CGMacros: Dataset For Personalized Nutrition and Diet Monitoring

Team-2, ***Data Explorers*** - Ganga, Monalisa, Rohitha, Shravani, Shruti

Objective: Uncover meaningful relationships between dietary habits, glucose responses, physical activity, and gut health.
Tools Used: Python (Jupyter Notebook), Tableau, Excel

Significance Of Dataset:

The CGMacros dataset contains multimodal information from two continuous glucose monitors (CGM), food macronutrients, food photographs, and physical activity, in addition to anonymized participant demographics, anthropometric measurements and health parameters from blood analyses and gut microbiome profiles.

Dataset Background:

The CGMacros dataset, publicly available on [PhysioNet], was designed to support research in personalized nutrition and automated diet monitoring. It includes rich, multimodal data collected from 45 adults over a 10-day period, integrating continuous glucose monitoring (CGM), dietary intake (with macronutrient breakdown), physical activity metrics, heart rate, and gut microbiome profiles. Unlike many existing datasets limited to single-sensor inputs or self-reported logs, CGMacros uniquely captures the dynamic relationship between food composition, glucose response, and physiological signals, enabling deeper analysis of how lifestyle and dietary choices influence metabolic health.

Approach for data analysis:

To uncover meaningful relationships between glucose responses, dietary intake, physical activity, and gut health, we followed a structured multi-step analysis approach:

1. **Demographic Profiling:** We began by analyzing participant demographics such as age, gender, BMI classification, and ethnicity to understand the composition of the study group. This helped identify potential baseline factors influencing health markers and glucose regulation.

2. **Impact of Dietary Habits:** Meals were analyzed by type (breakfast, lunch, dinner, snack) and macronutrient composition (carbs, protein, fat, fiber). We examined how these components correlate with postprandial glucose responses at 60 and 120 minutes, revealing trends such as breakfast-induced spikes and snack-related stability.

3. **Impact of Physical Activity:** To assess the impact of physical activity on post-meal glucose responses, we calculated the average activity calories burned per minute in the 60-minute window following each meal. Based on this metric, meals were categorized into three activity levels: Low, Moderate, and High. This allowed us to examine how varying intensities of physical exertion influence glycemic outcomes across participants.

4. **Impact of Gut Health:** We integrated gut microbiome features such as metabolic fitness, inflammatory activity, and fermentation pathways. Sankey charts and categorical comparisons helped trace how gut profiles relate to glucose outcomes and overall metabolic health.

5. **Cross-Domain Correlation:** Throughout the analysis, we layered variables across domains (e.g., BMI->HbA1c->Gut Health) to explore how combined factors contribute to metabolic outcomes. This multi-dimensional approach enabled a holistic understanding of personalized nutrition and lifestyle impact.

Key Findings and Insights:

1) Demographics:

The dataset includes 45 participants data, among which most participants in this dataset identify as Hispanic/Latino (75%), followed by White (15%), African American (7%), and Black/African American (2%).
BMI Analysis - Over half of the participants fall in the *Obese* category (51%), 18% in the *Normal* weight category, while 31% in the *Overweight* category.
Age distribution: The age distribution of participants is skewed toward older adults, with the majority in the 50s (37.78%) and 40s (22.22%) age groups. Participants in their 60s account for 17.78%, while younger age groups, 30s (11.11%), 20s

(8.89%), and 10s (2.22%), are less represented in the cohort.

Gender distribution - The participant cohort consisted of 64% females and 36% males, indicating a higher representation of female participants in the study.

Glucose Classification Distribution Across LDL Risk Groups: Glucose dysregulation (prediabetes and diabetes) is observed across all LDL risk categories, including those with optimal LDL levels. This highlights that good cholesterol alone does not ensure healthy glucose control. In contrast, normal glucose levels are more common in the optimal and near-optimal LDL groups, suggesting a partial overlap between healthy lipid and glucose profiles. Regular monitoring of both markers is essential for comprehensive metabolic risk assessment.

BMI and cholesterol trends vary by age and gender: Females show higher BMI than males in most age groups, especially in their 30s and 40s. Cholesterol levels peak for both genders in their 50s, with females showing slightly higher values overall. These patterns highlight the need for age- and gender-specific health interventions targeting weight and lipid management.

How age and BMI affect Diabetes Risk?

As people age, especially into their 50s and 60s, obesity risk increases due to metabolic slowdown, which in turn raises the likelihood of diabetes. However, normal BMI does not guarantee healthy glucose levels; individuals with healthy weight can still develop prediabetes or diabetes, influenced by factors like genetics, diet, and gut health. Even in their 20s, some adults already show early signs of excess weight and poor glucose regulation, emphasizing the need for early lifestyle intervention. Age and BMI are key risk factors, but effective prevention must also address broader contributors to metabolic health.

