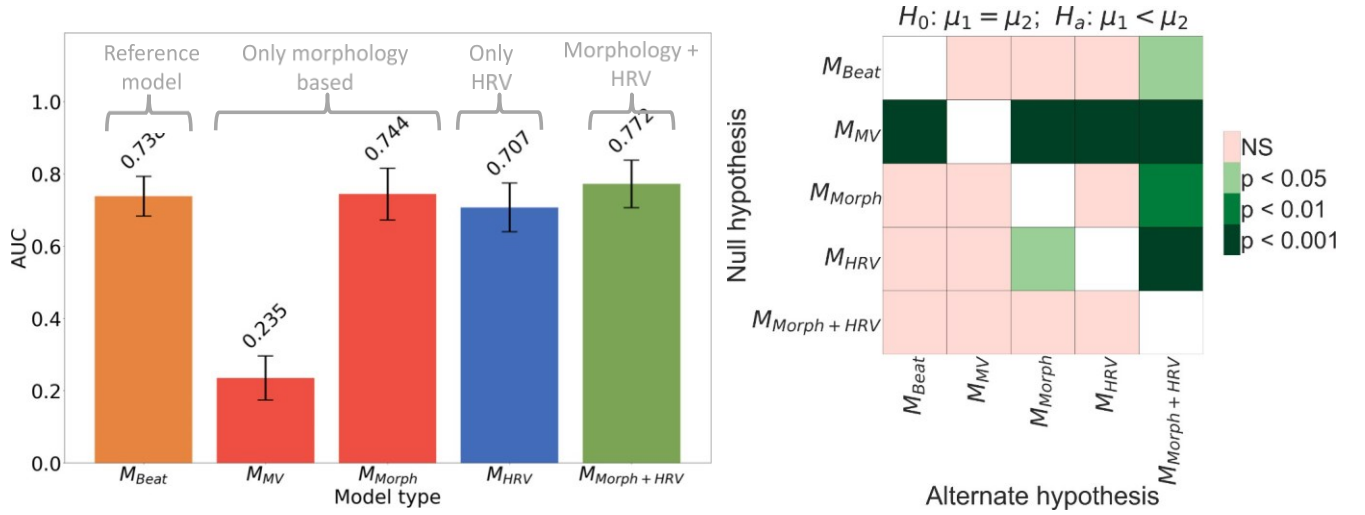


Fig. 6. Hyperglycemia detection: (left) model performance comparison across different aggregation approaches (right) pairwise student's *t*-test results.

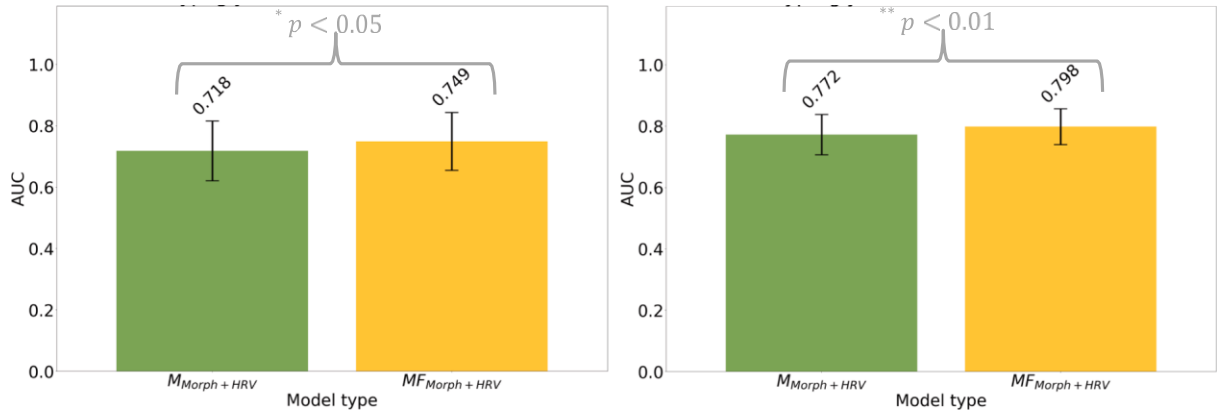


Fig. 7. Performance comparison of the baseline model against fusion model for: (left) hypoglycemia detection and (right) hyperglycemia detection.

In the case of the fusion model (Fig. 13), features extracted from thresholds other than the standard threshold of 70 (i.e., 55, 60, 65, 75, 80, 85, and 90) for hypoglycemia show high importance (61 %) and similarly for hyperglycemia, features extracted from thresholds other than the standard clinical threshold of 180 (i.e., 150, 165, 200, 225 and 250) for hyperglycemia exhibit high importance (57 %). This validates our hypothesis that morphological changes around the neighborhood of standard clinical thresholds for excursions provide useful information, explaining the improvement in the performance of the fusion models.

4. Discussion

4.1. Insights and observations

The relative importance of feature sets in the random forest-based VIP plots (Fig. 12 and Fig. 13) validates the importance of morphology features and HRV features. It also indicates a synergistic relationship between intra-beat and inter-beat information in detecting glycemic excursions. Fig. 14 presents a comparison of selected morphology features across hypoglycemia (< 70 mg/dL), normal (≥ 70 mg/dL and ≤ 180 mg/dL) and hyperglycemia (> 180 mg/dL) ranges using boxplots. This portrayal illustrates morphological changes occurring at different glycemic levels [58,59], contributing to glycemic state classifications as depicted in Figs. 10 and 11. In the case of the fusion-model, it is evident that incorporating morphological change information at different thresholds boosts performance. This observation is

supported by the RF-based importance plots (Fig. 13). A more detailed breakdown of the fusion features at individual threshold-level features further confirms that features from various glycemic thresholds contribute to hypoglycemia and hyperglycemia detection (Appendix F1).

Upon a more in-depth examination of misclassification errors for our best-performing model $MF_{HRV+Morphology}$ (Fig. 15), we observe an increasingly improved true detection rate (TP) for lower glucose readings, ranging from 57 % for readings between 65 mg/dL and 70 mg/dL to 84 % below 55 mg/dL. Similarly, for higher glucose readings the true detection rate ranges from 65 % for readings between 180 mg/dL and 200 mg/dL to 91 % for readings above 350 mg/dL. This indicates that the model demonstrates enhanced effectiveness in detecting severe glycemic excursions, thereby helping in preventing severe hypoglycemia and hyperglycemia.

4.2. Varying excursion thresholds

Learning morphological changes at varying glycemic excursion thresholds plays a key role for the fusion model in outperforming baseline models for hypoglycemia and hyperglycemia detection. Different excursion thresholds result in different numbers of events. Fig. 16 shows the percentage of hypoglycemia and hyperglycemia values within a patient profile for different excursion thresholds. The rate of change for hypoglycemia and hyperglycemia events varies across different patients for hypoglycemia ($pvalue = 0.05$) and hyperglycemia

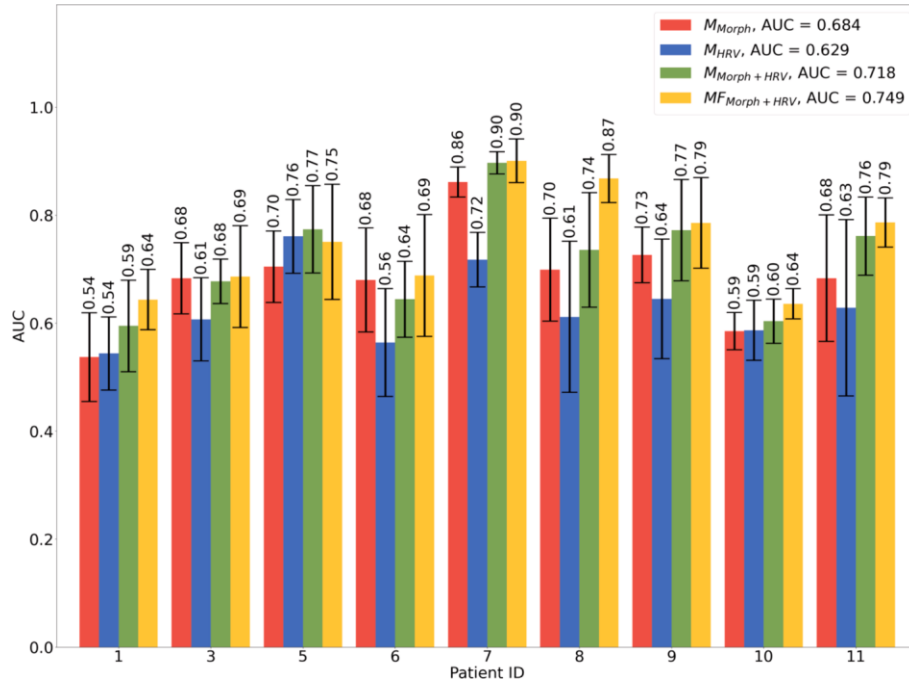


Fig. 8. Patient-wise performance between fusion-based approaches and best-performing interval-level aggregation approach for hypoglycemia detection.

