

FMGP: Foundation-Model-Guided Personalization for Cold-Start Continuous Glucose Forecasting

Asihati Hazaiti, Yiting Wei, *Student Member, IEEE*, and Nima TaheriNejad, *Member, IEEE*

Abstract—Continuous glucose forecasting is critical for diabetes management, yet substantial inter-patient heterogeneity limits the generalization of existing models to unseen patients. Current cold-start forecasting approaches primarily rely on meta-learning and few-shot adaptation, enabling rapid personalization through a shared model initialization while making limited use of patient-specific prior knowledge. To address this limitation, we propose Foundation-Model-Guided Personalization (FMGP), a novel framework that leverages foundation-model-derived patient semantic representations to drive model personalization. FMGP employs PubMedBERT to encode patient profiles into semantic representations and utilizes a semantic hypernetwork to generate personalized network parameters, establishing a new “Personalize Before Adapt” paradigm for cold-start forecasting. Simulations on three public datasets show that, under the cold-start setting with only $K = 5$ support samples, FMGP achieves RMSE values of 9.44, 17.64, and 17.97 mg/dL, respectively, outperforming all compared methods. Repeated patient-level evaluations further demonstrate stable predictive performance. Clinical analysis shows that 96.288% and 94.190% of predictions fall within Clarke Zones A+B on CGMacros and OhioT1DM, respectively, with no Zone E errors observed on either dataset. These results demonstrate that patient semantic representations can serve not only as auxiliary features but also as informative priors for model personalization before adaptation, providing a new direction for applying foundation models to personalized healthcare forecasting.

Index Terms—Continuous Glucose Forecasting, Cold-Start Learning, Patient Representation Learning, Foundation Models, Personalized Healthcare

I. INTRODUCTION

With the rapid advancement of precision medicine and personalized healthcare, medical artificial intelligence is shifting from population-level decision-making toward individualized patient-centered support [1]. By leveraging patient-specific characteristics, personalized healthcare aims to provide more accurate disease management strategies [2]. In diabetes care, Continuous Glucose Monitoring (CGM) forecasting enables early warning of hypo- and hyperglycemic events, supporting proactive intervention and improved glycemic control [3]. However, glucose regulation varies substantially across individuals due to differences in insulin sensitivity, dietary habits, physical activity, disease progression, and metabolic conditions [4]. This inter-patient heterogeneity challenges population-level forecasting models and limits their generalization to unseen patients.

Asihati Hazaiti and Yiting Wei contributed equally to this work.

This work was partially supported by Hector Stiftung. Asihati Hazaiti is with the Department of Informatics, University of Hamburg, Hamburg, Germany. Yiting Wei and Nima TaheriNejad are with the Institute of Computer Engineering (ZITI), Heidelberg University, Heidelberg, Germany. Corresponding author: Yiting Wei (e-mail: yiting.wei@uni-heidelberg.de).

To address this challenge, personalized glucose forecasting has received increasing attention. Traditional approaches typically require substantial patient-specific historical data for individualized fine-tuning [5], while meta-learning enables rapid adaptation to new patients with only a few labeled samples by learning transferable knowledge across patients [6]. However, existing meta-learning methods generally rely on a shared initialization, with personalization achieved only through subsequent adaptation [7]. This assumption overlooks patient-specific information already available before adaptation, including demographic, clinical, physiological, and behavioral characteristics [8]. Thus, cold-start forecasting involves not only limited samples but also limited patient awareness [9]. Existing approaches can learn from a few patient observations but lack an explicit understanding of the patient before adaptation, creating a representation gap between observed data and patient-specific characteristics [10].

Meanwhile, the rapid development of foundation models in healthcare has created new opportunities to bridge this gap. Biomedical foundation models such as PubMedBERT, ClinicalBERT, and GatorTron have demonstrated remarkable capabilities in learning transferable medical knowledge from large-scale biomedical literature and clinical corpora [11]. Such representations capture latent physiological patterns that are difficult to express using manually engineered features alone. Nevertheless, in existing glucose forecasting studies, patient semantic information is typically incorporated as auxiliary input features through simple feature concatenation, while its potential role in directly driving model personalization remains largely unexplored [12].