---

2) Impact Of Dietary Habits:

Pre- vs. Post-Meal Glucose Levels per Participant with Diabetes Status - Dexcom:

Post-meal glucose levels generally show a significant increase compared to pre-meal levels, with peaks exceeding 200 mg/dL for diabetic participants 35 & 39 .

This is expected, as diabetics often have impaired glucose regulation leading to higher postprandial spikes

Variability is high, with some normal (N) and prediabetic (P) participants also showing elevated post-meal levels, indicating potential early signs of glucose dysregulation.

Pre- vs Post-Meal Glucose Levels per Participant with Diabetes Status - Libre:

Post-meal glucose levels again rise notably, with peaks around 200-225 mg/dL for participant 35D. The pattern aligns with the Dexcom data, but some differences in magnitude of peaks suggest monitor-specific Variations. Diabetic participants consistently show higher post-meal levels, while normal (N) participants generally maintain lower levels, though some prediabetic (P) cases show intermediate responses.

Carbohydrate Intake and Glucose Response Vary by Meal Type:

Analysis of the relationship between carbohydrate intake and glucose levels, 60 minutes post-meal (measured via Dexcom CGM) reveals notable differences across meal types:

Breakfast meals are associated with the steepest glucose spikes. Carbohydrates consumed at breakfast tend to result in a sharper rise in glucose levels, potentially due to the faster digestion of liquid-based meals like protein shakes.

Dinner shows moderate glucose responses but with significant variability. While the average glucose rise is less steep, the wide range of responses, particularly the presence of some high spikes, may reflect variability in dinner composition and individual dietary choices.

Lunch produces a more consistent and moderate glycemic response. Since lunch meals were standardized (e.g., from Chipotle), the responses were more predictable, with a moderate slope indicating a balanced impact of carbs on glucose.

Snacks trigger the smallest glucose increases. The flatter response curve suggests that snacks, often smaller in portion or lower in glycemic index, are associated with milder postprandial glucose changes.

In **conclusion**, carbohydrate intake at breakfast leads to sharper glucose spikes than other meal types, likely due to faster-digesting food (protein shakes). Lunches (standardized from Chipotle) produce consistent responses, while dinners show high variability. Snacks result in the flattest glucose response, likely due to smaller portion size or food type.

Carbs vs Dexcom GL (60-min post-meal) by HbA1c Classification:

Post-meal glucose levels (60 minutes after eating) vary notably by HbA1c classification. Participants with diabetes exhibit consistently higher and more variable glucose spikes across all carbohydrate intake levels. Those with normal HbA1c have the lowest glucose responses, while prediabetic individuals fall in between. These findings highlight the heightened glycemic sensitivity in diabetic individuals and underscore the need for tailored dietary strategies and carbohydrate moderation for effective glucose management.

Macronutrient Balance Influences Post-Meal Glucose Response:

Protein-to-Carb Ratio:

A clear negative correlation is observed between the protein-to-carb ratio and glucose rise at 60 minutes post meal. Meals with a higher proportion of protein relative to carbohydrates are associated with smaller glucose spikes. This suggests that protein may play a moderating role in glycemic response, highlighting the potential benefits of macronutrient balancing in blood sugar regulation and metabolic health.

Fat-to-Carb Ratio:

Similarly, an inverse relationship exists between fat-to-carb ratio and glucose elevation. As the fat content relative to carbohydrates increases, the postprandial glucose rise tends to decrease. This indicates that dietary fat may help slow carbohydrate absorption, leading to a more gradual glucose response, an important consideration for individuals with impaired glucose tolerance.

Fiber-to-Carb Ratio:

Higher fiber-to-carb ratios are modestly linked to reduced glucose spikes. Although the relationship is less pronounced, meals with more fiber tend to produce lower post-meal glucose elevations. The generally low fiber content across meals suggests a nutritional opportunity to improve glycemic outcomes by incorporating more fiber-rich foods.

In **conclusion**, this analysis reinforces the value of macronutrient composition, particularly protein and fat content, in shaping glycemic responses and informs strategies for personalized dietary interventions.

---

3) Impact Of Physical Activity:

Impact of Physical Activity on Post-Meal Glucose Response (0–60 Minutes):

This analysis examines how varying levels of physical activity after meals influence changes in glucose levels over the following hour, as measured by continuous glucose monitoring (Dexcom):

1. Higher Activity Is Associated with Lower Glucose Spikes

Individuals in the high physical activity group exhibit the lowest median glucose increase, indicating that engaging in more activity after eating can help blunt the postprandial glucose response.

2. Greater Variability in Low and Moderate Activity Groups

Participants with low to moderate activity levels display wider spreads in glucose response, with several extreme cases showing spikes exceeding 200 mg/dL. This suggests inconsistent glycemic control when post-meal activity is limited.