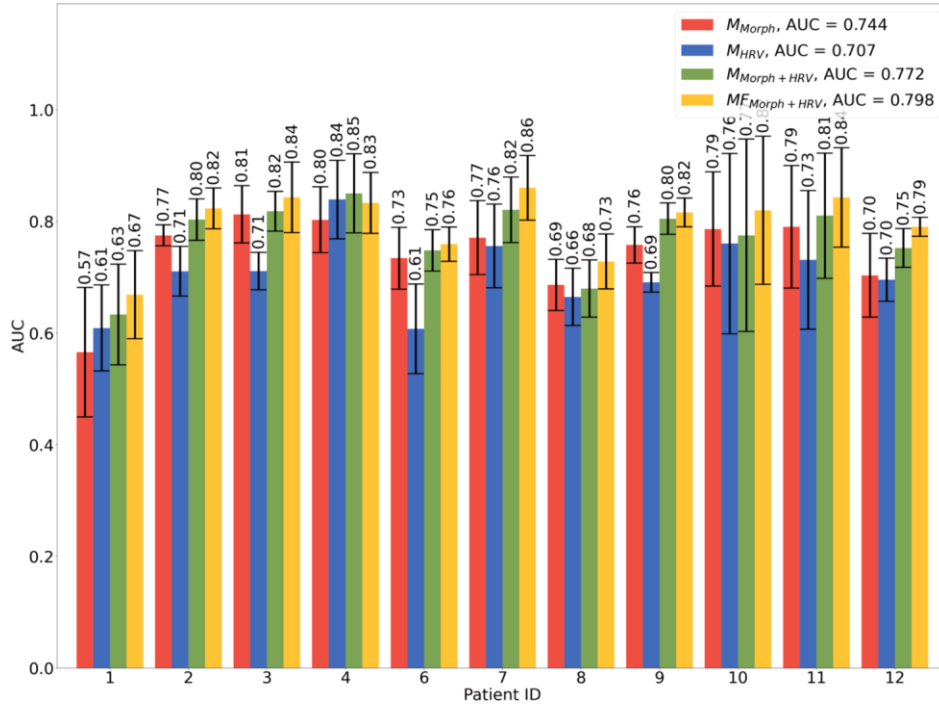


Fig. 9. Patient-wise performance between fusion-based approaches and best-performing interval-level aggregation approach for hyperglycemia detection.

($pvalue < 0.001$). However, when we compare model performance ($MF_{Morph+HRV}$) against rate of change of excursion events, we do not find a significant correlation for hypoglycemia ($pvalue = 0.38$) or hyperglycemia ($pvalue = 0.41$). This corroborates previous findings from the literature [33–36] that although ECG-based features effectively show signs of hypoglycemia and hyperglycemia, the changes observed for different ECG-based features vary across different patients and for different levels of glucose readings.

4.3. Applicability in real-world settings

In this work, we attempt to provide a realistic evaluation of using noninvasive wearables for the detection of glycemic excursion events in the real-world settings. To further understand this, we consider the current work on two aspects:

4.3.1. Model performance

Glucose-related risk prediction through historical CGM readings and noninvasive wearables is a well-researched area. Despite this, previous

Table 3

Performance comparison of different approaches used in the study for hypoglycemia detection and hyperglycemia detection.

Model Details	Hypoglycemia	Hyperglycemia
Beat-level (M_{Beat})	0.676 ± 0.10	0.721 ± 0.10
Majority-Voting (M_{MV})	0.25 ± 0.09	0.47 ± 0.07
Only Morphology features (M_{Morph})	0.684 ± 0.07	0.719 ± 0.07
Only HRV features (M_{HRV})	0.629 ± 0.08	0.708 ± 0.07
Morphology + HRV features ($M_{Morph+HRV}$)	0.718 ± 0.07	0.769 ± 0.07
Fusion (Morphology + HRV features) ($MF_{Morph+HRV}$)	0.749 ± 0.10	0.782 ± 0.10

research has adopted varying approaches for making predictions and reporting results of machine learning models. Predominantly, two main paradigms have been considered: (a) sample-based prediction: prediction is performed at each timestamp of the glucose reading, and (b) event-based prediction: consecutive CGM values (or timestamps) in the hypoglycemia or hyperglycemia range are considered as a single event [60]. Most studies [25,28,35,36] using noninvasive signals, have reported model performance for hypoglycemia and hyperglycemia detection within 1-to-5-minute intervals. These different approaches significantly impact the imbalance ratio and the performance metrics. For instance, the PPV reported in this work (Appendix T7) for a 1-minute interval shows that 9 out of 10 hypoglycemia detections are false alarms. However, evaluating the same model performance at the event level, results in only 2.07 false alarms per day. This approach also results in overlooking short-span hypoglycemia events. Additionally, variations in choosing glycemic thresholds for defining hypoglycemia and hyperglycemia events, different validation approaches (Table 4), and modeling strategies (precision medicine or population-level) further complicate the comparison of different studies and their relevance to real-world settings. Moreover, this article primarily aims to determine the type of information required to be extracted from the ECG signal for accurate glycemic excursion detection. The advantages and disadvantages of using specific intervals for making predictions and an optimal approach for providing real-time predictive alerts need further exploration.

4.3.2. Patient cohort

The dataset used for this study was collected from twelve subjects

with type-1 diabetes over 14 days, encompassing both day and night in free-living conditions. This aspect is critical because it captures natural glucose variations and patterns related to daily physical activities and eating habits. The twelve subjects in this study, aged between 20–40 years included 7 males and 5 females. Personalized models were developed for each individual. However, no significant correlation was found between model performance (for either hypoglycemia or hyperglycemia detection) and demographic factors such as age or gender.

Regarding glycemic profiles, the twelve patients had on average approximately 2 % readings falling within the hypoglycemia range and about 20 % in the hyperglycemia range. This distribution aligns with previous studies [9,61] involving data collected over longer periods and a much larger patient population. Hence, the current cohort is representative of the broader type-1 diabetes population in terms of glycemic profiles. Similarly to demographics, no significant correlation was found between model performance against the glycemic profiles of these participants.

This indicates that the approach used in this study is not limited to subjects of a specific age range, gender groups, or certain glycemic profiles. Our approach is generalizable to the broader diabetes population. A more detailed comparison of this article with other works along with highlighting key similarities and differences is presented in the next section.

4.4. Confounding factors impacting glucose dynamics

As we previously mentioned in the introduction section, multiple direct and indirect factors impact blood glucose levels and their fluctuations. Cardiac autonomic neuropathy (CAN) is an important factor that can lead to impaired heart responses to physiological stimuli like glycemic excursions. A common effect is impaired cardiac responses leading to loss of HRV among such patients [62,63]. This will severely impact glycemic excursion detection through PPG or ECG data as HRV derived from these signals will be different in patients with CAN. In addition to impacting heart rate, CAN among patients can cause abnormal blood pressure patterns, lead to prolonged QT-intervals (can be easily confounded with prolongation due to hypoglycemia) and create an overall imbalance between sympathetic and parasympathetic nervous system [64]. There is a need to conduct studies to assess

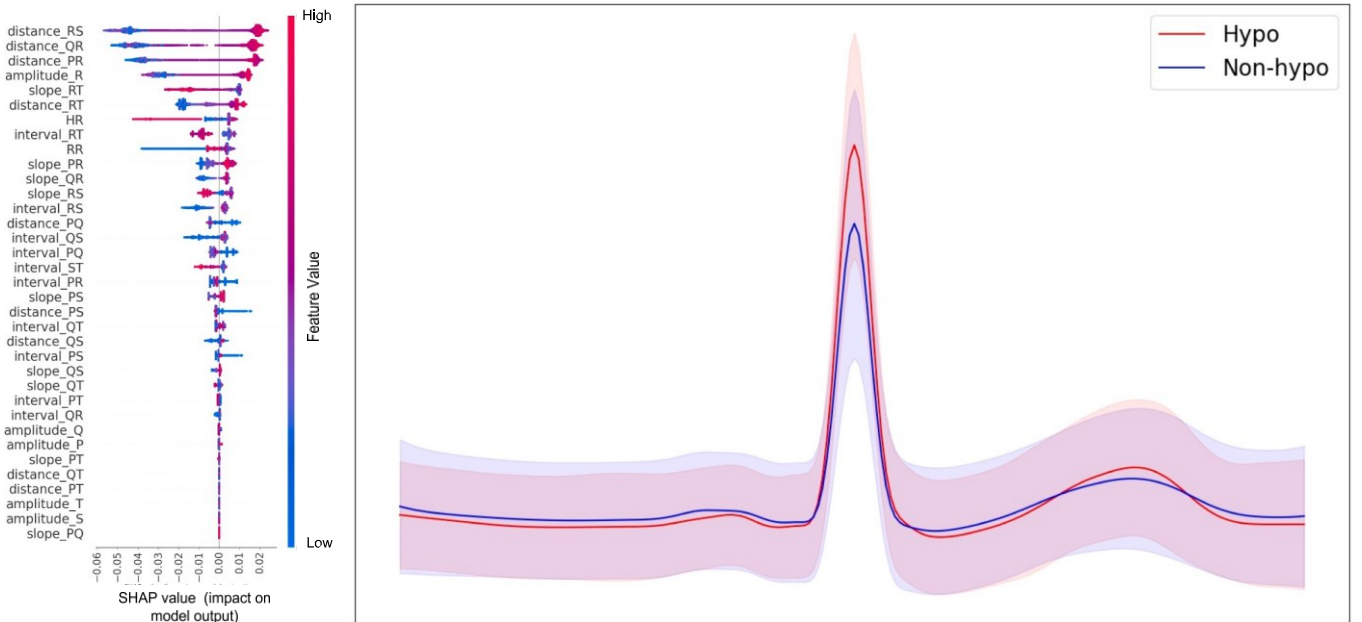


Fig. 10. SHAP plots showing how feature values drive model towards hypoglycemia prediction and associated comparison with raw beat-morphology observed for Patient ID: 5.