Based on these observations, we argue that personalization should begin before adaptation by leveraging patient semantic information as prior knowledge. A forecasting model should not only learn how a patient’s glucose changes, but also understand who the patient is. We therefore propose Foundation-Model-Guided Personalization (FMGP), a framework that uses foundation-model-derived patient representations to personalize cold-start continuous glucose forecasting before few-shot adaptation, shifting from the conventional “Adapt Then Personalize” paradigm to “Personalize Before Adapt.”

The main contributions are summarized as follows: 1) We propose FMGP, integrating patient semantic priors, enhanced temporal representations, and few-shot information for personalized cold-start glucose forecasting. 2) We introduce a semantic-driven personalization mechanism that transforms patient semantics into patient-specific regression-head initialization and layer-wise FiLM parameters for model-level personalization. 3) We extend the “Personalize Before Adapt”

paradigm to integrate pre-adaptation semantic initialization, K-shot parameter adaptation, and prediction-time support-memory retrieval. 4) We validate FMGP on CGMacros, OhioT1DM, and DiaTrend, demonstrating its effectiveness and generalizability across patient populations.

II. RELATED WORK

Recent advances in deep learning have significantly improved continuous glucose forecasting by leveraging temporal patterns in CGM data. Zhu et al. developed Population-Specific TFT, incorporating demographic and physiological information to account for patient heterogeneity [13]. Gowriswari et al. proposed a hybrid TCN-LSTM architecture to capture both short- and long-term glucose dynamics [14]. Piao et al. proposed GluADFL for cross-patient learning, enabling cold-start forecasting and patient-specific fine-tuning for unseen patients [15]. Lim et al. developed a virtual CGM framework that incorporates behavioral features, such as dietary intake and physical activity, and achieves personalization through individual-specific fine-tuning [16]. These methods primarily rely on subsequent fine-tuning or adaptation to achieve personalization. Patient representation learning has also emerged as an important direction in precision health-care. Dwivedi et al. proposed UniStruct to model structured patient information from electronic health records, improving patient-state representations and downstream generalization [17]. However, existing glucose forecasting methods mainly focus on adaptation-based personalization, while patient representations are typically used only as auxiliary inputs. How to directly leverage patient semantic information to drive model personalization before adaptation remains largely unexplored.

III. MATERIALS AND METHOD

A. Dataset

a) *CGMacros Dataset*: The CGMacros dataset [18] is a publicly available multimodal dataset designed for personalized nutrition and metabolic monitoring. It contains data from 45 participants, including 15 healthy individuals, 16 pre-diabetic individuals, and 14 participants with Type 2 Diabetes (T2D). For each participant, approximately 10 consecutive days of CGM data were collected together with demographic, anthropometric, clinical, dietary, and microbiome information. Its rich patient-level metadata makes it well suited for personalized glucose forecasting.

b) *OhioT1DM Dataset*: The OhioT1DM dataset [19] is a widely used benchmark for blood glucose forecasting and diabetes management research. It contains eight weeks of real-world data collected from 12 individuals with Type 1 Diabetes (T1D) receiving insulin pump therapy and CGM. In addition to CGM measurements recorded at 5-minute intervals, the dataset includes insulin, meal, activity, sleep, and other physiological and behavioral information. Its longitudinal and multimodal characteristics make it a challenging benchmark for personalized and cold-start glucose forecasting.

