Evaluating the Zero-Shot Predictive Ability of Large Language Models for Continuous Glucose Monitoring Data

Junyan Tan (Samuel)[1], Irina Gaynanova[1]

University of Michigan School of Public Health, Department of Biostatistics

Introduction

- **Motivation**: Increasing clinical usage of Continuous Glucose Monitoring (CGM) demands more accurate glucose forecasting.
- **Limitation of Existing Methods**: Traditional models require extensive training and overlook critical patient demographics (e.g., diabetes type, lifestyle), limiting clinical usefulness.
- **Proposed Solution**: We introduce a zero-shot Large Language Model (LLM) approach, converting CGM readings and demographic data into text prompts for direct glucose-level forecasting without additional training.

Data

Following the CGM forecasting benchmark established previously[1], we pre-processed five public CGM datasets with summarized information below.

- Interpolation & Segmentation
  - Identify gaps in each subject's CGM measurements.
  - Impute missing values with linear interpolation if the gap is below a threshold.
  - Break the time series into separate segments when the gap exceeds the threshold.
- Data Splitting
  - Split each dataset into training, validation, and in-distribution (ID) test sets using 90% of subjects.
  - For each subject, partition data chronologically, with validation and ID test sets each spanning 16 hours.
  - Use the remaining 10% of subjects to form an out-of-distribution (OD) test set.