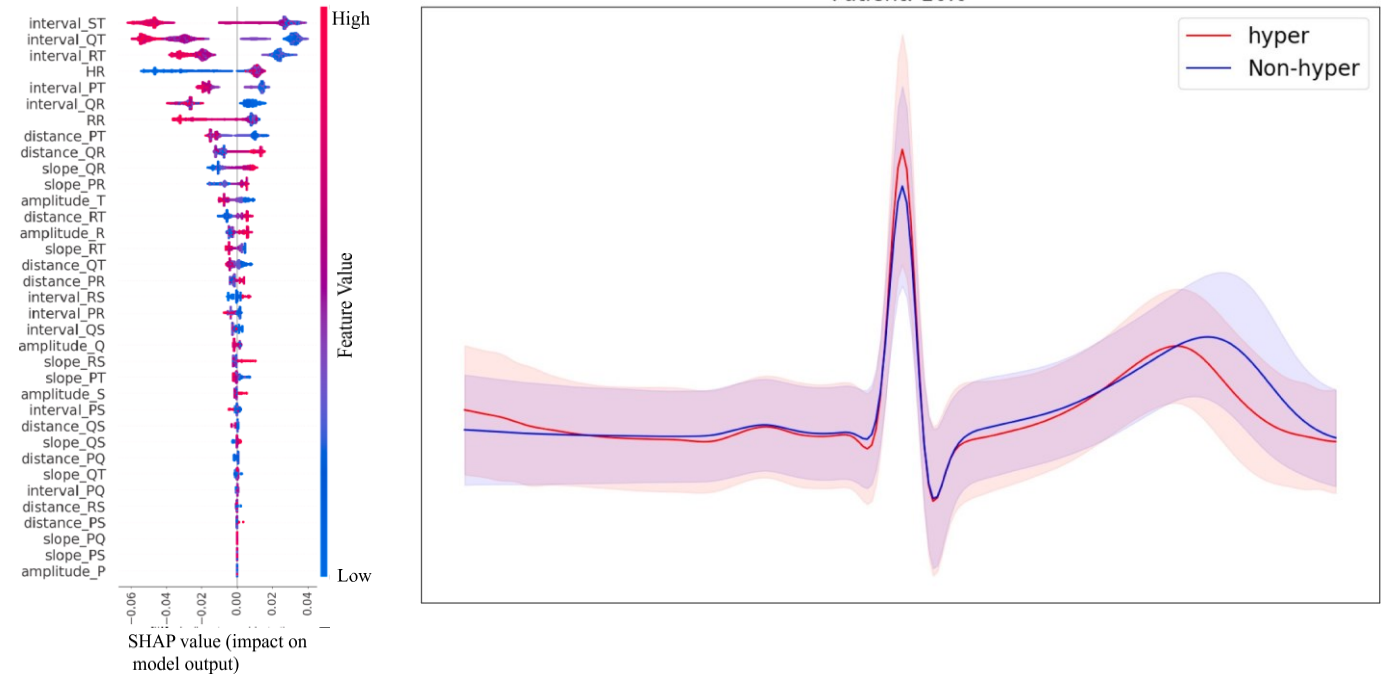


Fig. 11. SHAP plots showing how feature values drive model towards hyperglycemia prediction and associated comparison with raw beat-morphology observed for Patient ID: 10.

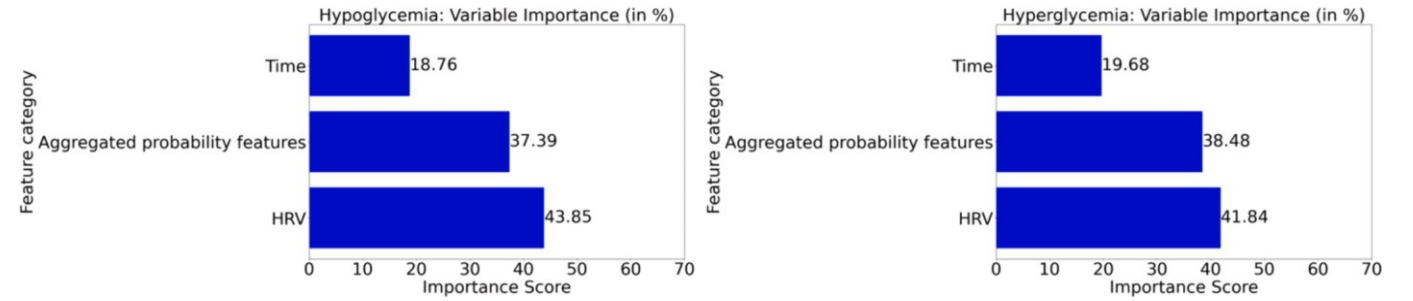


Fig. 12. Interval-level: Relative importance of feature categories for (left) Hypoglycemia and (right) Hyperglycemia detection.

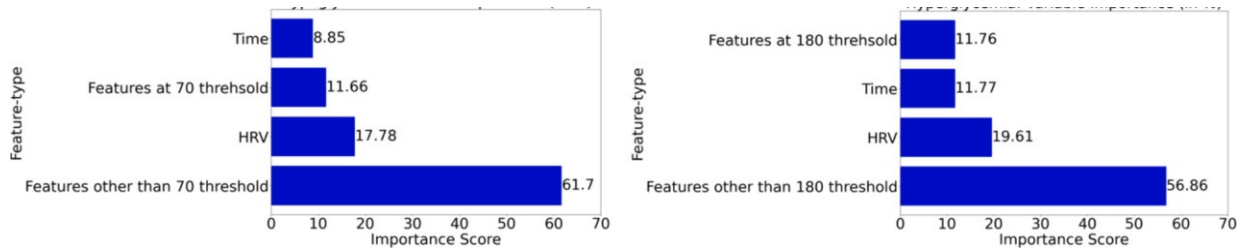


Fig. 13. Fusion-level: Relative importance of feature categories for (left) Hypoglycemia and (right) Hyperglycemia detection.

glycemic excursion detection performance in the presence of comorbidities (e.g., severe neuropathy), and other confounding conditions such as hypoglycemia unawareness which can lead to fatal consequences. Additionally, electrodermal activity which represents the electrical response of sweat glands to sympathetic innervation, was shown to have less pronounced changes during mild hypoglycemia as compared to severe hypoglycemia [36].

4.5. Comparison to previous works

We provide a comparative analysis of previous works studying

noninvasive glycemic predictions, involving hypoglycemia detection, hyperglycemia detection, and actual glucose value prediction. Specifically, we focus on works using ECG signals and related features as the input (Table 4). A predominant theme in the literature relied on morphology features like QT-interval (QT), corrected QT-interval (QTc), and HRV features as inputs in model development. More recently, the advent of deep learning approaches has allowed pattern extraction on ECG signals through CNN, RNN and LSTM-based models. These works have reported a higher prediction performance than earlier works using handcrafted features. However, comparing prediction performance across different studies is challenging due to the differences in data

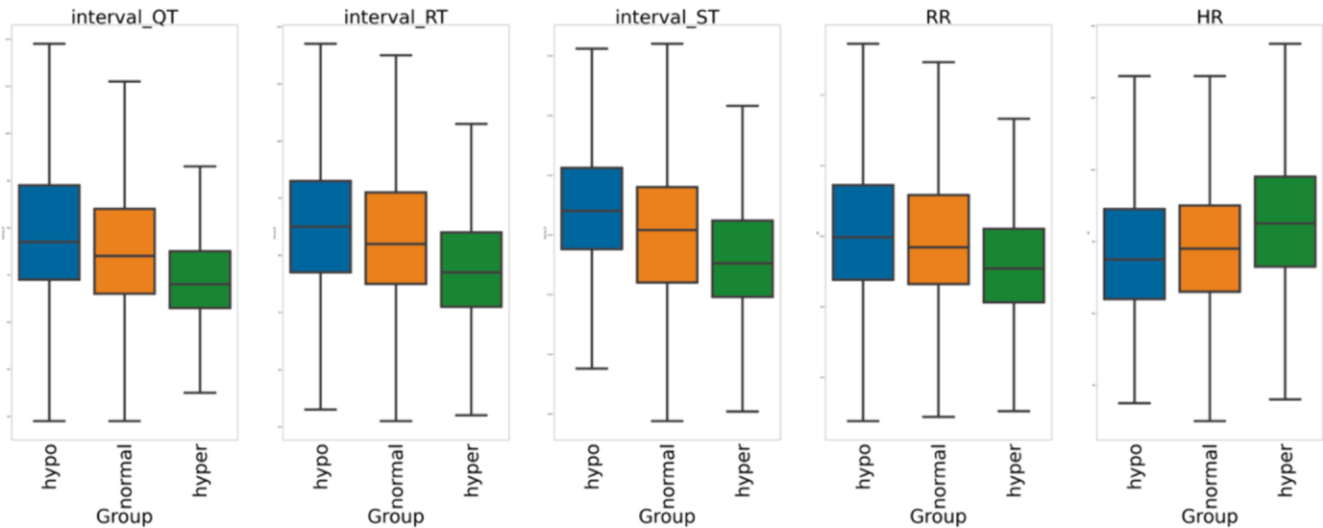


Fig. 14. Comparing HRV features across hypoglycemia, normal and hyperglycemia glucose ranges.