TABLE I: Patient Information Used for Semantic Profile Construction

Category	Features
Demographic	Age, Sex, BMI
Clinical	HbA1c, Fasting Glucose
CGM Statistics	Mean Glucose, Standard Deviation (SD), Coefficient of Variation (CV), Time in Range (TIR), Time above Range (TAR), Time below Range (TBR)
Behavioral	Meal Mode, Physical Activity, Circadian Features
Derived Phenotypes	Stable Glucose, High Variability, Frequent Hyperglycemia, Hypoglycemia Risk

SD: Standard Deviation; CV: Coefficient of Variation; TIR: Time in Range; TAR: Time Above Range; TBR: Time Below Range.

B. FMGP Overview

As illustrated in Fig. 1, FMGP consists of a Semantic Phase and a Temporal Phase. In the Semantic Phase, a patient profile is constructed from available patient information and historical glucose records, and encoded by a pretrained PubMedBERT. The resulting semantic representation is transformed by a semantic hypernetwork into patient-specific regression-head initialization and layer-wise FiLM parameters. In the Temporal Phase, a CGM-pretrained TCNFormer models extended historical CGM sequences, complemented by episode-level glucose dynamics. During K-shot personalization, a small support set is used to adapt both the regression head and FiLM parameters while keeping the shared backbone and semantic hypernetwork fixed. The support samples are further encoded into a patient-specific memory for query-to-support retrieval. Together, these components enable personalized cold-start forecasting for previously unseen patients.

C. Semantic Phase

1) *Patient Profile Construction*: We construct a patient profile containing demographic, clinical, historical glycemic, and behavioral information, as shown in Table I. To enhance clinical interpretability, phenotype tags are derived from these statistics, including stable glucose, high glucose variability, frequent hyperglycemia, and potential hypoglycemia risk. Finally, the patient information is organized into a unified natural-language template to form a patient-specific semantic description for subsequent representation learning.

2) *Patient Representation Learning*: After constructing the patient profile, we employ a frozen PubMedBERT encoder to extract patient semantic representations. Let P_i denote the profile text of patient i . The corresponding patient semantic representation is obtained as $s_i = E_{\text{PubMedBERT}}(P_i)$, where $E_{\text{PubMedBERT}}$ denotes the frozen PubMedBERT encoder and s_i represents the resulting patient semantic embedding.

In this work, the hidden representation corresponding to the [CLS] token is used as the patient-level semantic representation. To improve computational efficiency, all patient embeddings are precomputed and cached prior to meta-training, thereby avoiding repeated inference through the language model during episodic optimization. FMGP uses patient semantic representations as conditioning signals to directly drive parameter generation and network modulation, enabling patient-specific personalization prior to adaptation.

