Counterfactual Explainable AI for Diabetes Management

Hassan Ghasemzadeh, Asiful Arefeen

Embedded Machine Intelligence Lab (EMIL), College of Health Solutions, Arizona State University

ghasemzadeh.com

Paper Link

INTRODUCTION

Maintaining normal blood glucose levels through lifestyle behaviors is central to good health and disease prevention. Frequent exposure to dysglycemia (i.e., abnormal glucose events like hyperlycemia and hypoglycemia) leads to chronic complications. Therefore, a tool capable of predicting dysglycemia and offering users actionable feedback about how to make changes in their diet, exercise, and medication to prevent abnormal glycemic events could have significant societal impacts.

We propose ***GlyCoach***, a novel framework for generating counterfactual explanations that can provide insights into why a model made a particular prediction by generating hypothetical instances that are similar to the original input but lead to a different prediction outcome.

GlyCoach can preserve user preferences for keeping some features unchanged and making the technology more user friendly and realistic.

With **87% sensitivity** in the simulation-aided validation, GlyCoach surpasses the state-of-the-art techniques for generating counterfactual explanations by at least 10%. Besides, counterfactuals from GlyCoach exhibit a **32% improved proximity** compared to previous research.

AIM

Our aim is to establish a novel method to generate counterfactuals to help diabetes patients avoid hyperglycemic events beforehand.

Properties of the counterfactual suggestions

The generated counterfactuals (X_T^*) must have the following properties:

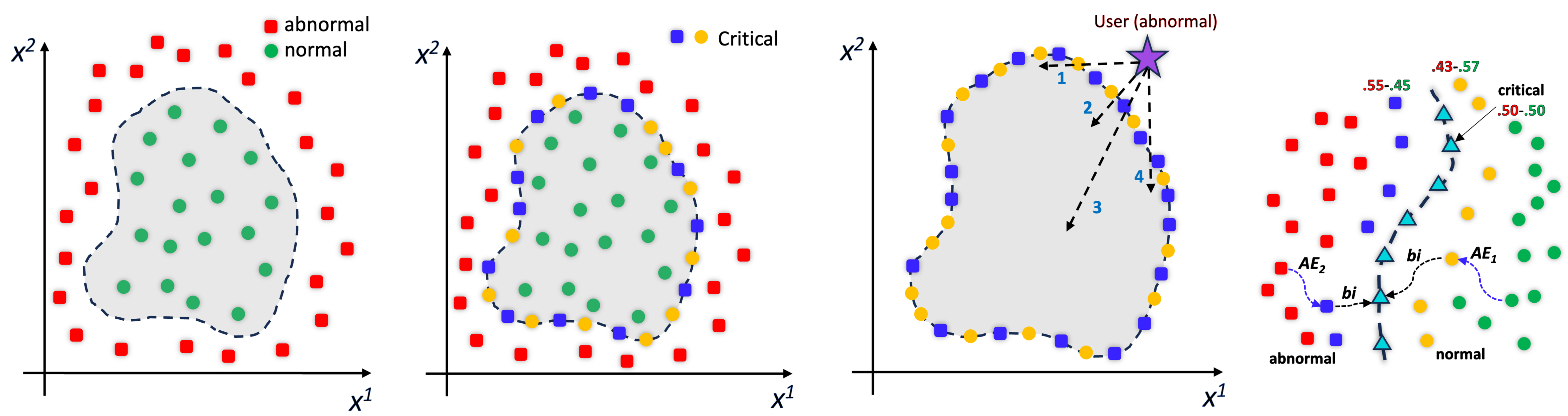
- ***Interventional:*** X_T^* must change the class of the initial prediction $f(X_T)$ from an abnormal class to the normal class
- ***Minimal:*** The counterfactual suggestion X_T^* must be minimally distant from the original hyperglycemic sample
- ***Realistic:*** The counterfactuals X_T^* must follow the original data manifold and must not make unrealistic suggestions
- ***Partial:*** X_T^* must reflect user preferences to keep some features unchanged

$$\min_{X_T^*} \Big[CE\big(f(X_T^*), \vec{n}\big) + d\big(X_T^*, X_T\big) + d\big(X_T^*, X\big) + d\big(X_T^*, X_T\big) \cdot R(X_T) \Big]$$

METHOD

Accessing the decision boundary

- Defining the decision boundary is a challenging task. We produce adversarial borderline instances to get access to the decision boundary.
- We tweak the basic autoencoder by modifying its loss function to produce critical instances right along the decision boundary.
- The algorithm is followed by a bisection method to add refinement to the produced critical samples.