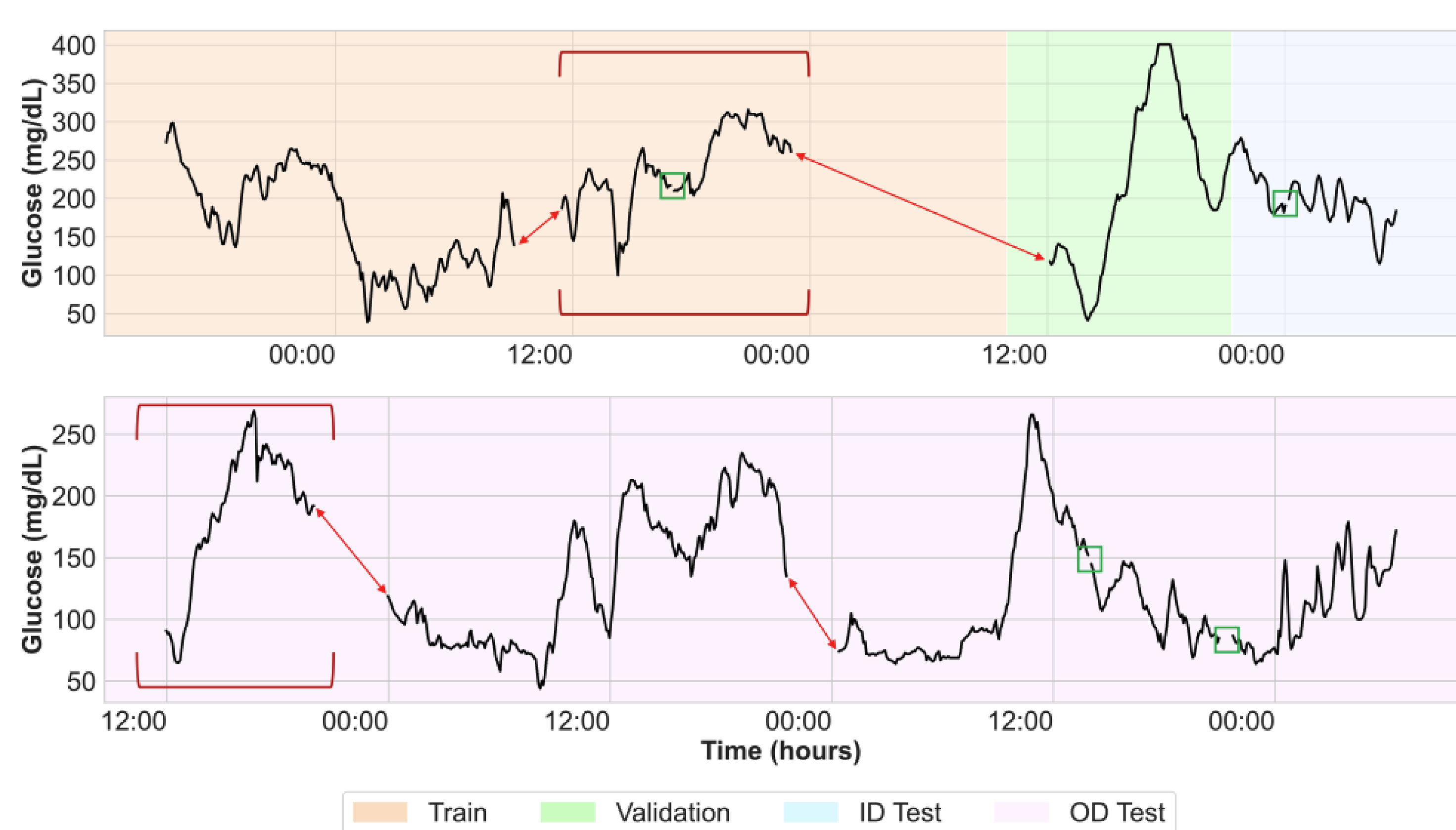

Figure 1. Example data processing on Weinstock et al.[6]

Dataset Information

Table 1. Demographic information (average) for each dataset (Raw data only).

Dataset	Diabetes	# of Subjects	Age	Sex (M/F)
Broll2021[2]	Type 2	5	NA	NA
Colás2019[3]	Type 1	208	59	103/104
Dubosson2018[4]	Mixed	9	NA	6/3
Hall2018[5]	Type 1	57	48	25/32
Weinstock2016[6]	Type 1	200	68	106/94