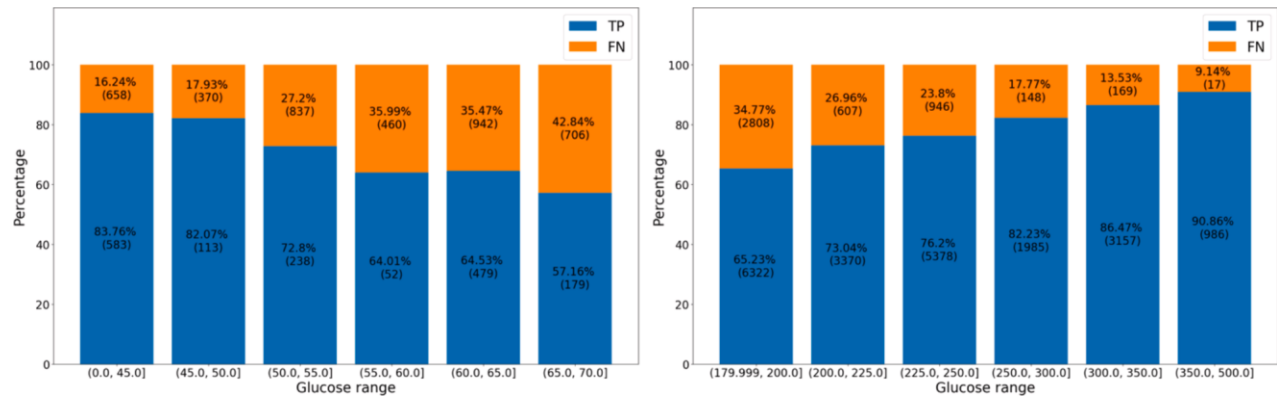


Fig. 15. Analysis of misclassification errors vs. glucose values for: (left) hypoglycemia detection (right) hyperglycemia detection.

A. Varying excursion thresholds

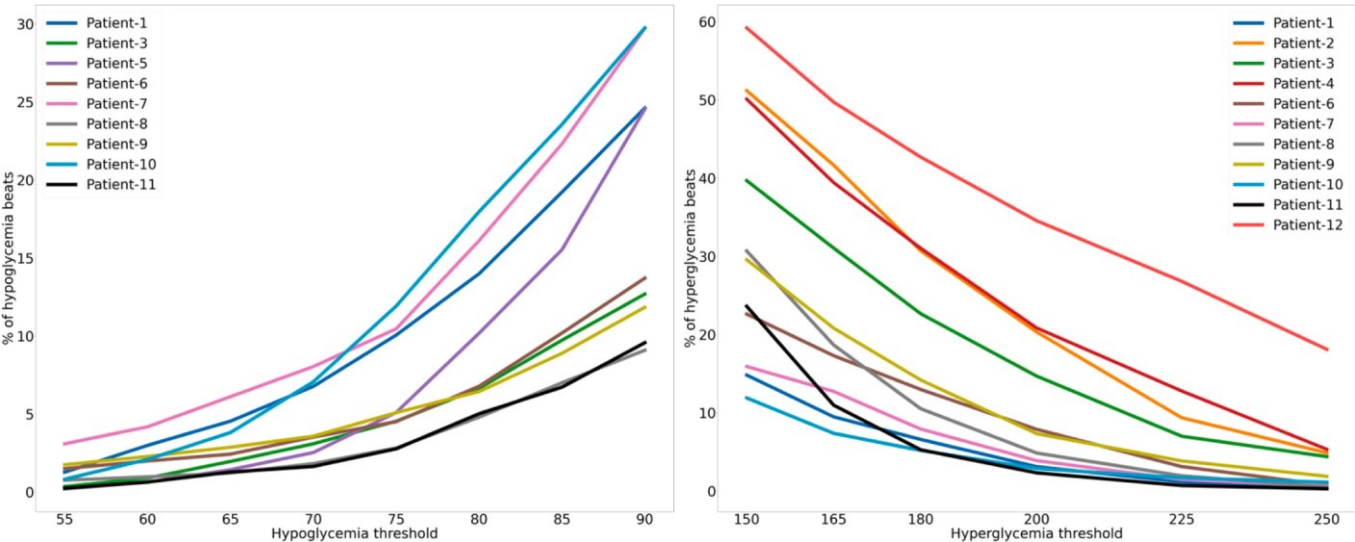


Fig. 16. Number of glycemic excursion events (individual beats) at different excursion thresholds for: (left) hypoglycemia detection and (right) hyperglycemia detection.

Table 4

Comparison of previous literature for noninvasive glucose (hypoglycemia, hyperglycemia) prediction using ECG signal.

Paper	Year	ECG features in the study	Cohort Details	Prediction Tasks	Validation approach	Metrics Used	Performance reported
Hypoglycemia detection based on cardiac repolarization features [22]	2011	Corrected QT-interval, RT-amplitude ratio, RR interval, T-wave slope, T-wave distance onset/offset	Patients: 22 Time: 1-hour recordings Condition: Controlled setting	Hypoglycemia	–	–	Exploratory analysis. Changes detected in 15/22 hypoglycemia events
Genetic-Algorithm-Based Multiple Regression with Fuzzy Inference System for Detection of Nocturnal Hypoglycemic Episodes [21]	2011	Heart Rate (HR), corrected QT (QTc), change in HR and QTc	Patients:16 Time: Overnight time recordings	Hypoglycemia	Patient-based	Sensitivity, Specificity	Sensitivity = 75 % Specificity = 50 %
Non-invasive hypoglycemia monitoring system using extreme learning machine for Type 1 diabetes [20]	2016	Heart Rate (HR), corrected QT (QTc), change in HR and QTc	Condition: Controlled setting		Patient-based	Sensitivity, Specificity	Sensitivity = 78 % Specificity = 60 %
Deep Learning Framework for Detection of Hypoglycemic Episodes in Children with Type 1 Diabetes [65]	2016	Heart Rate (HR), corrected QT (QTc)			Patient-based	Sensitivity, Specificity, Gamma	Sensitivity = 80 % Specificity = 50 % Gamma = 68 %
Paper	Year	ECG features in the study	Cohort Details	Prediction Tasks	Validation approach	Metrics Used	Performance reported
A multiparameter model for non-invasive detection of hypoglycemia [66]	2019	Heart Rate (HR), QT-interval	Patients:20 Time: ~1–2 h Condition: Controlled setting	Patients:20 Time: ~1–2 h Condition: Controlled setting	Random Splitting	AUC, Sensitivity, Specificity	AUC = 94% Sensitivity = 75%, Specificity = 98%
Precision Medicine and Artificial Intelligence: A pilot study on deep learning for hypoglycemic events detection based on ECG [25]	2020	CNN-based morphology features	Patients:8 Time: 14 days Condition: Free living	Hypoglycemia	Random Splitting	Sensitivity, Specificity, Accuracy	Sensitivity = 85%, Specificity = 85%, AUC = 85%
Hyperglycemia Identification using ECG in Deep learning era [67]	2021	Morphology features (distances, slopes, amplitudes)	Patients: 1119 Time: ~2-minute recordings Condition: Controlled setting	Hypoglycemia	Random Splitting	AUC, Sensitivity, Specificity	AUC = 94.53% Sensitivity = 87.57% Specificity = 85.04%
Paper	Year	ECG features in the study	Cohort Details	Prediction Tasks	Validation approach	Metrics Used	Performance reported
Non-invasive Monitoring of Three Glucose Ranges Based on ECG By Using DBSCAN-CNN [26]	2021	CNN-based morphology features	Patients:16 Time: ~1–2 h Condition: Controlled setting	Hypoglycemia, Normal, Hyperglycemia	Random Splitting	Accuracy	Accuracy = 82%
A Prediction Algorithm for Hypoglycemia Based on Support Vector Machine Using Glucose Level and Electrocardiogram[68]	2022	Corrected QT-interval, 5-HRV features (time-domain and frequency domain)	Patients:16 Time: Overnight time recordings Condition: Controlled setting	Hypoglycemia	Random Splitting	Sensitivity, Specificity, Accuracy	Sensitivity = 91%, Specificity = 87%, Accuracy = 89%
Detection of hypoglycemia and hyperglycemia using noninvasive wearable sensors: ECG and accelerometry [27]	2022	HRV features (time-domain)	Patients:5 Time: 14 days Condition: Free-living	Hypoglycemia, Hyperglycemia	Time-based splitting	AUC, Sensitivity, Specificity	Hypoglycemia: AUC = 76% Sensitivity = 69% Specificity = 69% Hyperglycemia: AUC = 82% Sensitivity = 74% Specificity = 74%
Paper	Year	ECG features in the study	Cohort Details	Prediction Tasks	Validation approach	Metrics Used	Performance reported
Noninvasive blood glucose monitoring using spatiotemporal ECG and PPG feature fusion and Weight-based Choquet Integral Multimodel approach [29]	2023	CNN-based spatial morphology features, Temporal statistical features	Patients:16 Time: Overnight time recordings (~ 103 days) Condition: Controlled setting	Glucose values	Random Splitting	RMSE, MARD, Parke's Error Grid Analysis	RMSE = 1.56 (mmol/L) MARD = 13.88 Zone (A+B) = 99.38%
Blood glucose estimation based on ECG signal [69]	2023	CNN-based morphology features	Patients: 3 Time: 8 days Condition: Free living	Glucose values	Random splitting	RMSE	RMSE = 0.47 mg/dL R ² = 82%
This Work	2023	Handcrafted Morphology (distances, intervals, amplitudes, slopes) features	Patients: 12 Time: 14 days Condition: Free living	Hypoglycemia, Hyperglycemia	Time-based splitting	AUC	Hypoglycemia: AUC = 76% Hyperglycemia: AUC = 80%

collection, underlying population characteristics (medical history, age, race, sex, etc.), study protocols, validation approaches used, etc. For example, data collected in controlled sedentary settings for a short duration differs significantly from data observed in free-living conditions. This disparity is critical, impacting data quality, as patterns observed in free-living conditions can deviate significantly from the controlled settings due to the influence of various external factors like motion artifacts, food consumption, etc.

