A large sensor foundation model pretrained on continuous glucose monitor data for diabetes management

Presenter: [Shovito Barua Soumma]

Date: Oct 22, 2025

- September 16, 2025
- npj Health Systems

Introduction

- Traditional AI models for glucose forecasting are task-specific, require patient-level tuning, and often fail to generalize.
- Introduces CGM-LSM, a Large Sensor Model inspired by large language models (LLMs) like GPT-2
  - captures universal glucose patterns across **diabetes types, ages, and genders**
  - Performs better in zero shot settings compared to SOTA in longer horizons

Method

- 5.9 M CGM readings from 592 diabetic patients (T1D & T2D) collected at 5-min intervals.
- **Instance Construction:** Each sample = **24 h input (288 points)** → predict **next 2 h (24 points)**; overlapping sliding windows used for data expansion.

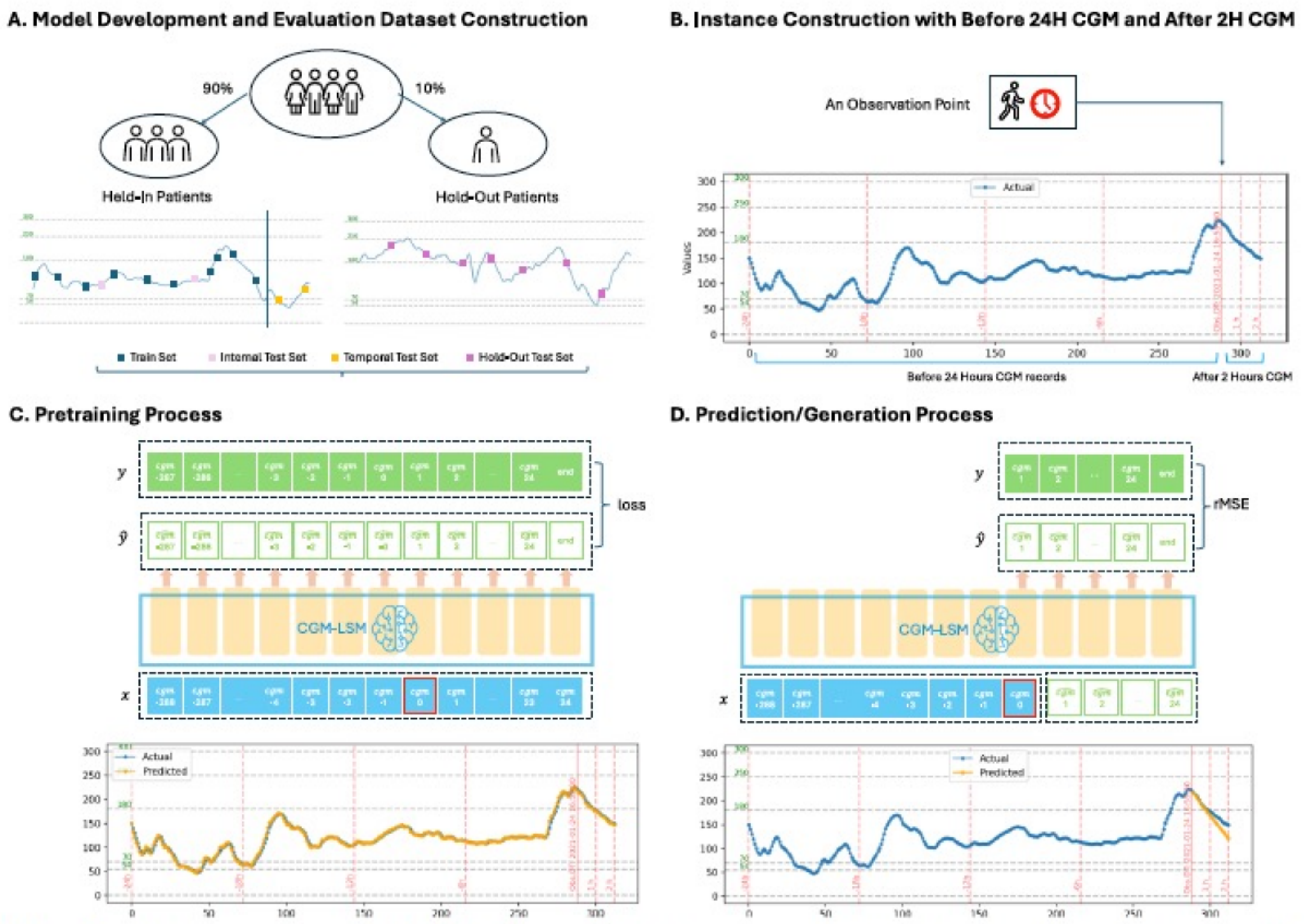


Fig. 1 | The workflow of dataset construction and CGM-LSM development. **A** The instances selection process to construct internal test set, temporal test set, held-out test set, and training set. **B** For one instance, the input-output pair construction process. Each instance is a combination of a patient and an instance datetime. The instance contains 288 CGM records before 24 h and 24 CGM records after 2 h. **C** Th pretraining process for one instance given 26-h CGM records. **D** The prediction (generation) process for one instance with 24-h CGM records.

Method

Known patients, known time period
Known patients, future time periods
Unknown patients

Internal	Same patients	Random samples	Checks in-sample accuracy
Temporal	Same patients	Future time periods	Tests prediction stability over time
Held-out	New patients	Entirely new people	Tests model generalization to unseen patients