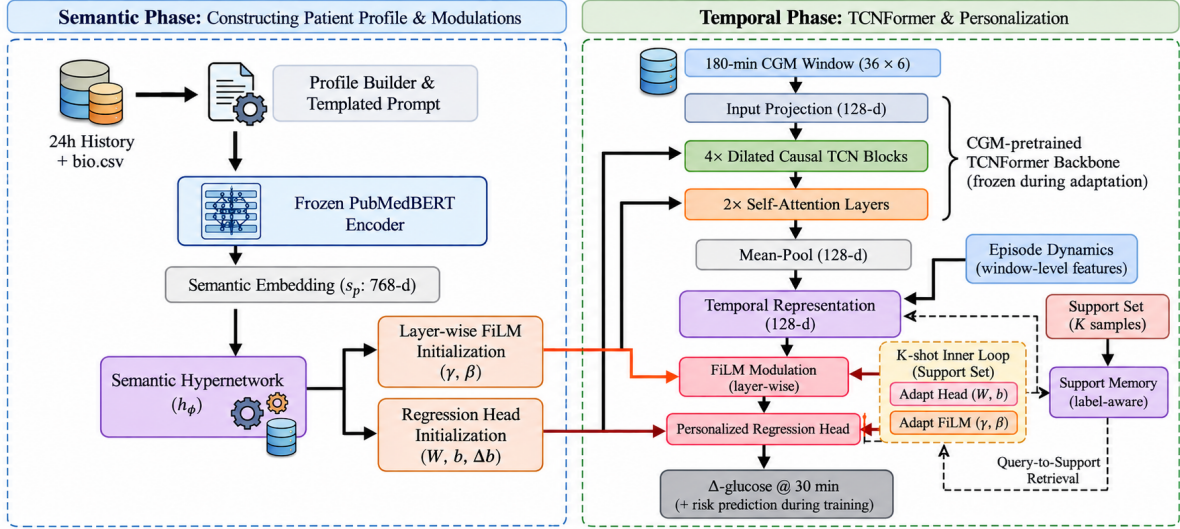


Fig. 1: Framework Overview of FMGP: Semantic-Guided Personalized Modulation for Temporal Glucose Prediction.

3) *Semantic Hypernetwork*: To transform patient semantic representations into model personalization capabilities, we design a semantic hypernetwork. Given the patient semantic embedding s_i , the hypernetwork generates layer-specific Feature-wise Linear Modulation (FiLM) modulation parameters as $(\gamma_l, \beta_l) = H_\phi^{(l)}(s_i)$, where H_ϕ denotes the semantic hypernetwork, and γ_l and β_l represent the scaling and shifting parameters for the l -th layer, respectively.

FiLM is then applied to modulate the hidden representations as $\hat{h}_l = \gamma_l \odot h_l + \beta_l$, where h_l and $\hat{h}_l$ denote the original and modulated hidden representations, respectively. The proposed TCNFormer consists of four Temporal Convolutional Network (TCN) blocks and two self-attention layers. Consequently, the semantic hypernetwork generates six sets of hierarchical FiLM parameters, allowing patient semantics to influence temporal representation learning at multiple network depths.

In addition to intermediate-layer modulation, the semantic hypernetwork also generates personalized initialization parameters for the regression head as $W_i^{(0)} = \gamma_h \odot W + \beta_h$ and $b_i^{(0)} = b + \Delta b$, where W and b denote the shared meta-parameters, and γ_h , β_h , and Δb are generated from the patient semantic embedding. Semantic information not only participates in feature representation learning but also directly contributes to prediction parameter generation, enabling patient-specific forecasting capabilities before adaptation begins.

D. Temporal Phase

1) *Personalized TCNFormer*: We adopt a personalized TCNFormer as the backbone for glucose forecasting. For each prediction sample, the model receives 36 time steps (180 minutes) of historical observations, which are projected into a 128-dimensional latent space. Four Dilated Causal TCN residual blocks with dilation factors of 1, 2, 4, and 8 are then used to capture multi-scale temporal patterns while preventing future information leakage. Two Transformer self-attention layers,

each with four attention heads and GELU activation, further model long-range temporal dependencies. To characterize the current glycemic state, an episode-dynamics branch extracts statistical and temporal features from the input window and integrates them with the backbone representation through residual fusion. All features are derived from the current input window to avoid future information leakage. Layer-wise FiLM parameters generated by the semantic hypernetwork modulate intermediate representations, enabling patient semantics to guide temporal representation learning at multiple levels. Finally, the resulting representation is fed into a personalized functional regression head to predict the glucose change over the next 30 minutes, together with an auxiliary glycemic-risk prediction used during training.

2) *K-shot Adaptation*: Although semantic modulation provides patient awareness, additional adaptation is required to capture patient-specific glucose dynamics. For each unseen patient, the first 24 hours of historical data are used to construct the patient profile and semantic representation. Subsequently, K samples are selected from the remaining time series as the support set, while the rest form the query set for evaluation.