From a modeling perspective, the focus of prediction (e.g., hypoglycemia or hyperglycemia detection or glucose value prediction) and validation methods are key aspects that vary across different works in the literature. Even when concentrating on the same prediction task, such as hypoglycemia detection, multiple studies have reported performance at the individual beat-level, intervals of one minute, five-minute or fixed sequence of beats. In this work, we adhere to the most commonly used 1-minute interval for reporting our results. Given that ECG data and glucose readings are temporal data streams, it is imperative to employ an appropriate validation scheme to provide more realistic estimates of performance when deploying the model. Although commonly used, random splitting can give an overly optimistic performance estimate. To underscore this, we compare the results of our model using random splitting-based validation. (Appendix T6).

4.6. Limitations

As previously discussed, multiple direct and indirect factors impact blood glucose levels and its fluctuations. It is important to study and evaluate these changes in blood glucose levels in the presence of other comorbidities. However, the scope of this study was limited to the ECG signal alone. An important factor to consider in technology-based excursion detection is diabetes-related autonomic neuropathy, as it can impair cardiac responses during glycemic excursions increasing the risk of severe hypoglycemia [70]. The subjects in this study were not specifically tested for cardiac autonomic neuropathy (CAN) due to their young age and shorter duration since diabetes diagnosis [71]. Nevertheless, understanding the effect of CAN on glycemia-induced morphological changes is important and is a limitation of the current study. Another limitation of our work is the relatively modest size of our dataset. Our dataset comprised recordings from twelve participants over a period of 14-days. Among these twelve patients, only nine were included in the hypoglycemia detection model, and eleven in the hyperglycemia detection model, based on the number of glycemic excursion events they experienced. In total, our analysis included 215 hypoglycemia episodes and 314 hyperglycemia episodes. To overcome this limitation, we implemented a 5-fold temporal cross-validation mechanism to ensure accurate estimation of model performance despite the constraints in data availability. Most prior works (from Table 4) have employed random data splitting for model training, neglecting temporal correlations in the data, which may inflate model performance due to temporal correlations, potentially leading to overly optimistic results. Appendix T5 presents the higher performance results achieved when random splitting-based validation is used instead of a more conservative validation approach adopted in this paper. Despite the many advantages of the 5-fold temporal cross-validation approach, it may pose challenges for including patients with an extremely low number of glycemic excursion events. This is because splitting the data into 5-folds for cross-validation could result in some folds lacking glycemic excursion events. Training a binary classification model requires both positive and negative class events, and when no positive class events are present, training a binary classifier becomes impossible. As a result, we cannot use such patients for analysis in this study. We acknowledge that the low precision reported in this work can be a deterrent to the adoption of technology among patients. However, our main focus was on extracting key features from ECG signals to detect glycemic excursions and exploring the hypothesis that ECG morphological changes occur at different glycemic thresholds, which can

improve predictive model performance. For giving out real-world predictive alerts, a comprehensive evaluation based on both, clinical as well as human behavior aspects is necessary. To address these challenges, we plan to focus on sustained hypoglycemia [12,61] and consider longer intervals for evaluating predictive alerts in our future research.

Significant improvements in model performance are necessary for noninvasive technologies to replace CGM devices effectively in patients with diabetes. Using multi-modal data such as PPG, electrodermal activity (EDA), BioZ, skin temperature etc. in addition to ECG can enhance performance to acceptable levels and will be the focus of our future efforts. All the patients in the current study were aged between 20–40 years. It would be valuable to collect and analyze data from younger individuals (aged 0–20 years) and older populations (>50 years) with type-1 diabetes as it can cover different lifestyles as well as risk associated with specific age-groups.

5. Conclusion

This article proposes a noninvasive approach to detecting glycemic excursions using ECG data. We explore the use of, intra-beat (morphology features) information and inter-beat (HRV features) information, evaluating different aggregation approaches for predictive efficacy. Our findings indicate that, models incorporating morphology-based features along with HRV features significantly outperform those using only morphology features or HRV features independently both hypoglycemia and hyperglycemia. The combination of morphology and HRV features achieves an average AUC of 72 % for hypoglycemia detection and 77 % for hyperglycemia detection. Our proposed fusion model adeptly learns morphological patterns at multiple glycemic thresholds, enhancing its ability to learn the morphological changes more effectively in the hypoglycemia and hyperglycemia ranges. The fusion model achieves an average AUC of 75 % for hypoglycemia detection and 78 % for hyperglycemia detection. These results underscore several key insights: (1) ECG signal is a promising noninvasive alternative for detecting glycemic excursions, (2) HRV is a useful supplement to morphology features for detecting glycemic excursions, (3) Higher performance of the fusion model indicates a need to explore personalized glycemic thresholds for hypoglycemia and hyperglycemia. These findings have important clinical implications for non-invasive diabetes monitoring and management.

CRedit authorship contribution statement

Darshit Dave: Writing – review & editing, Writing – original draft, Visualization, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Kathan Vyas:** Writing – review & editing, Visualization, Validation, Methodology, Data curation, Conceptualization. **Gerard L. Cote:** Writing – review & editing, Project administration, Investigation, Data curation, Conceptualization. **Madhav Erraguntla:** Writing – review & editing, Writing – original draft, Supervision, Resources, Methodology, Conceptualization.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Madhav Erraguntla reports financial support was provided by National Science Foundation. Gerard Cote reports financial support was provided by National Science Foundation. Darshit Dave reports financial support was provided by National Science Foundation. Kathan Vyas reports financial support was provided by National Science Foundation.

Data availability

The data that has been used is confidential.

Acknowledgement

This research was supported by NSF award 2037383 (SenSE: Multi-modal noninvasive wearable sensors and machine learning for

predicting critical glycemic events). Dr. Ricardo Osuna-Gutierrez contributed to the idea that glycemic threshold of 70 mg/dL is not universal and provided feedback for our ideas and effort.

Appendix

Appendix T1

Minimum information about clinical artificial intelligence modeling via the MI-CLAIM checklist

Study design (Part 1)	Completed	Notes
The clinical problem in which the model will be employed is clearly detailed in the paper.	X (Introduction)	Hypoglycemia and hyperglycemia detection using ECG
The research question is clearly stated.	X (Introduction)	1) Is intra-beat, inter-beat or a combination of both is need for detection of glycemic excursions? Are morphological patterns observed at varying glucose thresholds?
The characteristics of the cohorts (training and test sets) are detailed in the text.	X (Methods and Materials)	We do not have separate train and test cohorts. But we apply a five-fold temporal cross-validation for training personalized models. Details in Fig. 1.
The cohorts (training and test sets) are shown to be representative of real-world clinical settings.	X (Methods and Materials)	Detailed about validation provided in methods and materials section (Fig. 1). Also, provided explanations on its real-world relevance.
The state-of-the-art solution used as a baseline for comparison has been identified and detailed.	X (Discussion)	Detail comparison to state-of-the-art works with relevant differences provided in Discussion section.
Data and optimization (Parts 2, 3)	Completed	Notes
The origin of the data is described and the original format is detailed in the paper.	X (Methods and materials, Appendix)	Details on protocols and relevant patient characteristics provided in Appendix
Transformations of the data before it is applied to the proposed model are described.	X (Methods and Materials)	
The independence between training and test sets has been proven in the paper.	X (Methods)	We use a temporal 5-fold cross-validation to mimic a real-world setting and ensuring no data leakage.
Details on the models that were evaluated, and the code developed to select the best model are provided.	X (Methods and Materials)	
Is the input data type structured or unstructured?	Structured	
Model performance (Part 4)	Completed	Notes
The primary metric selected to evaluate algorithm performance (e.g., AUC, F-score, etc.), including the justification for selection, has been clearly stated.	X (Methods and materials, Appendix)	Primarily Area under the receiver operating curve (AUC-ROC). Other metrics provided in Appendix.
The primary metric selected to evaluate the clinical utility of the model (e.g., PPV, NNT, etc.), including the justification for selection, has been clearly stated.	X (Methods and materials)	
The performance comparison between baseline and proposed model is presented with the appropriate statistical significance.	X (Results, Discussion, Appendix)	Our proposed approach provides robust performance. Comparison with other works in Discussion section and some extra comparison in the Appendix
Model examination (Part 5)	Completed	Notes
Examination technique 1	X (Methods and Results, Discussion)	Shapley Additive Explanations (SHAP) summary plots to show the impact of a predictor on the model performance as a function of the predictor value. Random Forests based variable importance plots (VIMP) to show impact of feature categories on the output variable.
A discussion of the relevance of the examination results with respect to model/algorithm performance is presented.	X (Discussion)	
A discussion of the feasibility and significance of model interpretability at the case level if examination methods are uninterpretable is presented.	X (Results, Discussion)	Detailed explanation on the relevance of the results discussed
A discussion of the reliability and robustness of the model as the underlying data distribution shifts is included.	X (Discussion)	Discussion on the study cohort in this article and the relevant of model performance in real-world settings.
Reproducibility (Part 6): choose appropriate tier of transparency	Completed	Notes
Tier 1: complete sharing of the code	X	We are in the process to upload the entire python code on our GitHub repository soon. However, sharing of data will not be made possible at this time. The authors plan to release the entire dataset in the future through appropriate channels.
Tier 2: allow a third party to evaluate the code for accuracy/fairness; share the results of this evaluation		
Tier 3: release of a virtual machine (binary) for running the code on new data without sharing its details		
Tier 4: no sharing		

Appendix T2

Demographics data on subject who are part of the study.