Two types of interventions

- *Minimal intervention- uses nearest neighbor* (identifies the solution with minimal change)
- *Constrained intervention- uses grid search* (identifies the solution that reflects user preferences)

Modified loss functions of the autoencoders

$$\mathcal{L}_1 = \sum_{\forall X_n} \min \Big[||X_n - AE_1(X_n)||^2 + \alpha \times CE\Big(f\big(AE_1(X_n)\big), \vec{a}\Big) \Big]$$

$$\mathcal{L}_2 = \sum_{\forall X_a} \min \Big[||X_a - AE_2(X_a)||^2 + \alpha \times CE\Big(f\big(AE_2(X_a)\big), \vec{n}\Big) \Big]$$

RESULTS

Visualizing the decision boundary

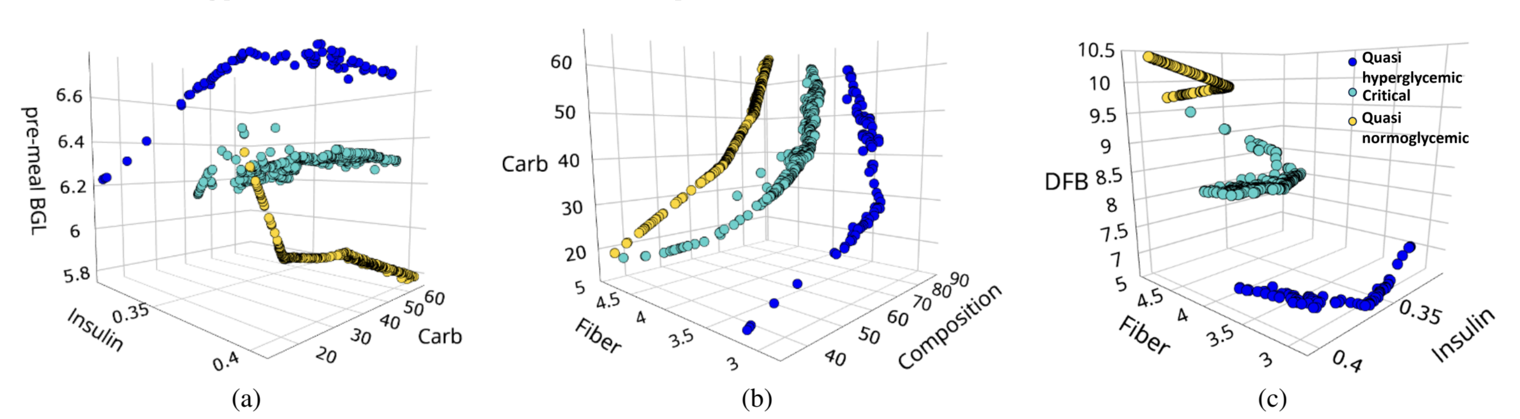


Figure 4: Outputs of the autoencoders and the bisection algorithm, i.e., the borderline instances, are shown across three plots. The plots display six out of the seven features, grouped as (a) pre-meal BGL (i.e. SBGL), insulin intake, CHO, (b) CHO, fiber intake, CHO composition, and (c) DFB (i.e. time elapsed since last insulin), fiber intake, insulin intake. This visualization effectively demonstrates the precise placement of critical instances between the quasi-hyperglycemic and normoglycemic samples.

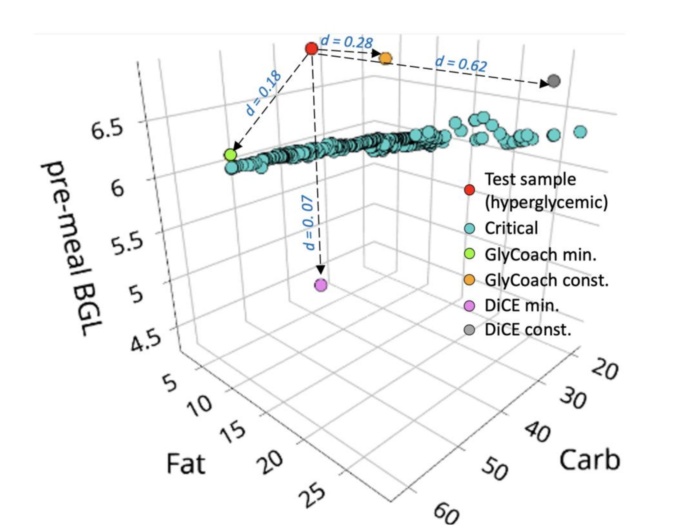


Figure 5: Trajectory comparison for DiCE, minimal and constrained interventions of GlyCoach.