For each episode, personalized FiLM parameters and regression-head initialization parameters are generated from the patient semantic embedding. The support set is then used to adapt both the regression head and layer-wise FiLM parameters, while the shared backbone and semantic hypernetwork remain fixed. Specifically, the regression head is updated as $(W_i^{(t+1)}, b_i^{(t+1)}) = (W_i^{(t)}, b_i^{(t)}) - \alpha \nabla_{W_i, b_i} \mathcal{L}_{S_i}$, while the FiLM parameters are updated as $\gamma_{i,l}^{(t+1)} = \gamma_{i,l}^{(t)} - \alpha_f \nabla_{\gamma_{i,l}} \mathcal{L}_{S_i}$ and $\beta_{i,l}^{(t+1)} = \beta_{i,l}^{(t)} - \alpha_f \nabla_{\beta_{i,l}} \mathcal{L}_{S_i}$. Here, W_i and b_i denote the patient-specific regression-head parameters, $\gamma_{i,l}$ and $\beta_{i,l}$ denote the FiLM parameters of the l -th fusion layer, α and α_f are their corresponding inner-loop learning rates, and $\mathcal{L}_{S_i}$ denotes the support-set loss for patient i .

The same K -shot samples are also encoded into a patient-

specific support memory containing their representations, observed glucose changes, and glycemic-risk information. For each query, scaled dot-product attention retrieves relevant support information, which is residually integrated into the query representation before prediction.

E. Experimental Protocol

We adopt a repeated patient-level cold-start meta-learning evaluation protocol to assess the generalizability of FMGP to unseen patients. For CGMacros, patients are divided into meta-training, meta-validation, and meta-test sets using a 30/5/10 split, while a 6/2/4 patient-level split is used for OhioT1DM. Meta-training patients are used for episodic meta-training, meta-validation patients for model selection and early stopping, and meta-test patients only for final evaluation. All partitions are performed at the patient level to ensure that test patients remain completely unseen during training.

To reduce the variability associated with a single patient split, we conduct 50 rounds of repeated subject-wise random splitting and evaluation. In each round, all methods use identical patient partitions, random seeds, support sets, and query sets. Final results are aggregated across the 50 patient-level evaluation rounds, with 95% confidence intervals reported.

IV. RESULTS AND DISCUSSION

A. Comparison with Existing Methods

As shown in Table II, FMGP demonstrates strong performance for 30-minute CGM forecasting on both CGMacros and OhioT1DM. On CGMacros, FMGP achieves an Root Mean Square Error (RMSE) of 9.44 mg/dL and an Mean Absolute Error (MAE) of 6.70 mg/dL, reducing the two errors by approximately 7.5% and 8.7%, respectively, compared with the Transformer baseline, and outperforming Zero-order Hold, Simple Linear Regression, and Mamba-style SSM + Adaptive Gating. On the more challenging OhioT1DM dataset, FMGP achieves an RMSE of 17.64 mg/dL, outperforming all compared existing methods. In particular, FMGP reduces RMSE by approximately 8.5% compared with the strongest baseline, GluNet. FMGP also outperforms Time-LLM/Chronos, which leverages large-scale time-series foundation models, further demonstrating the advantage of explicit patient-semantic personalization for cold-start glucose forecasting. It should be noted that the compared methods employ different evaluation protocols. On OhioT1DM, the predefined split uses approximately 40 days of training data followed by 10 days of test data, whereas LOPO-CV uses the first 70% of each patient's chronological data for fine-tuning and the remaining 30% for testing. In contrast, FMGP is evaluated using 50 repeated patient-level splits and uses only five target-patient samples for adaptation when $K = 5$, imposing a stricter constraint on the amount of patient-specific data available and more closely reflecting the cold-start setting. The 95% confidence intervals of RMSE are 9.21–9.67 mg/dL on CGMacros and 17.19–18.09 mg/dL on OhioT1DM, indicating stable performance across different patient-level splits.

To further evaluate the consistency of patient-level improvements, Table III compares FMGP with GluNet, the strongest

baseline on OhioT1DM, across all 12 subjects. FMGP achieves lower RMSE for every subject, with a mean paired difference of -1.5963 mg/dL ($\Delta\text{RMSE} = \text{RMSE}_{\text{FMGP}} - \text{RMSE}_{\text{GluNet}}$), which is statistically significant (95% CI: -2.8445 to -0.3481 mg/dL; paired t -test, $p = 0.0168$). This consistent improvement indicates that the overall performance gain is not driven by only a small subset of patients, further supporting the stability and competitive performance of FMGP for few-shot personalized glucose forecasting.