Subject ID	Age	Gender	Body weight (lb)	Height	Race/Ethnicity	HbA1C
1	41	M	152	5ft 10in	white	5.5
2	24	M	195	6ft 3in	white	6.8
3	34	M	224	5ft 8in	white	6.9
4	34	F	157	5ft 4in	white	6.4
5	35	F	170	5ft 11in	White	5.4
6	30	M	200	6ft4in	White	5.9
7	31	F	132	5ft2in	White	—
8	30	M	160	5ft 10in	Multiple	6.5
9	30	F	182	5ft 6in	White	5.8
10	29	M	183	6ft 2in	white	6.0
11	35	M	185	6ft 2in	White	5.8
12	37	F	138	5ft 5in	White	7.2

Appendix T3

Hypoglycemia profile of patients who are part of the study.

Patient ID	Total CGM readings	Hypoglycemia CGM readings	Hypoglycemia events/episodes
1	2408	139	33
2	3068	8	2
3	2840	91	19
4	2251	5	1
5	3195	108	24
6	3379	134	22
7	2815	171	18
8	3268	58	12
9	3758	128	20
10	3412	234	59
11	3610	61	7
12	2637	4	2

Appendix T4

Hyperglycemia profile of patients who are part of the study.

Patient ID	Total CGM readings	Hyperglycemia CGM readings	Hyperglycemia events/episodes
1	2408	165	19
2	3068	856	41
3	2840	660	39
4	2251	586	37
5	3195	26	2
6	3379	437	25
7	2815	166	11
8	3268	355	36
9	3758	497	37
10	3412	145	13
11	3610	184	15
12	2637	1070	41

Appendix T5

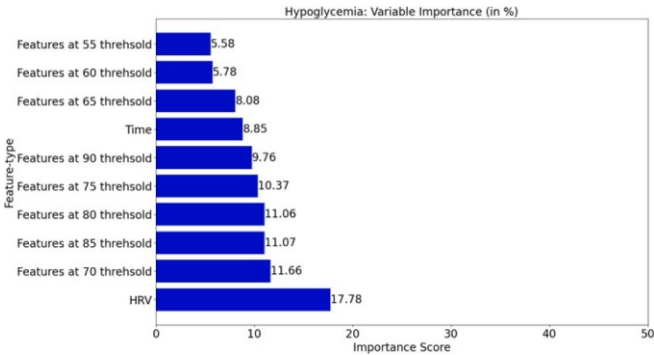
HRV features used in the study.

Feature	Description
MeanNN (ms)	Mean NN Intervals
CVNN	Coefficient of variation of NN intervals $\frac{SDNN}{MeanNN}$
CVSD	Coefficient of variation of successive differences $\frac{RMSSD}{MeanNN}$
SDNN (ms)	Standard deviation of NN intervals
SDSD (ms)	Standard deviation of Successive differences
RMSSD (ms)	Root mean square of successive interval differences
Pnni_20 (%)	Proportion of successive NN intervals that differ by > 20 ms
Pnni_50 (%)	Proportion of successive NN intervals that differ by > 50 ms
HR_Mad	Mean absolute deviation of NN intervals
MedianNN (ms)	Median NN Intervals
MinNN	Minimum of NN Intervals

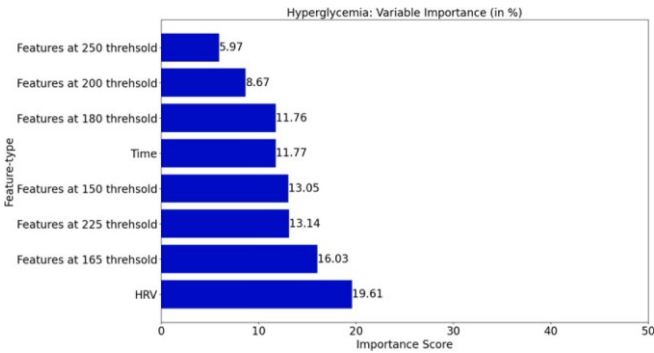
(continued on next page)

Appendix T5 (continued)

Feature	Description
MaxNN	Maximum of NN Intervals
MCVNN	Ratio of $\frac{MadNN}{MedianNN}$
IQRNN	Inter-quartile range of NN intervals
Prc20NN	20th percentile of NN intervals
Prc80NN	80th percentile of NN intervals
HTI	HRV triangular index. Integral of the density of the RR interval histogram divided by its height
TINN	Baseline width of the RR interval histogram



Appendix F1. Fusion-level: Detailed importance of feature categories for hypoglycemia detection.

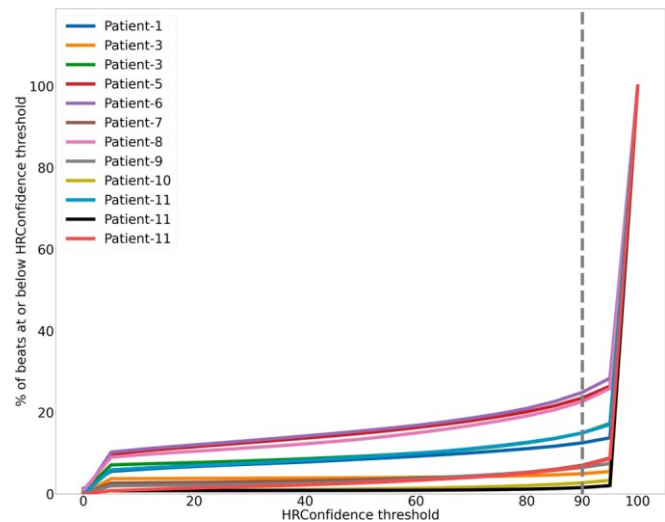


Fusion-level: Detailed breakdown of relative importance of feature categories for hyperglycemia detection

Appendix T6

Performance comparison for different model types and glycemic detection condition in the (proposed) temporal-splitting validation vs random-splitting (at CGM level) validation approach.

Model Details/Validation Approach	Hypoglycemia		Hyperglycemia	
	Time splitting	Random splitting	Time splitting	Random splitting
Beat-level (M_{Beat})	0.676 ± 0.10	0.867 ± 0.10	0.737 ± 0.10	0.868 ± 0.10
Only Morphology features (M_{Morph})	0.684 ± 0.07	0.901 ± 0.07	0.743 ± 0.07	0.934 ± 0.07
Only HRV features (M_{HRV})	0.629 ± 0.08	0.815 ± 0.07	0.707 ± 0.07	0.845 ± 0.07
Morphology + HRV features ($M_{Morph+HRV}$)	0.718 ± 0.07	0.913 ± 0.07	0.772 ± 0.07	0.941 ± 0.07
Fusion (Only –Morphology features) (MF_{Morph})	0.740 ± 0.10	0.929 ± 0.10	0.789 ± 0.10	0.946 ± 0.10
Fusion (Morphology + HRV features) ($MF_{Morph+HRV}$)	0.749 ± 0.10	0.932 ± 0.10	0.798 ± 0.10	0.946 ± 0.10



Appendix F2. Percentage of detected beats at various HRConfidence thresholds.

Appendix T7

Performance comparison for different model types with additional evaluation metrics.

Model type	Hypoglycemia						
	AUC	Sensitivity	Specificity	Precision	F1-score	Hypoglycemia events	
Only morphology	0.68 ± 0.09	0.63 ± 0.1	0.67 ± 0.06	0.08 ± 0.03	0.14 ± 0.04	5519	135,014
Only HRV	0.62 ± 0.14	0.54 ± 0.18	0.65 ± 0.07	0.03 ± 0.02	0.06 ± 0.03		
Morphology + HRV	0.71 ± 0.09	0.66 ± 0.08	0.68 ± 0.07	0.09 ± 0.03	0.15 ± 0.05		
Fusion: Morphology + HRV	0.75 ± 0.1	0.69 ± 0.08	0.70 ± 0.06	0.1 ± 0.03	0.17 ± 0.05		
Model type	Hyperglycemia						
	AUC	Sensitivity	Specificity	Precision	F1-score	Hyperglycemic events	
Only morphology	0.68 ± 0.09	0.63 ± 0.1	0.67 ± 0.06	0.08 ± 0.03	0.14 ± 0.04	25,774	141,767
Only HRV	0.71 ± 0.07	0.66 ± 0.05	0.66 ± 0.05	0.26 ± 0.04	0.35 ± 0.05		
Morphology + HRV	0.77 ± 0.07	0.71 ± 0.06	0.71 ± 0.05	0.31 ± 0.04	0.40 ± 0.05		
Fusion: Morphology + HRV	0.80 ± 0.06	0.73 ± 0.05	0.73 ± 0.05	0.32 ± 0.05	0.42 ± 0.05		

References

[1] H. Sun, P. Saeedi, S. Karuranga, et al., IDF Diabetes Atlas: Global, regional and country-level diabetes prevalence estimates for 2021 and projections for 2045, *Diabetes. Res. Clin. Pract.* 183 (2022) 109119, <https://doi.org/10.1016/j.diabres.2021.109119> [published Online First: 20211206].

