Data-Driven Insights from Wearable Sensors for Early Detection of Chronic Disease Events

Joshua Olaoye
Computer Science Department
San Diego State University

Introduce problem

- Glycemic variability(GV), a key indicator of prediabetic conditions, often goes unnoticed due to the limitations of intermittent glucose monitoring
    - Standard clinical tools (e.g., HbA1c) fail to capture short-term glucose fluctuations like GV.

Background

"**Prediabetes** is an intermediate state of hyperglycemia with glycemic parameters above normal but below the diabetes threshold… Prediabetes is a condition defined as having blood glucose levels above normal but below the defined threshold of diabetes. It is considered to be an at risk state, with high chances of developing diabetes. While, prediabetes is commonly an asymptomatic condition, there is always presence of prediabetes before the onset of diabetes. The elevation of blood sugar is a continuum and hence prediabetes can not be considered an entirely benign condition."[2]
Early intervention, during the prediabetes state, in patients can reduce the risk of progression to Type 2 diabetes by 58%, according to the Diabetes Prevention Program [3]. In addition to diabetes, prediabetes is also associated with an increased risk of macrovascular disease, chronic kidney disease, and other metabolic complications[2]
Glycemic variability(GV) refers to swings in blood glucose levels, over minutes, hours, or days. Unlike HbA1c, which captures long-term average glucose, GV reflects the dynamic nature of glucose regulation and the body's ability to respond to internal and external stressors.
High glycemic variability is often seen before sustained hyperglycemia, allowing it to be a key marker for detecting metabolic dysfunction before diabetes develops.

Problem Motivation

97.6 million adults in the U.S. (38% of the population) have prediabetes, and an estimated 90% of these adults are undiagnosed [1]. Prediabetes can lead to severe health consequences, including type 2 diabetes and cardiovascular disease[1], and early intervention can reduce the risk of progression to Type 2 diabetes by 58%. "Unlike HbA1c, which measures long-term average glucose, GV reflects dynamic fluctuations that impair endothelial function and trigger chronic inflammation through pathways like NF-κB activation"[4]. Glycemic Variability (GV), the dynamic measure of glucose fluctuations, is a proven early marker of metabolic dysfunction and cardiovascular risk.
The standard for GV measurement, continuous glucose monitoring (CGM), is expensive and inaccessible in many settings. On the other hand, wearable devices offer a scalable, non-invasive, and much more accessible alternative by tracking physiological markers like HRV, skin temperature, and EDA. These markers correlate strongly with GV, bringing about the idea of using these markers to provide real-time insights into glucose variability.

Approach

1. Data Collection
- Source: The BIG IDEAs Lab Glycemic Variability and Wearable Device dataset, which includes physiological data recorded from wearable devices for participants at risk of prediabetes, the dataset for this project utilizes the data of 9 participants.[1]
- Features:
    - Glucose: Continuous glucose readings.
    - Heart Rate (HR): Beats per minute.
    - Heart Rate Variability (HRV): Variability in inter-beat intervals.
    - Electrodermal Activity (EDA): Skin conductivity changes reflect autonomic nervous system activation.
    - Skin Temperature (TEMP): Peripheral body temperature fluctuations.
    - Tri-Axial Accelerometry (ACC): X, Y, and Z axes movement intensities.

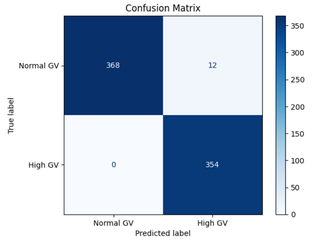
2. Preprocessing
- To ensure consistent formatting and alignment of all sensor data:
    - Converted timestamps into datetime format and aligned data points across modalities.
    - Resampled physiological signals to retain original timestamps for each participant, preserving granularity.
- Rolling Metrics to compare to timestamped data:
    - Calculated 1-hour rolling mean, standard deviation (SD), and coefficient of variation (CV) for glucose data.
    - Defined high GV events as CV > 33% at each timestamp.

3. Feature Engineering
- Extracted features relevant to GV detection:
    - HRV Metrics: Calculated the standard deviation of inter-beat intervals over rolling windows.
    - Activity Levels: Aggregated accelerometry data to summarize movement intensity over time.
- Contextual Features:
    - Skin temperature trends indicate metabolic stress.
    - EDA peaks reflect sympathetic nervous system activity.
    - Time of day to account for circadian effects on glucose metabolism.
- Combined all wearable metrics into a single dataset for machine learning.

4. Model Training
- Model: Logistic Regression
- Target Variable: Classified timestamps into:
    - High GV: Defined by CV > 33%.
    - Normal GV: Defined by CV ≤ 33%.
- Handling imbalanced and scarce data (few high GV cases compared to normal GV)
    - Applied Synthetic Minority Oversampling Technique (SMOTE) to balance the data
- Train-Test Split: Divided data into 80% training and 20% testing sets, ensuring a balanced representation of both classes.