B. Cross-Dataset Generalization Analysis

To further evaluate the generalizability of FMGP, we introduce DiaTrend [26] as an independent third dataset. DiaTrend includes 54 patients with T1D aged 19–74 years, comprising 27,561 days of CGM data and 8,220 days of insulin pump data, with CGM measurements recorded approximately every 5 min. It also provides patient-level information, including age, sex, HbA1c, CGM measurements, and insulin pump records, from two independent study cohorts.

As shown in Table IV, under the same 30-min, $K = 5$ few-shot setting, with evaluation conducted over 50 repeated random patient-level splits, FMGP achieves an RMSE of 17.97 mg/dL and an MAE of 12.97 mg/dL on DiaTrend. Compared with FKAN, the strongest baseline in terms of RMSE, FMGP further reduces RMSE by approximately 11.9%. FMGP also achieves a Pearson r of 0.958 and an R^2 of 0.918, with relatively narrow 95% confidence intervals across the reported metrics. Together with the results on CGMacros and OhioT1DM, these findings show that FMGP maintains stable few-shot personalized forecasting performance on an independent dataset with a larger cohort and richer longitudinal records, providing further evidence of its generalizability across patient populations.

C. Ablation Study and Semantic Personalization Analysis

1) *Ablation of the Personalization Mechanism:* As shown in Table VPooled Zero-shot (B0) provides a competitive baseline, while per-subject fine-tuning (B1) and Reptile (B4) perform relatively worse, indicating that conventional adaptation strategies do not consistently improve forecasting performance. FOMAML/ANIL (B2) performs strongly, achieving the best MAE. Directly using patient semantics as input features (B3) is also competitive, whereas FMGP achieves a lower RMSE, suggesting that semantic-driven FiLM modulation is more effective than simple feature concatenation for personalized forecasting. With three-seed ensembling, FMGP further reduces RMSE to 18.19 mg/dL and achieves the highest R^2 of 0.918, demonstrating consistent performance across different model initializations under the same patient split. Overall, these results support the effectiveness of combining patient-specific semantic conditioning with few-shot personalization in FMGP.

2) *Patient Semantic Information and Encoder Analysis:* Building on these results, Table VI further analyzes the contributions of different patient semantic information and encoder designs. Considering computational efficiency, this analysis is conducted over 10 repeated subject-wise splits. Removing

TABLE II: Performance Comparison with Existing Methods on CGMacros and OhioT1DM Datasets (30 mins, K=5)

Dataset	Method	Protocol	RMSE (mg/dL) (95% CI)	MAE (mg/dL) (95% CI)	Pearson r (95% CI)	R^2 (95% CI)
CGMacros	Transformer	Random split (patient-aware)	10.20	7.34	0.97	-
	Zero-order Hold [20]	Full-data evaluation	19.90	-	-	-
	Simple Linear Regression [20]	Full-data evaluation	30.59	-	-	-
	Mamba-style SSM + Adaptive Gating [21]	LOSO + within subject fine-tune	-	15.20	-	-
	FMGP (Ours)	Repeated patient-level split ($n = 50$) + few-shot	9.44 (9.21–9.67)	6.70 (6.51–6.88)	0.96 (0.96–0.97)	0.93 (0.92–0.93)
OhioT1DM	Gated Recurrent Unit (GRU) with Double Attention [22]	Predefined split	20.48	14.59	-	0.89
	GluNet (Dilated CNN-based Deep Neural Network) [23]	Predefined split	19.28	-	-	-
	BiT-MAML (Bi-LSTM + Transformer + MAML) [24]	LOPO-CV	24.89	18.35	-	-
	Time-LLM / Chronos [25]	Personalized	24.44	16.90	-	-
	FMGP (Ours)	Repeated patient-level split ($n = 50$) + few-shot	17.64 (17.19–18.09)	13.17 (12.81–13.53)	0.95 (0.95–0.95)	0.90 (0.89–0.90)