[2] T. Battelino, T. Danne, R.M. Bergenstal, et al., Clinical targets for continuous glucose monitoring data interpretation: recommendations from the international consensus on time in range, *Diabetes. Care* 42 (8) (2019) 1593–1603, <https://doi.org/10.2337/dc19-0028> [published Online First: 20190608].

[3] R.W. Beck, R.M. Bergenstal, T.D. Riddlesworth, et al., Validation of time in range as an outcome measure for diabetes clinical trials, *Diabetes. Care* 42 (3) (2019) 400–445, <https://doi.org/10.2337/dc18-1444>.

[4] Y.B. Zhu, R.K. Mehta, M. Erraguntla, F. Sasangohar, K. Qaraqe, Quantifying accelerometer-based tremor features of neuromuscular fatigue in healthy and diabetic adults, *IEEE. Sens. J.* 20 (19) (2020) 11183–11190, <https://doi.org/10.1109/Jsen.2020.2996372>.

[5] K. Zahed, F. Sasangohar, R. Mehta, M. Erraguntla, K. Qaraqe, Diabetes management experience and the state of hypoglycemia: national online survey study, *JMIR. Diabetes* 5 (2) (2020) e17890.

[6] K. Zahed, R. Mehta, M. Erraguntla, K. Qaraqe, F. Sasangohar, Understanding patient beliefs in using technology to manage diabetes: path analysis model from a national web-based sample, *JMIR. Diabetes* 8 (2023) e41501.

[7] O. Tyagi, Y. Zhu, C. Johnson, et al., Neural signatures of handgrip fatigue in type 1 diabetic men and women, *Front. Hum. Neurosci.* 14 (2020) 564969, <https://doi.org/10.3389/fnhum.2020.564969> [published Online First: 20201109].

[8] M. Qaraqe, M. Erraguntla, D. Dave, AI and machine learning in diabetes management: opportunity, status, and challenges, *Multiple. Perspect. Artificial. Intellig. Healthcare: Opportunities. Challenges* (2021) 129–141.

[9] D. Dave, D.J. DeSalvo, B. Haridas, et al., Feature-based machine learning model for real-time hypoglycemia prediction, *J. Diabetes. Sci. Technol.* 15 (4) (2021) 842–855, <https://doi.org/10.1177/1932296820922622> [published Online First: 20200601].

[10] A.C. Larme, J.A. Pugh, Attitudes of primary care providers toward diabetes: barriers to guideline implementation, *Diabetes. Care* 21 (9) (1998) 1391–1396, <https://doi.org/10.2337/diacare.21.9.1391>.

[11] S.H. Atkin, A. Dasmahapatra, M.A. Jaker, M.I. Chorost, S. Reddy, Fingerstick glucose determination in shock, *Ann. Intern. Med.* 114 (12) (1991) 1020–1024, <https://doi.org/10.7326/0003-4819-114-12-1020>.

[12] D. Dave, M. Erraguntla, M. Lawley, et al., Improved low-glucose predictive alerts based on sustained hypoglycemia: model development and validation study, *JMIR. Diabetes* 6 (2) (2021) e26909.

[13] D. Rodbard, Continuous glucose monitoring: a review of recent studies demonstrating improved glycemic outcomes, *Diabetes. Technol. Ther.* 19 (S3) (2017) S25–S37, <https://doi.org/10.1089/dia.2017.0035>.

[14] P. Jain, A.M. Joshi, S. Mohanty, Everything you wanted to know about noninvasive glucose measurement and control. *arXiv preprint arXiv:2101.08996* 2021.

[15] R. Jahromi, K. Zahed, F. Sasangohar, M. Erraguntla, R. Mehta, K. Qaraqe, Hypoglycemia detection using hand tremors: home study of patients with type 1 diabetes, *JMIR. Diabetes* 8 (1) (2023) e40990.

[16] L. Aljihmani, O. Kerdjidi, Y. Zhu, et al., Classification of fatigue phases in healthy and diabetic adults using wearable sensor, *Sensors (Basel)* 20 (23) (2020) 6897, <https://doi.org/10.3390/s20236897> [published Online First: 20201203].

[17] L. Aljihmani, O. Kerdjidi, G. Petrovski, et al., Hand tremor-based hypoglycemia detection and prediction in adolescents with type 1 diabetes, *Biomed. Signal.*