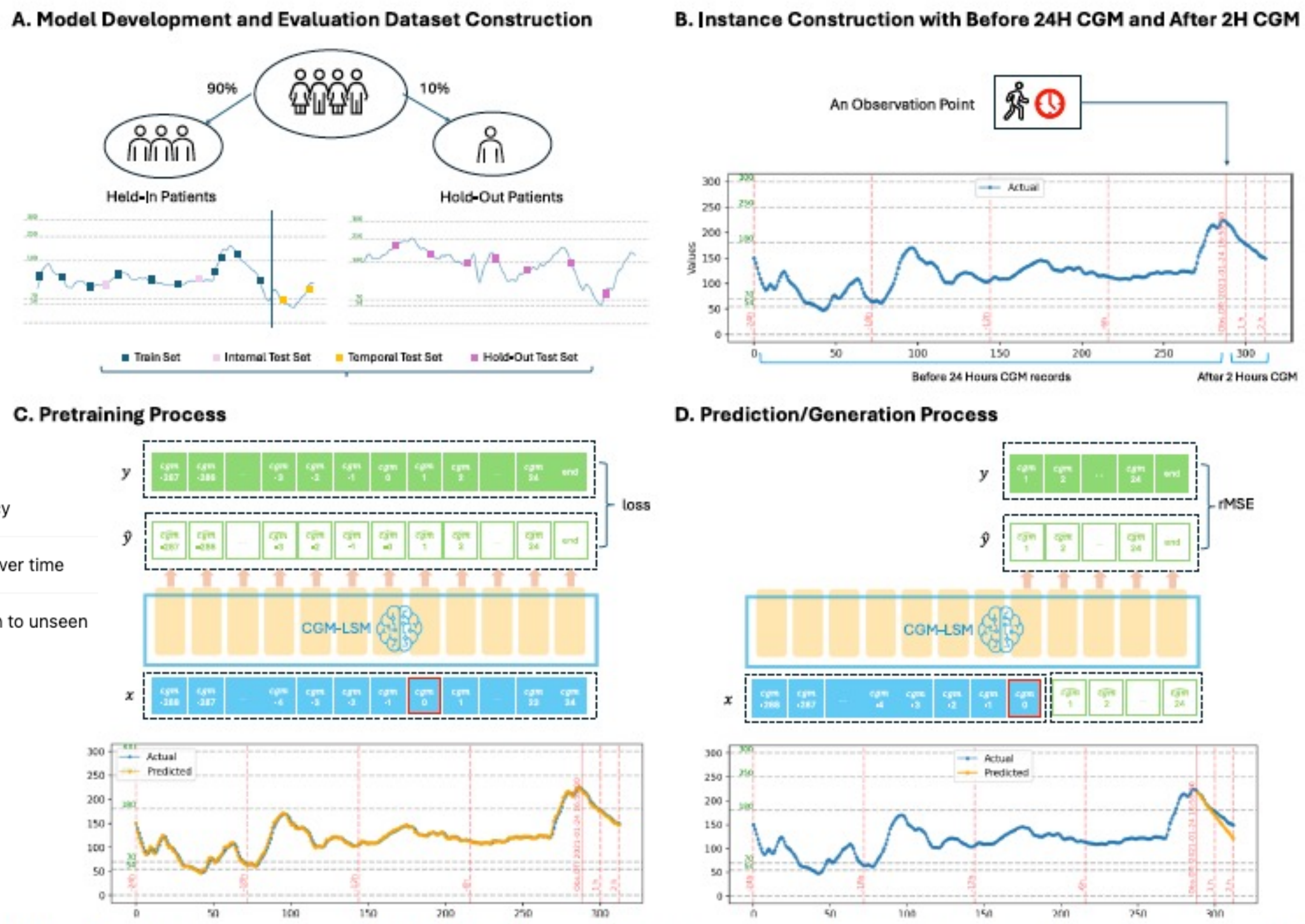


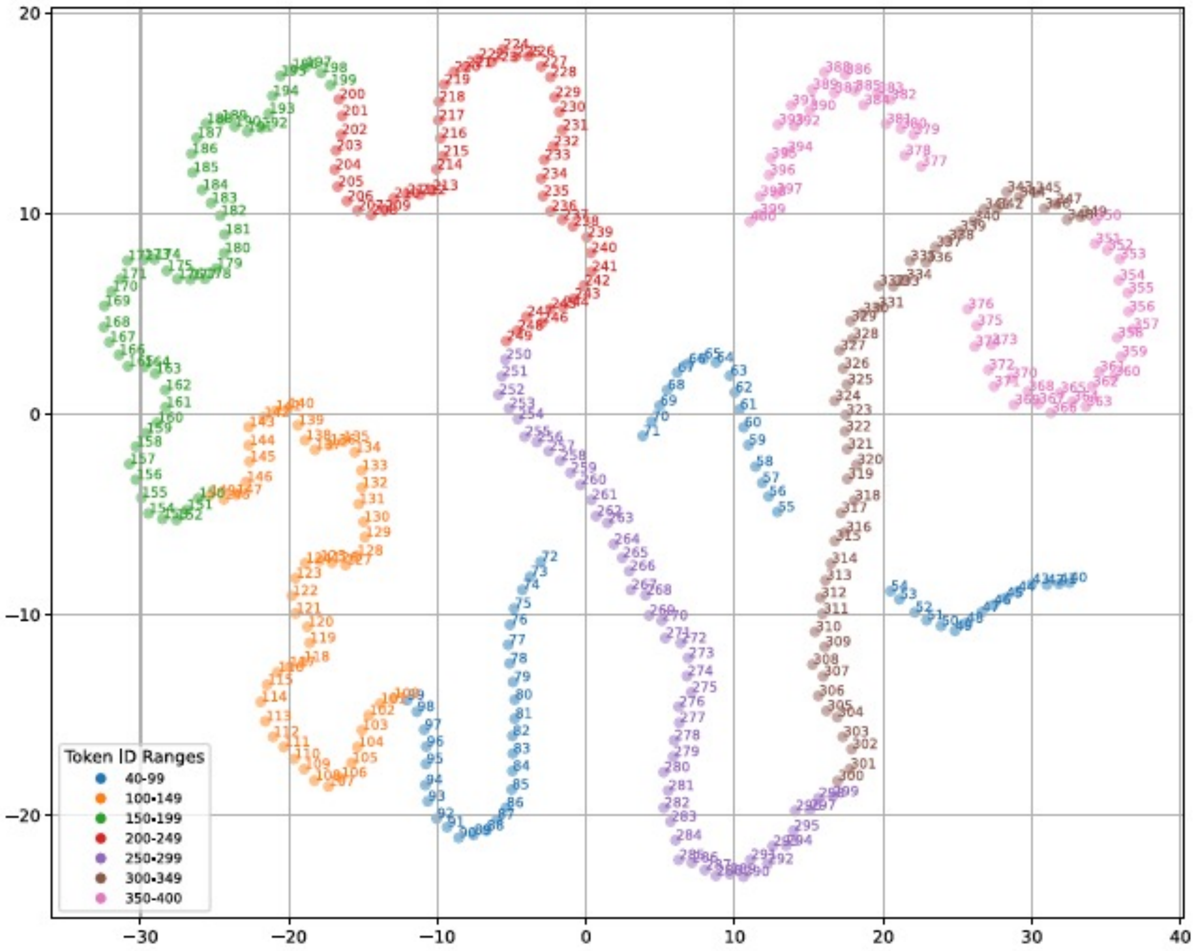
Fig. 1 | The workflow of dataset construction and CGM-LSM development. A The instances selection process to construct internal test set, temporal test set, held-out test set, and training set. **B** For one instance, the input-output pair construction process. Each instance is a combination of a patient and an instance datetime. The instance contains 288 CGM records before 24 h and 24 CGM records after 2 h. **C** T

