# Executive Summary: Predicting Glucose Levels using Smartwatch Data

Noah Gillespie, Margaret Swerdloff, Oladimeji Olaluwoye, S. C. Park, Daniel Visscher
GitHub: https://github.com/NoahGillespie/CGM

## 1. Introduction

High blood glucose levels can lead to serious health complications. The ability to predict these events, particularly for individuals with prediabetes, could be a valuable tool for preventative healthcare. Continuous glucose monitoring (CGM) is a rapidly developing technology that allows for real-time interstitial blood sugar measurements, offering advantages over traditional methods. Our primary goal is to develop a model that can predict potential high glucose events (e.g. interstitial glucose level exceeding the 95th percentile), using CGM, smartwatch, and food intake data.

## 2. Dataset and Processing

We leveraged a large dataset (Cho et al., 2023; Bent et al., 2021) containing CGM data alongside various smartwatch sensor readings, including accelerometer and heart rate, and food logs. These were collected from 16 prediabetic individuals over 8-10 days. A high-pass filter was used to extract activity level from accelerometer data. Large language models (LLMs) were used to match descriptions to web-scraped glycemic index (GI) data (from https://glycemicindex.com/) and classify unmapped foods.

## 3. Models

- **Baseline Model (Logistic Regression)**: Predicts the occurrence of high blood glucose events based on carbohydrate intake.
- **Improved Logistic Regression Model**: Based on carbohydrate intake and GI.
- **ADL Model (Autoregressive Distributed Lag Model):** Predict future glucose levels as a time series, trained on past glucose, food, and smartwatch data.
    - **Food Only:** Predicts glucose level as a time series using food data (nutrition and GI data) to understand food choices' impact.
    - **Smartwatch Only:** Predicts glucose level as a time series using smartwatch data (movement data, skin temperature, electrodermal activity, and heart rate).
    - **Combined:** Combines food and smartwatch data for more accurate predictions.

## 4. Results

The baseline model with carbohydrates showed sensitivity and specificity (rate of correctly predicting resp. high and non-high glucose events) of 33% and 84%. Adding GI doubled the sensitivity to 67%, while showing a minor dip in specificity to 72%.

Time series modeling with ADL performed well for some models, predicting peaks and glucose trends. The best model for participants with low variance data used both the smartwatch and food or GI data. When participant data was noisy, the best models used only the carb or GI data. The hyperparameters of the ADL model were tuned for each participant. The top 5 RMSEs are:

1. Participant 5: 10.62 mg/dL
2. Participant 1: 14.19 mg/dL

3. Participant 12: 16.34 mg/dL
4. Participant 8: 16.54 mg/dL
5. Participant 16: 16.82 mg/dL

**5. Future Directions**

A future direction is to incorporate additional sensor data (e.g., blood volume pulse) to further enhance model performance. Additionally, it would be interesting to investigate the feasibility of predicting the time to reach/recover from the peak glucose level after a meal. Finally, a mobile application could be developed that integrates these models to provide real-time glucose level predictions and personalized dietary recommendations for users.

**References:**


Bent, B., Cho, P.J., Henriquez, M. et al. Engineering digital biomarkers of interstitial glucose from noninvasive smartwatches. npj Digit. Med. 4, 89 (2021). https://doi.org/10.1038/s41746-021-00465-w

Brand-Miller, J. (2024, May 1). *GI Group*. Glycemic Index – Glycemic Index Research and GI News. https://glycemicindex.com/

Cho, P., Kim, J., Bent, B., & Dunn, J. (2023). BIG IDEAs Lab Glycemic Variability and Wearable Device Data (version 1.1.2). *PhysioNet*. https://doi.org/10.13026/zthx-5212.

Goldberger, A., Amaral, L., Glass, L., Hausdorff, J., Ivanov, P. C., Mark, R., ... & Stanley, H. E. (2000). PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals. Circulation [Online]. 101 (23), pp. E215–e220.