- Process. Control 78 (2022) 103869, <https://doi.org/10.1016/j.bspc.2022.103869>. ARTN 103869.
- [18] W. Villena Gonzales, A.T. Mobashsher, A. Abbosh, The progress of glucose monitoring—a review of invasive to minimally and non-invasive techniques, devices and sensors, *Sensors (Basel)* 19 (4) (2019) 800, <https://doi.org/10.3390/s19040800> [published Online First: 20190215].
- [19] F.A. Koopman, M.A. van Maanen, M.J. Vervoordeeldonk, P.P. Tak, Balancing the autonomic nervous system to reduce inflammation in rheumatoid arthritis, *J. Intern. Med* 282 (1) (2017) 64–75, <https://doi.org/10.1111/joim.12626> [published Online First: 20170526].
- [20] S.H. Ling, P.P. San, H.T. Nguyen, Non-invasive hypoglycemia monitoring system using extreme learning machine for Type 1 diabetes, *ISA. Trans* 64 (2016) 440–446, <https://doi.org/10.1016/j.isatra.2016.05.008> [published Online First: 20160613].
- [21] S.S. Ling, H.T. Nguyen, Genetic-algorithm-based multiple regression with fuzzy inference system for detection of nocturnal hypoglycemic episodes, *IEEE. Trans. Inf. Technol. Biomed* 15 (2) (2011) 308–315, <https://doi.org/10.1109/TITB.2010.2103953> [published Online First: 20110224].
- [22] Hypoglycemia detection based on cardiac repolarization features, 2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society; 2011. IEEE.
- [23] S.L. Cichosz, J. Frystyk, L. Tarnow, J. Fleischer, Combining information of autonomic modulation and CGM measurements enables prediction and improves detection of spontaneous hypoglycemic events, *J. Diabetes. Sci. Technol* 9 (1) (2015) 132–137, <https://doi.org/10.1177/1932296814549830> [published Online First: 20140912].
- [24] T.F. Christensen, S.L. Cichosz, L. Tarnow, et al., Hypoglycaemia and QT interval prolongation in type 1 diabetes - bridging the gap between clamp studies and spontaneous episodes, *J. Diabetes. Complications* 28 (5) (2014) 723–778, <https://doi.org/10.1016/j.jdiacomp.2014.03.007>.
- [25] M. Porumb, S. Stranges, A. Pescapè, L. Pecchia, Precision medicine and artificial intelligence: a pilot study on deep learning for hypoglycemic events detection based on ECG, *Sci. Rep* 10 (1) (2020) 170, <https://doi.org/10.1038/s41598-019-56927-5> [published Online First: 20200113].
- [26] J. Li, I. Tobore, Y. Liu, A. Kandwal, L. Wang, Z. Nie, Non-invasive Monitoring of three glucose ranges based on ECG by using DBSCAN-CNN, *IEEE. J. Biomed. Health. Inform* 25 (9) (2021) 3340–3350, <https://doi.org/10.1109/JBHI.2021.3072628> [published Online First: 20210903].
- [27] D. Dave, K. Vyas, K. Branan, et al., Detection of hypoglycemia and hyperglycemia using noninvasive wearable sensors: ECG and accelerometry, *J. Diabetes Sci. Technol.*, 2022, 19322968221116393.
- [28] V. Lehmann, S. Foll, M. Maritsch, et al., Noninvasive hypoglycemia detection in people with diabetes using smartwatch data, *Diabetes Care* 46 (5) (2023) 993–997, <https://doi.org/10.2337/dc22-2290>.
- [29] J. Li, J. Ma, O.M. Omisore, et al., Noninvasive blood glucose monitoring using spatiotemporal ECG and PPG feature fusion and weight-based Choquet integral multimodel approach, *IEEE. Trans. Neural. Netw. Learn. Syst* (2023), <https://doi.org/10.1109/TNNLS.2023.3279383> [published Online First: 20230608].
- [30] PUA-MOS: End-to-End Point-wise Uncertainty Weighted Aggregation for Moving Object Segmentation. 2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), IEEE, 2022.
- [31] P. Divilly, N. Zaremba, Z. Mahmoudi, et al., Hypo-METRICS: Hypoglycaemia—MEasurement, ThResholds and ImpaCtS—A multi-country clinical study to define the optimal threshold and duration of sensor-detected hypoglycaemia that impact the experience of hypoglycaemia, quality of life and health economic outcomes: The study protocol, *Diabet. Med.* 39 (9) (2022) e14892.
- [32] Z. Mahmoudi, S. Del Favero, P. Jacob, P. Choudhary, R.C. Hypo, Toward an optimal definition of hypoglycemia with continuous glucose monitoring, *Comput. Methods. Programs. Biomed* 209 (2021) 106303, <https://doi.org/10.1016/j.cmpb.2021.106303> [published Online First: 20210727].
- [33] M. Koenenman, M. Olde Bekkink, L.V. Meijel, S. Bredie, B. de Galan, Effect of hypoglycemia on heart rate variability in people with type 1 diabetes and impaired awareness of hypoglycemia, *J. Diabetes. Sci. Technol* 16 (5) (2022) 1144–2119, <https://doi.org/10.1177/19322968211007485> [published Online First: 20210415].
- [34] C.R. Andreasen, A. Andersen, P.G. Hagelqvist, et al., Sustained heart rate-corrected QT prolongation during recovery from hypoglycaemia in people with type 1 diabetes, independently of recovery to hyperglycaemia or euglycaemia, *Diabetes. Obes. Metab* 25 (6) (2023) 1566–1575, <https://doi.org/10.1111/dom.15005> [published Online First: 20230220].
- [35] B. Bent, P.J. Cho, M. Henriquez, et al., Engineering digital biomarkers of interstitial glucose from noninvasive smartwatches, *NPJ. Digit. Med* 4 (1) (2021) 89, <https://doi.org/10.1038/s41746-021-00465-w> [published Online First: 20210602].
- [36] M. Maritsch, S. Foll, V. Lehmann, et al., Smartwatches for non-invasive hypoglycaemia detection during cognitive and psychomotor stress, *Diabetes. Obes. Metab* 26 (3) (2024) 1133–2116, <https://doi.org/10.1111/dom.15402> [published Online First: 20231212].
- [37] W.R. Russell, A. Baka, I. Björck, et al., Impact of diet composition on blood glucose regulation, *Crit. Rev. Food. Sci. Nutr.* 56 (4) (2016) 541–590.
- [38] N.G. Boul'ee, C. Robert, G.J. Bell, et al., Metformin and exercise in type 2 diabetes: examining treatment modality interactions, *Diabetes. Care* 34 (7) (2011) 1469–1474.
- [39] C. Lloyd, J. Smith, K. Weinger, Stress and diabetes: a review of the links, *Diabetes. Spectrum* 18 (2) (2005) 121–217.
- [40] G.P. Kenny, R.J. Sigal, R. McGinn, Body temperature regulation in diabetes, *Temperature* 3 (1) (2016) 119–145.
- [41] A. Grandinetti, D.C. Chow, D.M. Sletten, et al., Impaired glucose tolerance is associated with postganglionic sudomotor impairment, *Clin. Auton. Res.* 17 (2007) 231–233.
- [42] P.G. Gandhi, H.R. Gundu, Detection of neuropathy using a sudomotor test in type 2 diabetes, *Degenerat. Neurol. Neuromuscular Dis.* (2015) 1–7.
- [43] A.Z. Woldaregay, E. Arsand, T. Botsis, D. Albers, L. Manykina, G. Hartvigsen, Data-driven blood glucose pattern classification and anomalies detection: machine-learning applications in type 1 diabetes, *J. Med. Internet. Res* 21 (5) (2019) e11030, <https://doi.org/10.2196/11030> [published Online First: 20190501].
- [44] B. Norgaard, G. Quer, B.K. Beaulieu-Jones, et al., Minimum information about clinical artificial intelligence modeling: the MI-CLAIM checklist, *Nat. Med.* 26 (9) (2020) 1320–2134, <https://doi.org/10.1038/s41591-020-1041-y>.
- [45] Evaluation of ambulatory ECG sensors for a clinical trial on outpatient cardiac rehabilitation. IEEE/ICME International Conference on Complex Medical Engineering, IEEE, 2010.
- [46] J.-H. Kim, R. Roberge, J. Powell, A. Shafer, W.J. Williams, Measurement accuracy of heart rate and respiratory rate during graded exercise and sustained exercise in the heat using the Zephyr BioHarness™, *Int. J. Sports. Med.* (2012) 497–501.
- [47] A particle filter framework for the estimation of heart rate from ECG signals corrupted by motion artifacts. 2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), IEEE, 2015.
- [48] P. Valensi, F. Extramiana, C. Lange, et al., Influence of blood glucose on heart rate and cardiac autonomic function, *DESIR Study. Diabet. Med.* 28 (4) (2011) 440–449, <https://doi.org/10.1111/j.1464-5491.2010.03222.x>.
- [49] D. Makowski, T. Pham, Z.J. Lau, et al., NeuroKit2: A Python toolbox for neurophysiological signal processing, *Behav. Res. Methods* 53 (4) (2021) 1689–1696, <https://doi.org/10.3758/s13428-020-01516-y> [published Online First: 20210202].
- [50] D.R. Roberts, V. Bahn, S. Ciuti, et al., Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure, *Ecography* 40 (8) (2017) 913–929.
- [51] E.I. Georga, V.C. Protopappas, D. Ardigo, D. Polyzos, D.I. Fotiadis, A glucose model based on support vector regression for the prediction of hypoglycemic events under free-living conditions, *Diabetes. Technol. Ther* 15 (8) (2013) 634–643, <https://doi.org/10.1089/dia.2012.0285> [published Online First: 20130713].
- [52] J. Jeon, P.J. Leimbiger, G. Baruah, M.H. Li, Y. Fossat, A.J. Whitehead, Predicting glycaemia in type 1 diabetes patients: experiments in feature engineering and data imputation, *J. Healthc. Inform. Res* 4 (1) (2020) 71–90, <https://doi.org/10.1007/s41666-019-00063-2> [published Online First: 20191210].
- [53] A. Adams, P. Vamplew, Encoding and decoding cyclic data, *S. Pac. J. Nat. Sci.* 16 (01) (1998).
- [54] M.H. Imani, E. Bompard, P. Colella, T. Huang, Forecasting Electricity price in different time horizons: an application to the Italian electricity market, *IEEE Trans. Ind. Appl.* 57 (6) (2021) 5726–5736, <https://doi.org/10.1109/Tia.2021.3114129>.
- [55] T. Danne, R. Nimri, T. Battelino, et al., International Consensus on Use of Continuous Glucose Monitoring, *Diabetes. Care* 40 (12) (2017) 1631–1640, <https://doi.org/10.2337/dc17-1600>.
- [56] P.G. Jacobs, P. Herrero, A. Facchinetti, et al., Artificial intelligence and machine learning for improving glycaemic control in diabetes: best practices, pitfalls and opportunities, *IEEE. Rev. Biomed. Eng.* (2023).
- [57] S.M. Lundberg, S.I. Lee, A Unified Approach to Interpreting Model Predictions. Advances in Neural Information Processing Systems 30 (Nips 2017) 2017:30.
- [58] Effects of hyperglycemia on variability of RR, QT and corrected QT intervals in Type 1 diabetic patients. 2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 2013. IEEE.
- [59] J.P. Singh, M.G. Larson, C.J. O'Donnell, et al., Association of hyperglycemia with reduced heart rate variability (The Framingham Heart Study), *Am. J. Cardiol* 86 (3) (2000) 309–312, [https://doi.org/10.1016/s0002-9149\(00\)00920-6](https://doi.org/10.1016/s0002-9149(00)00920-6).
- [60] F. D'Antoni, L. Petrosino, A. Marchetti, et al., Layered meta-learning algorithm for predicting adverse events in type 1 diabetes, *IEEE. Access* 11 (2023) 9074–9094.
- [61] The OhioT1DM dataset for blood glucose level prediction: Update 2020. CEUR workshop proceedings, NIH Public Access, 2020.
- [62] J. Huang, A.M. Yeung, D.C. Klonoff, D. Kerr, Regarding, “detection of hypoglycemia and hyperglycemia using noninvasive wearable sensors: ECG and accelerometry”, *J. Diabetes. Sci. Technol.* 17 (6) (2023) 1722.
- [63] T. Richardson, A. Rozkovec, P. Thomas, J. Ryder, C. Meckes, D. Kerr, Influence of caffeine on heart rate variability in patients with long-standing type 1 diabetes, *Diabetes. Care* 27 (5) (2004) 1127–1131.
- [64] J.M. Gonin, M.M. Kadrofske, S. Schmaltz, E.J. Bastyr III, A.I. Vinik, Corrected QT interval prolongation as diagnostic tool for assessment of cardiac autonomic neuropathy in diabetes mellitus, *Diabetes. Care* 13 (1) (1990) 68–71.
- [65] Deep learning framework for detection of hypoglycemic episodes in children with type 1 diabetes. 2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 2016. IEEE.
- [66] O. Elvebak, C. Tronstad, K.I. Birkeland, et al., A multiparameter model for non-invasive detection of hypoglycemia, *Physiol. Meas* 40 (8) (2019) 085004, <https://doi.org/10.1088/1361-6579/ab3676> [published Online First: 20190903].
- [67] R. Cordeiro, N. Karimian, Y. Park, Hyperglycemia Identification Using ECG in Deep Learning Era, *Sensors (basel)* 21 (18) (2021) 6263, <https://doi.org/10.3390/s21186263> [published Online First: 20210918].
- [68] J.-U. Park, Y. Kim, Y. Lee, E. Urtnasan, K.-J. Lee, A prediction algorithm for hypoglycemia based on support vector machine using glucose level and electrocardiogram, *J. Med. Syst.* 46 (10) (2022) 68.

- [69] K. Fellah Arbi, S. Soulimane, F. Saffih, M.A. Bechar, O. Azzoug, Blood glucose estimation based on ECG signal, *Phys. Eng. Sci. Med.* 46 (1) (2023) 255–264, <https://doi.org/10.1007/s13246-022-01214-3> [published Online First: 20230103].
- [70] J. Stephenson, P. Kempler, P.C. Perin, J. Fuller, Group EICS, Is autonomic neuropathy a risk factor for severe hypoglycaemia? The EURODIAB IDDM Complications Study, *Diabetologia* 39 (1996) 1372–2136.
- [71] V. Spallone, D. Ziegler, R. Freeman, et al., Cardiovascular autonomic neuropathy in diabetes: clinical impact, assessment, diagnosis, and management, *Diabetes. Metab. Res. Rev.* 27 (7) (2011) 639–653.