Method

- Decoder based transformer (autoregressive)
- Discrete/categorical tokenization

Token ID	Glucose bin (mg/dL)
14	110–115
15	115–120
16	120–125
17	125–130

$$L(\theta) = \sum_i \log p_\theta(s_i | s_1 \ldots s_{i-1})$$



Results

Table 1 | Data description for the Welldoc dataset, showing the number of instances and patient counts across subsets by diabetes type (Type 1 Diabetes [T1D] and Type 2 Diabetes [T2D]), age groups, and gender

	Records (Patients)	Type 1 Diabetes (T1D)						Type 2 Diabetes (T2D)					
		Complete dataset	Dataset	Training	Internal	Temporal	Held-Out	Complete dataset	Dataset	Training	Internal	Temporal	Held-out
Patients	15,961,183 (592)	7,788,836 (291)	779,111 (290)	560,184 (257)	61,979 (256)	68,762 (257)	88,186 (33)	8,172,347 (301)	817,125 (301)	597,975 274)	66,257 (273)	73,293 (273)	79,600 (27)
Age group													
18–39 years old	3,425,955 (129)	2,901,601 (109)	290,615 (108)	209,054 (94)	23,321 (94)	25,837 (94)	32,403 (14)	524,354 (20)	52,077 (20)	39,252 (19)	4,440 (19)	4908 (19)	3,477 (1)
40-64 years old	8,040,123 (299)	3,249,786 (123)	324,649 (123)	231,576 (108)	25,408 (107)	28,412 (108)	39,253 (15)	4,790,337 (176)	478,441 (176)	347,663 (158)	38,481 (158)	42,483 (158)	49,814 (18)
65+ years old	4,495,105 (164)	1,637,449 (59)	163,847 (59)	119,554 (55)	13,250 (55)	14,513 (55)	16,530 (4)	2,857,656 (105)	286,607 (105)	211060 (97)	23,336 (96)	25902 (96)	26,309 (8)
Gender group													
Female	7,133,594 (283)	4,150,047 (165)	415,869 (164)	295,753 (144)	32,910 (144)	36,478 (144)	50,728 (20)	2,983,547 (118)	297,557 (118)	224,985 (109)	24,768 (108)	27,582 (108)	20,222 (9)
Male	8,827,589 (309)	3,638,789 (126)	363,242 (126)	264,431 (113)	29,069 (112)	32,284 (113)	37,458 (13)	5,188,800 (183)	519,568 (183)	372,990 (165)	41,489 (165)	45,711 (165)	59,378 (18)