Methodology

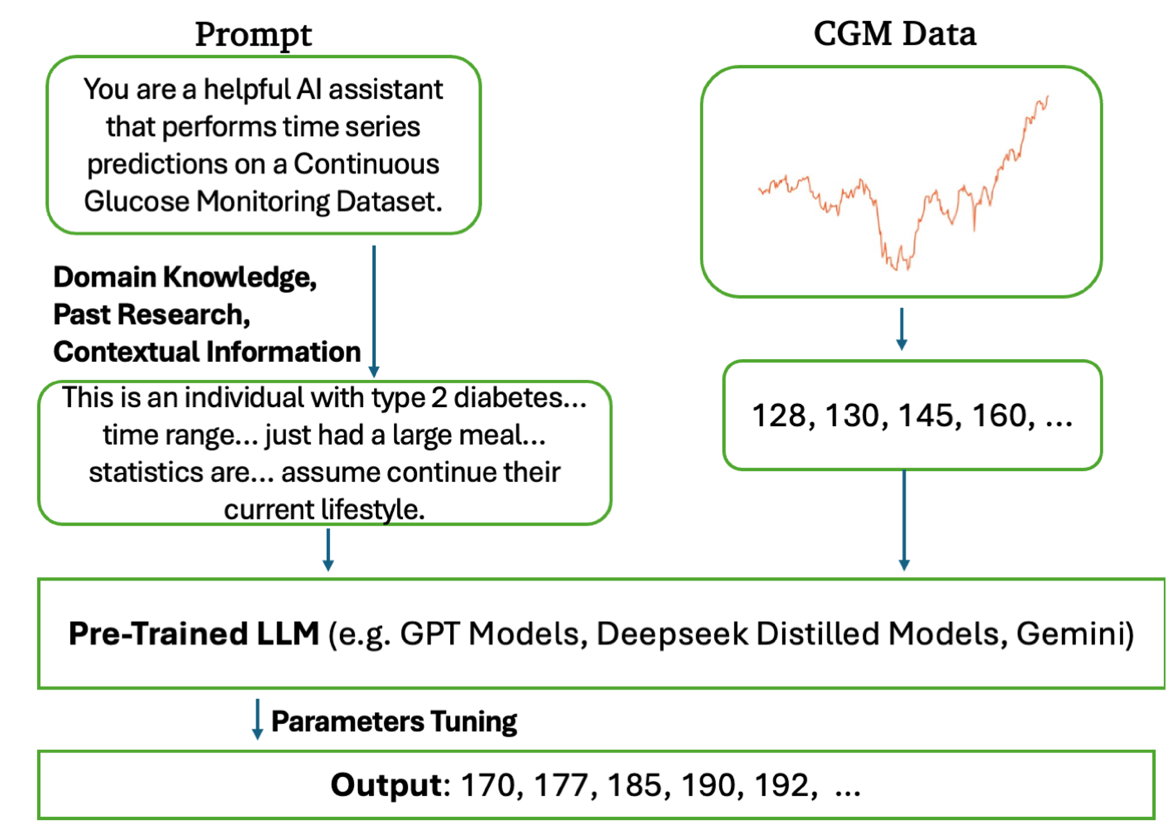

Figure 2. Baseline Workflow

Baseline Approach: We use a structured prompt to guide the LLM:

- **Context:** Patient background (type, demographics, insulin therapy).
- **Constraints:** Enforce valid CGM ranges (40–400 mg/dL).
- **Scenario:** Highlight events (big meal, exercise).
- **Statistics:** Provide key metrics (min, max, mean, std).
- **Output:** Instruct LLM to return only numerical forecasts.

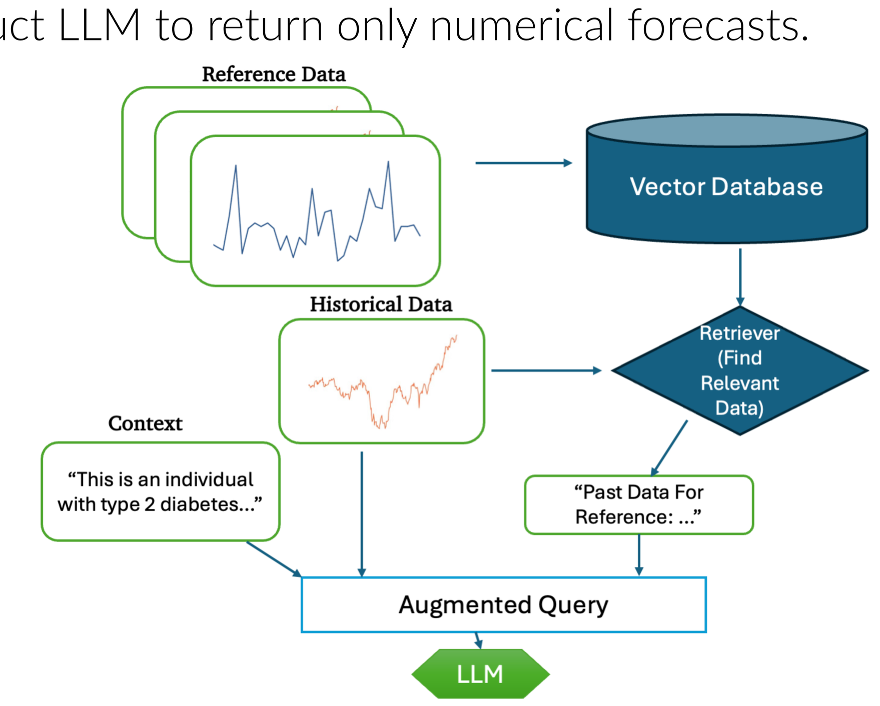

Figure 3. Overview of RAG Process

Retrieval Augmented Generation (RAG): A novel approach to improve prediction accuracy by supplementing the prompt with additional information:

- Store historical CGM data in a vector database.
- Retrieve similar data based on embeddings.
- Append relevant info to the prompt for improved accuracy.

