GlucoBench: Curated List of Continuous Glucose Monitoring Datasets with Prediction Benchmarks

Renat Sergazinov[1], Elizabeth Chun[1], Valeriya Rogovchenko[1], Nathaniel Fernandes[2], Nicholas Kasman[2], Irina Gaynanova[1]

[1]Department of Statistics, [2]Department of Electrical and Computer Engineering Texas A&M University

International Conference On Learning Representations

March 25, 2024

Continuous glucose monitors

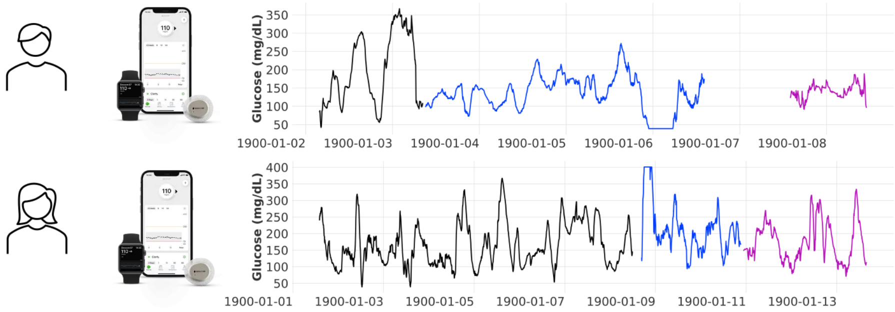


Figure: Sample of glucose curves captured by the Dexcom G4 Continuous Glucose Monitoring (CGM) system, with dates de-identified for privacy [5].

Problem

Table: Summary of the glucose prediction models by dataset and model type.

Type	Diabetes	# of datasets	# of deep	# of shallow	# of Physiological
Open	Type 1	9	13	3	2
Simulation	Type 1	12	3	3	6
Proprietary	Mixed	22	7	8	7

Limitations:

1. **Lack of Benchmarks:** few open datasets, no tasks, no pre-processing tools;
2. **Open-Source Shortage:** 38 out of 45 surveyed methods released code;
3. **Narrow Focus:** exclusion of Type 2 diabetes from the datasets.

Our approach: data

Table: Proposed suite of open datasets.

Dataset	Diabetes	CGM	# of Subjects		Age		Sex (M / F)	
	Overall	Overall	Raw	Processed	Raw	Processed	Raw	Processed
Broll [1]	Type 2	Dexcom G4	5	5	NA	NA	NA	NA
Colas [2]	Mixed	MiniMed iPro	208	201	59	59	103 / 104	100 / 100
Dubosson [3]	Type 1	MiniMed iPro2	9	7	NA	NA	6 / 3	NA
Hall [4]	Mixed	Dexcom G4	57	56	48	48	25 / 32	NA
Weinstock [5]	Type 1	Dexcom G4	200	192	68	NA	106 / 94	101 / 91

Our approach: pre-processing and tasks

Systematic pre-processing across datasets:

1. Interpolation and Segmentation: linear interpolation or segment division.
2. Covariates Scaling and Encoding: scaling and label encoding.
3. Data Splitting: chronologically ordered + out-of-distribution set.

To create a fair comparison and highlight main difficulties in CGM prediction,we create the following **task setup**:

1. **In-distribution fit:** for patients in training data;
2. **Out-distribution fit:** for new patients (cold start);
3. **Inclusion of covariates:** support for covariates.

Results

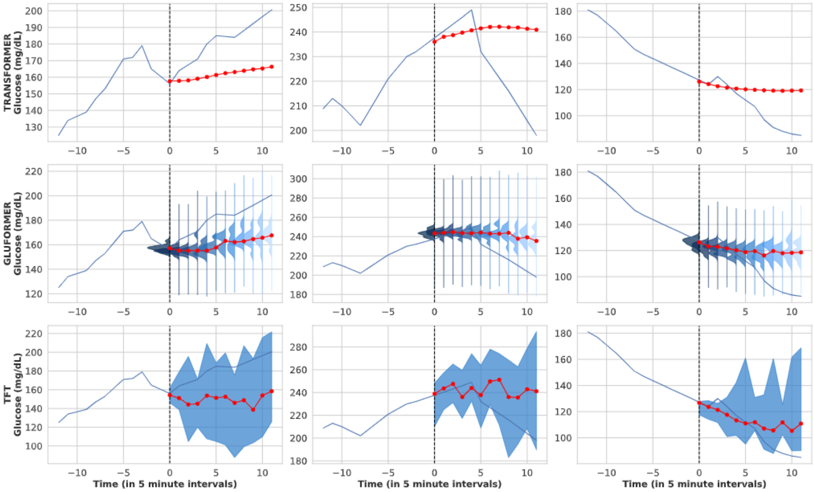


Figure: Model forecasts on Weinstock [5] dataset.