Evaluation

Evaluated the model(both training(2934 values) and testing(734 values) sets) on the following metrics:
- Accuracy: 99%, indicating overall correctness of predictions.
- Precision: For High GV: 98%, showing minimal false positives.
- Recall: For High GV: 100%, demonstrating the model identified all true positives.
- F1-Score: Balanced measure of precision and recall, at 0.99.
- AUC (Area Under the Curve): 0.99, reflecting excellent discrimination between High GV and Normal GV.



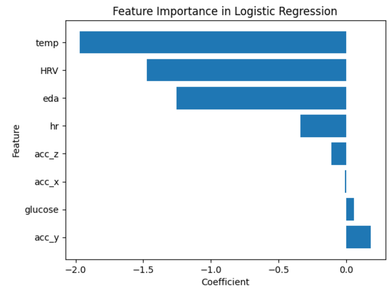
True Positives: 100% of High GV cases were correctly identified.
False Positives: Only 12 cases were misclassified as High GV.
False Negatives: Zero, no High GV case was missed.

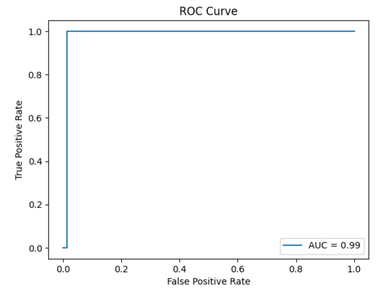
Results

The trained logistic regression model demonstrated strong predictive capabilities in detecting High Glycemic Variability events:
- **Accuracy**: 99% across both training and test datasets, confirming the model's overall reliability.
- **Recall (Sensitivity):** 100% for High GV cases, ensuring no true positive cases were missed.
- **Precision**: 98% for High GV cases, indicating minimal false positives.
- **F1-Score**: 0.99, showcasing a balanced performance between precision and recall.
- **AUC (Area Under the Curve):** 0.99, reflecting near-perfect separation between High GV and Normal GV classes.

Feature Importance

- **Skin Temperature**: The strongest predictor, with lower values correlating with High GV events.
- **HRV (Heart Rate Variability):** Significant negative predictor, reduced HRV indicates metabolic stress.
- **EDA (Electrodermal Activity)**: Peaks in EDA were linked to glycemic excursions, making it a key feature.
- **Glucose Metrics (CV)**: Confirmed as the foundation for defining High GV events, with higher CV thresholds indicating metabolic instability.
- **ACC(X, Y, Z):** not a strong predictor in comparison to the other features, but were important in explaining glucose dynamics(high/low activity periods)
    - high activity is linked to lower GV (physical movement increases insulin sensitivity and stabilizes blood glucose levels), while low activity is linked to higher GV





Conclusions

- This project highlights the feasibility and potential of wearable sensor data for early detection of high glycemic variability(GV) in prediabetic individuals.
- Using physiological markers such as heart rate variability (HRV), skin temperature, and electrodermal activity (EDA), combined with glucose variability metrics(CV), the proposed model achieved:
    - 99% accuracy and an AUC of 0.99, underscoring its reliability.
    - 100% recall for high GV events, ensuring no true positives were missed.
    - Effective integration of multimodal data, demonstrating that non-invasive wearables can bridge critical gaps in early GV detection.
- The model proposes a non-invasive, scalable alternative to CGMs enabling early detection and intervention

Shortcomings:
- While the model performed well, limitations include reliance on a small, controlled dataset and the need for real-world validation. Future research should focus on testing across diverse populations, integrating additional features like diet and stress, and deploying adaptive models for continuous monitoring.

Merging wearable technology with machine learning opens new pathways for innovation in personalized preventative healthcare.

References

[1] Cho, Peter, et al. "BIG IDEAs Lab Glycemic Variability and Wearable Device Data" (version 1.1.2). PhysioNet (2023), https://doi.org/10.13026/zthx-5212.
[2] Bansal, Nidhi. "Prediabetes diagnosis and treatment: A review." World journal of diabetesvol. 6,2 (2015): 296-303. doi:10.4239/wjd.v6.i2.296
[3] Diabetes Prevention Program Research Group. "10-year follow-up of diabetes incidence and weight loss in the Diabetes Prevention Program Outcomes Study." The Lancet374.9702 (2009): 1677-1686.
[4] Rodbard, David. "Glucose variability: A review of Clinical Applications and Research Developments." Diabetes Technology & Therapeutics, vol. 20, no. S2, June 2018, https://doi.org/10.1089/dia.2018.0092.