Example from the Broll Dataset

You are looking at CGM data of a type 2 diabetes patient. Each reading has a 5-minute interval, and the first reading is recorded at <timestamp>. For the given 24-hour interval, the glucose statistics are: . . .
<RAG input>: Relevant historical data from the same or similar profiles: . . .
<RAG input end>
Glucose levels range from 40–400 mg/dL. Assume the patient continues with their current lifestyle.
Return only numerical forecasts; no additional commentary.

Results

- **Comparable Performance**: Preliminary results indicate that LLMs perform at least on par with traditional methods.
- **Trend Awareness**: LLMs effectively capture the trend and rate of change in CGM data.
- **Periodicity Recognition**: They recognize the inherent periodic nature of CGM data, even in out-of-distribution forecasts.
- **Prompt Sensitivity**: LLMs respond significantly to prompt modifications, demonstrating awareness of the forecasting context.

Context Recognition

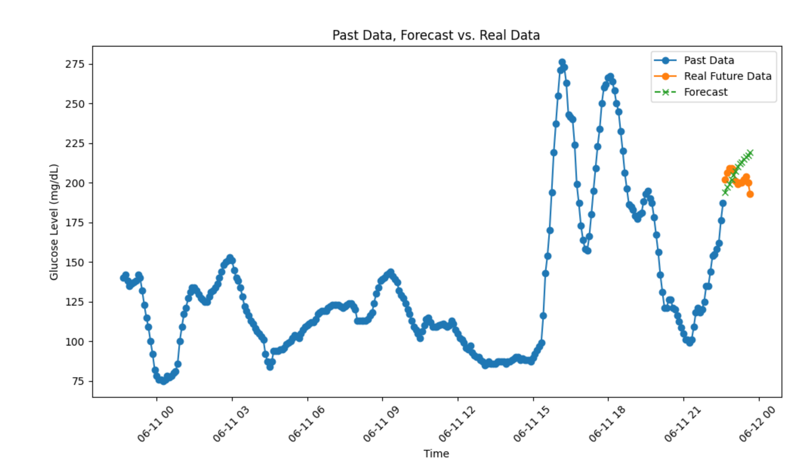

Figure 4. Given Context: "Patient is type 2, on insulin therapy"

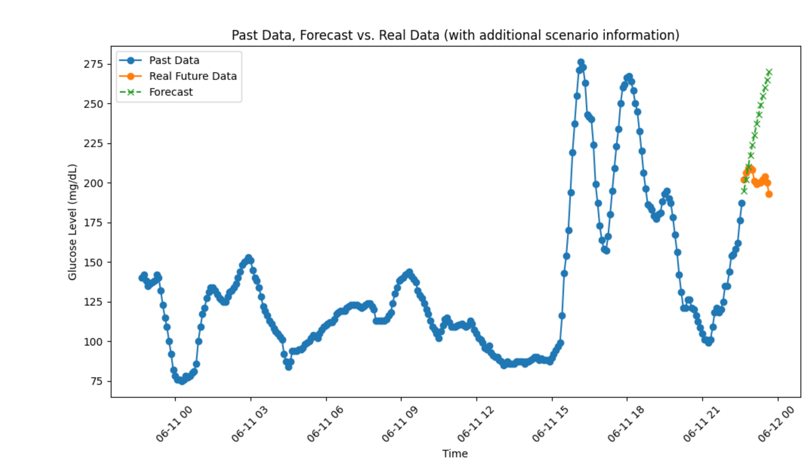

Figure 5. Given Context: "Patient is type 2, not on insulin therapy"