In-distribution performance

Table: In-distribution performance.

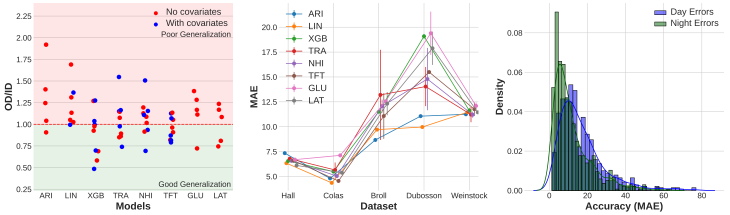
Accuracy	Broll		Colas		Dubosson		Hall		Weinstock	
	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE
ARIMA	**10.53**	**8.67**	5.80	4.80	13.53	11.06	8.63	7.34	13.40	11.25
Linear	11.68	9.71	**5.26**	**4.35**	**12.07**	**9.97**	7.38	6.33	13.60	11.46
Latent ODE	14.37	12.32	6.28	5.37	20.14	17.88	**7.13**	**6.11**	13.54	11.45
Transformer	15.12	13.20	6.47	5.65	16.62	14.04	7.89	6.78	**13.22**	**11.22**
Uncertainty	Lik.	Cal.	Lik.	Cal.	Lik.	Cal.	Lik.	Cal.	Lik.	Cal.
Gluformer	**-2.11**	**0.05**	**-1.07**	0.14	**-2.15**	**0.06**	**-1.56**	**0.05**	**-2.50**	0.08
TFT	–	0.16	–	**0.07**	–	0.23	–	0.07	–	**0.07**

Out-of-distribution performance

Table: with- vs. without-covariates, performance increase and decrease shown.

Accuracy	Broll		Colas		Dubosson		Hall		Weinstock	
	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE
Linear	-14.82%	-13.34%	+5.54%	+5.75%	+2.84%	+0.61%	+6.17%	+5.09%	-1.54%	-1.08%
Transformer	-15.14%	-14.64%	+30.31%	+37.56%	+64.99%	+73.82%	-5.06%	-5.4%	+9.33%	+12.41%
Uncertainty	Lik.	Cal.	Lik.	Cal.	Lik.	Cal.	Lik.	Cal.	Lik.	Cal.
TFT	–	+94.6%	–	+114.61%	–	+7.57%	–	+16.84%	–	-21.55%

Key takeaways



(a) Generalization (b) Error by population (c) Daytime vs. nighttime error

Figure: Analysis of errors by: (a) OD versus ID, (b) population diabetic type (healthy → Type 2 → Type 1), (c) daytime (9:00AM to 9:00PM) versus nighttime (9:00PM to 9:00AM).

Key takeaways cont.

Key takeaways:

1. Model Performance Variation Factors:
   1. **Dataset Size:** deep learning models excel on larger datasets.
   2. **Patient Composition:** healthy subjects being easier to predict than those with diabetes.
   3. **Time of Day:** daytime predictions are more challenging.
2. Model Generalizability:
   1. **Deep learning** models generally show better generalization.
   2. Performance typically drops on out-of-distribution (OD) data.
3. Impact of Covariates:
   1. Integrating covariates is non-trivial, and currently **no model is able to take full advantage of covariates.**

References

[1] S. Broll, J. Urbanek, D. Buchanan, E. Chun, J. Muschelli, N. M. Punjabi, and I. Gaynanova. Interpreting blood glucose data with r package iglu. *PloS one*, 16(4):e0248560, 2021.

[2] A. Colás, L. Vigil, B. Vargas, D. Cuesta-Frau, and M. Varela. 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, 2019.

[3] F. Dubosson, J.-E. Ranvier, S. Bromuri, J.-P. Calbimonte, J. Ruiz, and M. Schumacher. The open d1namo dataset: A multi-modal dataset for research on non-invasive type 1 diabetes management. *Informatics in Medicine Unlocked*, 13:92–100, 2018.

[4] H. Hall, D. Perelman, A. Breschi, P. Limcaoco, R. Kellogg, T. McLaughlin, and M. Snyder. Glucotypes reveal new patterns of glucose dysregulation. *PLoS biology*, 16(7):e2005143, 2018.

[5] R. S. Weinstock, S. N. DuBose, R. M. Bergenstal, N. S. Chaytor, C. Peterson, B. A. Olson, M. N. Munshi, A. J. Perrin, K. M. Miller, R. W. Beck, et al. Risk factors associated with severe hypoglycemia in older adults with type 1 diabetes. *Diabetes Care*, 39(4):603–610, 2016.

Thank You!