3. Glucose Declines Observed Across All Groups

Some individuals experienced negative glucose deltas, where glucose levels decreased after eating. This could result from small meal size, high insulin sensitivity, or immediate light activity after meals.

In **conclusion,** physical activity following meals plays a critical role in modulating glucose spikes. Higher levels of activity are associated with more favorable and stable glucose outcomes, reinforcing the value of movement as part of post-meal routine, particularly for individuals aiming to manage blood sugar levels.

Impact of Physical Activity on Glucose Spikes by HbA1c Classification:

This analysis explores how post-meal physical activity influences glucose spikes across different HbA1c groups (Normal, Prediabetes, Diabetes):

- Higher Activity Reduces Glucose Spikes: Across all HbA1c classifications, increased physical activity consistently lowers median glucose rises, with the effect most pronounced in the diabetes group.
- Normal HbA1c Group Shows Most Stability: Individuals with normal HbA1c exhibit the lowest and most consistent glucose responses, even with low activity levels.
- Greater Variability in Diabetes Group: Participants with diabetes show larger fluctuations and more frequent high spikes, especially at low or moderate activity levels.
- Prediabetic Responses Fall in Between: Glucose responses in the prediabetes group are intermediate, showing improvement with higher activity but with less stability than the normal group.

In **conclusion**, post-meal physical activity helps reduce glucose spikes in all groups, with the greatest benefit observed in individuals with diabetes or prediabetes. Promoting even moderate activity after meals may support better glycemic control.

Impact of Physical Activity on Glucose Spikes by BMI Classification:

This analysis examines how post-meal physical activity affects 60-minute glucose spikes across BMI groups (Normal, Overweight, Obese):

- Higher BMI implies Greater Glucose Spikes: Obese participants consistently show higher and more variable glucose rises, especially at low activity levels. Overweight individuals fall in the middle, while those with normal BMI exhibit the lowest and most stable responses.
- Physical Activity Lowers Spikes Across All Groups: Increased activity leads to reduced median glucose spikes in all BMI categories, with fewer high outliers, highlighting the positive effect of movement.

• Exercise Mitigates BMI-Related Risk: Even in obese individuals, high activity levels help blunt glucose spikes, though their responses remain elevated compared to normal BMI participants.
In **conclusion**, post-meal physical activity improves glucose control for everyone, but is especially important for individuals with higher BMI.

Relationship Between METs and Active Calories Burned per Minute:

There is a strong positive correlation between METs, and calories burned per minute, confirming that higher-intensity physical activities lead to greater energy expenditure. As METs increase, so does calories burn, with most participants following this consistent upward trend.

Higher physical activity is linked to lower glucose levels, especially in diabetics:

Both Libre and Dexcom data show a downward trend in glucose with increasing METs. The decline is steeper for individuals with diabetes, indicating that physical activity plays a crucial role in glucose control, particularly for those with elevated HbA1c.

---

4) Impact Of Gut Health:

Link Between Metabolic Fitness, Inflammation, and Gut Microbiome Health:

Our analysis highlights a clear progression: individuals with poor metabolic fitness tend to exhibit higher levels of inflammation, which in turn is associated with reduced gut microbiome health. This suggests that addressing metabolic and inflammatory issues may be key to improving gut health, reinforcing the importance of upstream interventions in personalized health strategies.

Gender Differences in Microbial Balance and HbA1c Levels:

This analysis explores the relationship between the ratio of beneficial to harmful gut microbes and HbA1c levels, a marker of long-term blood glucose control. The trend varies by gender, men generally show a slight decrease in HbA1c as microbial balance improves, while women display a flatter or slightly increasing trend. These patterns suggest a potential gender-specific link between gut microbiome composition and glucose regulation. However, due to the wide variability and overlapping confidence intervals, further investigation using multivariate analysis (e.g., adjusting for age or BMI) is warranted.

Average Gut Function Profile:

The radar chart showed a well-balanced gut function across four key dimensions, forming a nearly symmetrical shape. This indicates an overall stable and moderate gut health profile. Slightly higher scores in Digestive Efficiency and Microbial Diversity suggest a generally effective microbiome, supporting good nutrient absorption and immune function.

Impact of Gut Microbe Types on Fasting Glucose & Insulin Levels:

Higher levels of good gut microbes are linked to lower fasting glucose and insulin, while elevated bad microbes correlate with poorer metabolic markers. This suggests that gut microbiome composition plays a key role in glycemic control.

Microbial Balance by Participant: Dominance of Good vs Moderate and Bad Microbes:

Most participants exhibit a high proportion of moderate microbes, with smaller but variable levels of good and bad microbes. A few individuals show elevated bad microbe counts, highlighting the need for closer microbiome monitoring in those cases.