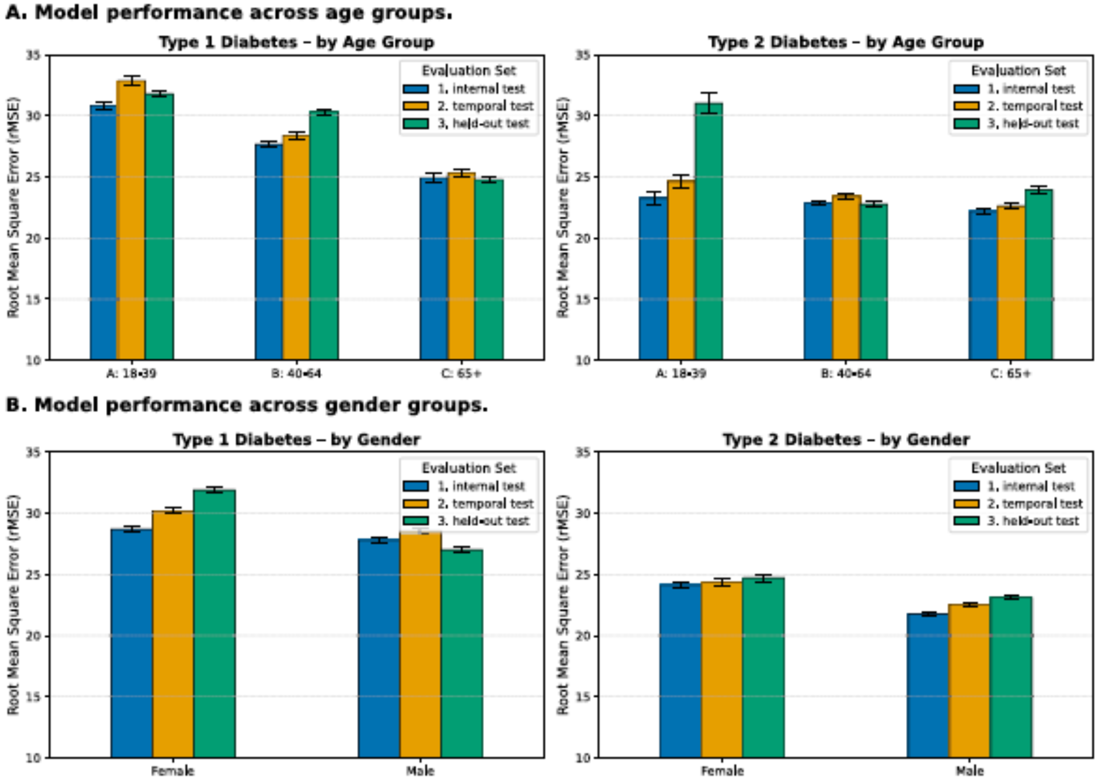
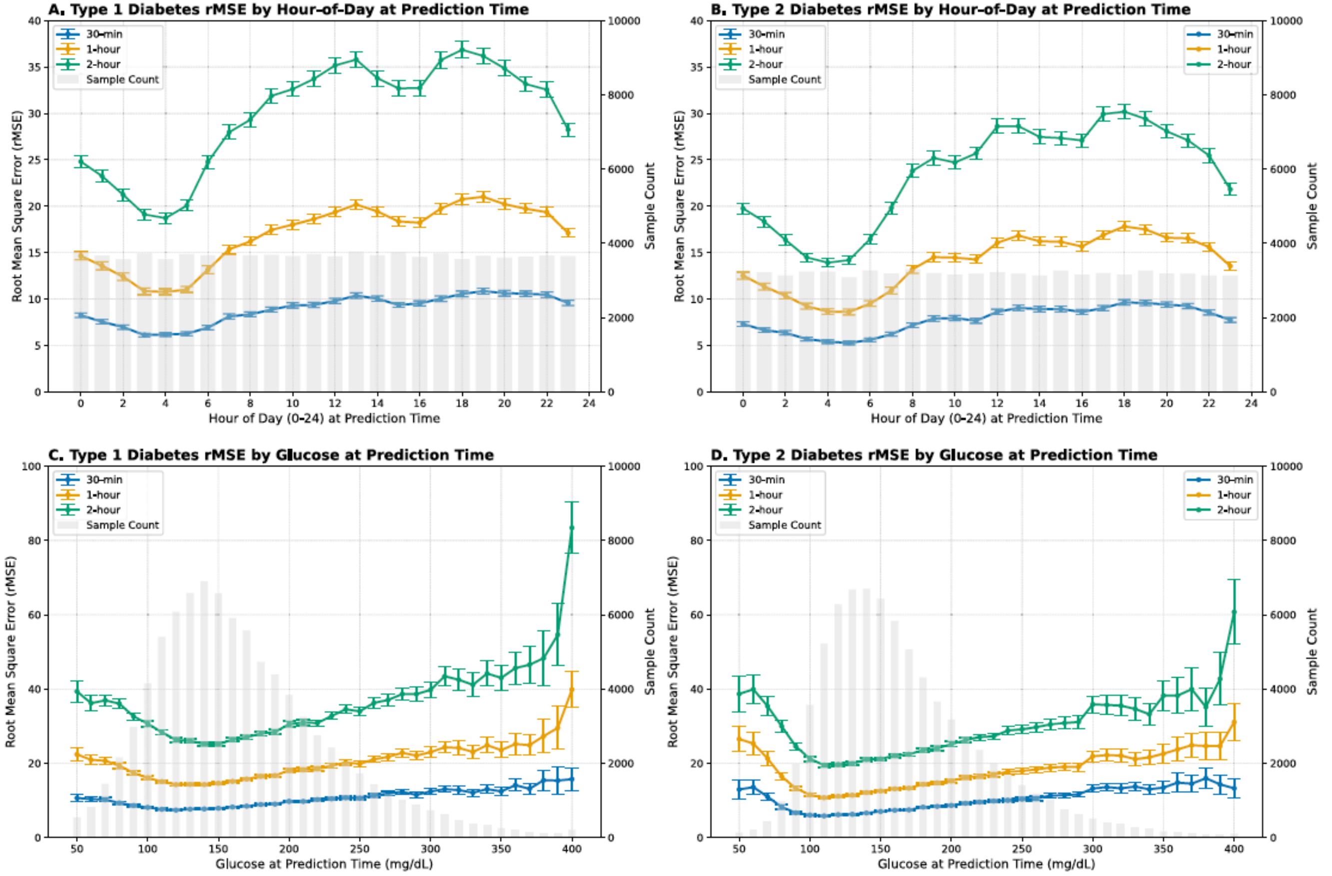