Table 1: Examples of minimal and constrained interventions made on hyperglycemic pre-meal contexts across the two datasets.

Pre-meal context	Intervention
Nutrition Absorption dataset	
Aaron's current blood glucose level is 6.8 mmol/L. His Lunch contains 60.5g CHO, 9.4g fat and 4.2g fiber. He took 0.32 unit bolus 8 minutes before lunch.	(**minimal**) Increase CHO by 1.3g, fat by 2.6g, take a bolus of 0.4 units and wait until blood glucose drops to 6 mmol/L to stay normoglycemic
Emily prepared a meal with 74.2g CHO, 6.4g fat and 3.1g fiber. She has already taken 0.32 unit bolus 5 minutes back and her glucose level is 7.2 mmol/L. Emily wants to eat her meal while avoiding another insulin dose to stay normoglycemic.	(**constrained**) She can remain normoglycemic just by reducing the carb amount to 21.3 grams.
OhioT1DM	
Sarah intends to eat 100 gram carbohydrate at 8 mmol/L pre-meal blood glucose level. Over the past hour, she completed total exercise of 10 units, administered 1.2 unit total basal and took 6.5 units total bolus 50 minutes ago.	(**minimal**) Reducing CHO to 84.5g, total basal to 0.64 units, bolus intake time to 20.4 minutes, increasing total exercise to 19.6 units, and waiting until blood glucose level drops to 7.6 mmol/L will prevent hyperglycemia.
Harry aims to avoid hyperglycemia without changing his planned 58g CHO meal. His blood glucose level is 7.5 mmol/L. Over the past hour, he took 1.1 units basal, did 4 units exercise and 10 units work. He took 2.8 units bolus 50 minutes ago.	(**constrained**) Harry needs to wait until his pre-meal blood glucose level becomes 4.3 mmol/L to avoid having hyperglycemia.

Table 2: Evaluating the explanations via external simulator sensitivity, L_2 normalized distance, PV penalty and Cover.

	Sensitivity (%)	$\hat{d}$	PV penalty	Cover (%)
Nutrition Absorption				
GlyCoach	**82.6**	0.239	17.47	96.8
DiCE	72.3	0.353	**8**	90.3
Baseline	41.7	**0.176**	35.87	**100**
OhioT1DM				
GlyCoach	**90.6**	**0.314**	**10.2**	90.3
DiCE	80.4	0.472	13.1	80.6
Baseline	70.2	0.518	33.95	**100**

Datasets

Nutrition Absorption

- 4 participants, 167 meal records, supplemented with artificial data
- Postprandial Glycemic responses were recorded using a CGM
- Dataset contains meal macronutrients (carb, fat, fiber), insulin dose, last insulin intake time, glycemic responses of all participants
- Task: How the user needs to modify his behavior/food intake to stay normoglycemic

OhioT1DM

- 12 Type 1 diabetes patients, monitored for 8 weeks
- Self reported CHO amount, sleep hours, exercise and work intensity
- Heart rate, skin response (GSR), acceleration, skin temperature were recorded on Empatica wristband
- Participants wore Medtronic 530G/630G insulin pumps and Medtronic Enlite CGM that transmits blood glucose level every 5 minutes
- 44 days for training and 12 days for testing

CONCLUSIONS

GlyCoach has enormous potential in digital health and chronic disease management. Developing an AI driven intervention is hard, however, deploying it in real world settings is even harder. The main limitation of GlyCoach is the lack of validation through user studies resulting in uncertain performance in a clinical setting. Therefore, the next steps for GlyCoach involve creating a mobile app for it and validating its performance through user studies and gather expert/user opinions to understand what they think about this technology and its performance.

ACKNOWLEDGEMENTS

Check to make sure you've acknowledged partner and funding agencies, either with text or with their logos.

College of Health Solutions

Arizona State University