Relationship Between Gut Microbial Load and Glucose Levels (Libre vs Dexcom):

Higher total microbial counts are modestly associated with lower average glucose levels across both CGM devices, indicating a potential protective role of gut microbiota in glycemic control.

Comparison of 2-Hour Postprandial Glucose Changes Across CGM Devices by Participant

Despite differences in placement (upper arm for Libre, abdomen for Dexcom) and sampling frequency (15-min vs. 5-min), both devices show similar trends in glucose response across participants. However, occasional discrepancies in the magnitude of glucose change suggest that each sensor may respond differently based on physiological factors, sensor calibration, or placement sensitivity. Comparing the readings provides a deeper understanding of postprandial glucose dynamics and helps validate consistency across monitoring technologies, which is especially important in clinical research and diabetes management.

---

5) Machine Learning:

We implemented machine learning to predict 2-hour post prandial glucose responses using Light Gradient Boosting Machine (LightGBM) algorithms with R^2 value of 0.78 and RMSE of 16.99 on the test data. We used Feature Importance to identify the variables influencing the predictions, and Hyperparameter Tuning to improve the performances of the model.

We have also used Linear Regression, k-Nearest Neighbors, Random Forest and XGBoost models to compare the performances and LightGBM performed well with balanced high R^2, lowest test error, and excellent generalization.
The study leverages the use of technology to gain a deeper understanding of how our bodies respond to food and use that knowledge to make informed decisions promoting better health and well-being. Early identification allows for targeted interventions to prevent or delay the onset of the disease.

Challenges Faced During the Analysis:

1. Missing or Incomplete Data: Several participants had missing values in key variables like insulin, HbA1c, or microbial counts, which limited the sample size for certain analyses.
2. Data Heterogeneity Across Devices:
Glucose data was collected from both Libre and Dexcom devices, which have slight differences in calibration and reporting frequency, complicating direct comparisons.
3. Limited Sample Size in Subgroups:
Some demographics or health-status subgroups had too few participants for robust stratified analysis.
4. This dataset provides high-frequency CGM (Continuous Glucose Monitoring) readings, it creates a dense time series. This granularity is excellent for machine learning and modeling, but it poses challenges such as:
   - Overplotting in scatter plots and line charts
   - Difficulty in identifying trends or patterns visually without aggregation
   - Increased processing time and memory usage
5. Categorizing the original list of 1979 microbes into Good, Moderate or Bad based on research literature posed a challenge due to the volume and complexity of microbial data.
6. Machine Learning:
   - Participant Variability: Wide variability in participant behavior, meal composition, and physiology added noise to prediction.
   - Feature Selection Complexity: Choosing the right set of input variables was critical. Irrelevant or redundant features could mislead the model.

Conclusions:

1. Metabolic Health Is Multifactorial
No single factor, be it BMI, LDL, or age, fully determines glucose control. Even individuals with normal weight or optimal cholesterol showed signs of dysglycemia, emphasizing the need for a holistic view of metabolic risk.

2. Diet Composition Strongly Influences Glucose Response
Meals with higher protein, fat, or fiber relative to carbs lead to smaller glucose spikes. Breakfast meals, especially shakes, caused the highest glycemic responses, while standardized lunches and snacks showed more stable outcomes. This supports personalized macronutrient strategies for better glycemic control.

3. Post-Meal Physical Activity Blunts Glucose Spikes
Higher activity after meals consistently reduced glucose spikes across age, BMI, and HbA1c groups. The impact was most pronounced in individuals with diabetes or obesity. Moderate activity can be a powerful, non-invasive tool for glycemic regulation.

4. Gut Microbiome Plays a Vital Role in Glycemic Regulation
Individuals with better microbial balance (more beneficial microbes) showed lower fasting glucose and insulin levels. Poor metabolic fitness was associated with inflammation and poorer gut health, underscoring the upstream influence of gut function on overall metabolic outcomes.

5. Demographics Inform Risk Patterns but Don't Dictate Outcomes
Obesity was most common among older adults, particularly women, with many showing signs of elevated glucose. However, dysglycemia was observed even in younger and normal-weight participants, stressing the need for early screening and proactive lifestyle strategies.

6. Machine Learning Implementing
The machine learning approach effectively identified key predictors of post-meal glucose response using multiple regression models. Among them, LightGBM delivered the best performance with strong generalization and accuracy. Despite limited improvement from hyperparameter tuning, the models highlighted important relationships between pre-meal glucose levels, dietary intake, and physiological metrics. This reinforces the value of data-driven modeling in understanding individual glucose dynamics and opens pathways for personalized nutrition and metabolic health management.