Fig. 3 | Model performance in 2-h glucose prediction across age and gender. A Model performance across age groups (18–39, 40–64, 65+), showing rMSE-2H values for Type 1 and Type 2 Diabetes prediction across different test sets. B Model performance across gender, showing rMSE-2H values for Type 1 and Type 2 Diabetes prediction across different test sets.



Results: Zero Shot settings (Benchmarking)

Table 2 | Comparative performance of predictive models for future glucose levels, using Root Mean Square Errors (rMSE)

Model	Dataset		rMSE-30m	rMSE-1h	rMSE-2h
LSTM	OhioT1DM		36.022 (0.551)	37.17 (0.513)	38.703 (0.461)
RNN			36.102 (0.552)	37.344 (0.512)	38.952 (0.458)
GRU			37.555 (0.554)	38.573 (0.517)	40.108 (0.463)
Transformer			27.886 (0.463)	30.869 (0.462)	36.653 (0.47)
Informer			35.197 (0.501)	36.962 (0.485)	40.204 (0.463)
Autoformer			36.08 (0.552)	38.352 (0.542)	41.395 (0.515)
CGMLSM			9.024 (0.168)	15.895 (0.283)	26.876 (0.43)
CGM-LSM	WellDoc T1D	Internal Test	8.403 (0.066)	16.049 (0.118)	28.277 (0.188)
		Temporal Test	9.155 (0.068)	17.013 (0.118)	29.426 (0.184)
		Held-Out Test	8.926 (0.056)	16.905 (0.101)	29.812 (0.16)
	WellDoc T2D	Internal Test	7.441 (0.055)	13.418 (0.094)	22.649 (0.147)
		Temporal Test	8.025 (0.058)	14.073 (0.095)	23.216 (0.143)
		Held-Out Test	7.772 (0.055)	13.877 (0.091)	23.494 (0.143)

Each entry displays the mean rMSE followed by the confidence interval width in parentheses, indicating the range within which the true mean is expected to lie with 95% confidence.

Limitation

- Need to study scaling laws (how data/model size affects performance)
  - Fix data (pf vs complexity)
  - Fix parameter (pf vs data)

- No inclusion of context (meals, insulin, exercise) (rare + discrete)