Scenario Adaptation

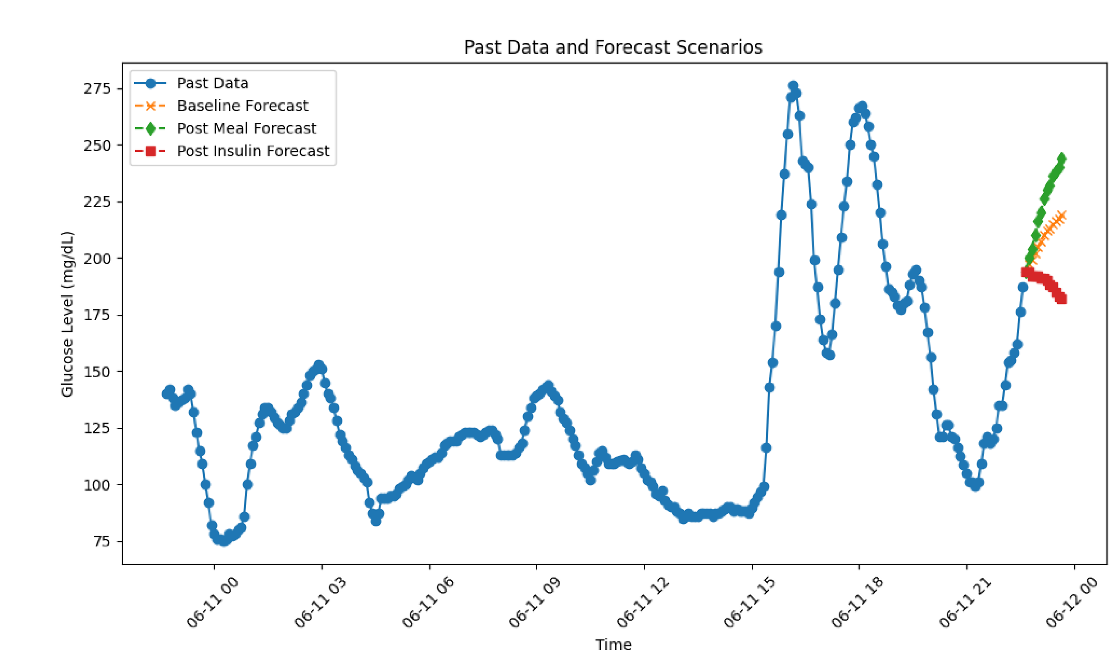

Figure 6. Scenario Prediction: The LLM adapts forecasts based on different real-life scenarios

Advantages Compared to Traditional Methods

- **Context-Aware**: While not necessarily the most accurate, LLMs remain valuable for incorporating contextual information and adapting predictions on the fly.
- **Zero-Shot Ready**: Requires little to no parameter tuning, enabling immediate use and lowering computational barriers.
- **Cost-Effective**: With advancements in distilled models (e.g., DeepSeek distilled Llama) and growing competition, LLM-based forecasting is becoming increasingly affordable.

[1] Sergazinov, R., et al. (2023). *GlucoBench: Curated List of Continuous Glucose Monitoring Datasets with Prediction Benchmarks*. *arXiv*.
[2] Broll, S., et al. (2021). Interpreting blood glucose data with R package iglu. *PLoS One, 16*(4), e0248560.
[3] Colás, A., et al. (2019). Detrended fluctuation analysis in the prediction of type 2 diabetes mellitus in patients at risk: Model optimization and comparison with other metrics. *PLoS One, 14*(12), e0225817.
[4] Dubosson, F., et al. (2018). The Open D1NAMO dataset: A multi-modal dataset for research on non-invasive type 1 diabetes management. *Informatics in Medicine Unlocked, 13*, 92–100.
[5] Hall, H., et al. (2018). Glucotypes reveal new patterns of glucose dysregulation. *PLoS Biology, 16*(7), e2005143.
[6] Weinstock, R. S., et al. (2016). Risk factors associated with severe hypoglycemia in older adults with type 1 diabetes. *Diabetes Care, 39*(4), 603–610.

 tansam@umich.edu