TABLE III: Paired Patient-Level Comparison between FMGP and the Strongest Baseline on the OhioT1DM Dataset

Subject	GluNet RMSE	FMGP RMSE	Δ RMSE	Subject	GluNet RMSE	FMGP RMSE	Δ RMSE
540	20.6175	19.4993	-1.1182	570	21.2634	13.7965	-7.4669
544	18.9648	17.1060	-1.8588	575	16.0126	15.6004	-0.4122
552	17.8949	17.0354	-0.8596	584	20.3351	18.5941	-1.7410
559	19.5561	17.4000	-2.1561	588	18.6562	17.0863	-1.5699
563	17.0754	16.4738	-0.6016	591	18.4227	18.3404	-0.0824
567	18.6109	17.5665	-1.0445	596	14.3037	14.0590	-0.2447

Note: All RMSE values are reported in mg/dL. Δ RMSE = $\text{RMSE}_{\text{FMGP}} - \text{RMSE}_{\text{GluNet}}$; negative values favor FMGP. Mean paired difference: Δ RMSE = -1.5963 mg/dL (95% CI: -2.8445 to -0.3481); paired t -test, $p = 0.0168$.

TABLE IV: Performance Comparison with Existing Methods on the DiaTrend Dataset (30 mins, K=5)

Method	RMSE (mg/dL) (95% CI)	MAE (mg/dL) (95% CI)	Pearson r (95% CI)	R^2 (95% CI)
Fourier Kolmogorov–Arnold Network (FKAN) [27]	20.40	-	-	-
MLP + LSTM + GRU [28]	22.50	16.38	-	-
DL Benchmark (8-model average) [29]	25.20	-	-	-
FMGP (Ours)	17.97 (17.65–18.29)	12.97 (12.72–13.23)	0.958 (0.956–0.961)	0.918 (0.913–0.924)

any patient-profile category degrades performance, with the largest degradation in RMSE observed when removing Derived Phenotypes (Δ RMSE = -0.2107 mg/dL), followed by Demographic (-0.1115 mg/dL) and Clinical information (-0.0481 mg/dL), while CGM Statistics and Behavioral information have smaller effects. These results indicate that FMGP integrates complementary patient information, with phenotype-level information playing a particularly important role in representing inter-patient differences.

To evaluate the representational value of PubMedBERT, we replace it with a lightweight Multilayer Perceptron (MLP) encoder using the same patient variables. The MLP achieves an RMSE of 8.3716 mg/dL, compared with 8.1886 mg/dL for the full FMGP (Δ RMSE = -0.1830 mg/dL), indicating that PubMedBERT provides additional representational value beyond the structured variables themselves. After removing CGM Statistics, the MLP variant remains slightly inferior to its corresponding PubMedBERT variant, suggesting that this advantage does not solely depend on CGM summary statistics. Finally, fine-tuning PubMedBERT yields an RMSE of 8.2581 mg/dL, slightly higher than 8.1886 mg/dL for the frozen version. Thus, under limited-data and few-shot settings, freezing PubMedBERT reduces unnecessary parameter updates and

TABLE V: Ablation Study of Personalization Strategies on the OhioT1DM Dataset (30-min Forecasting, Single Random Patient-Level Split)

Method	RMSE (mg/dL)	MAE (mg/dL)	Pearson r	R^2
ZOH (Naive Baseline)	22.3452	16.5700	0.9457	0.8755
B0 – Pooled Zero-shot	18.9052	14.1565	0.9580	0.9109
B1 – Per-subject Fine-tuning	21.2845	16.1533	0.9433	0.8871
B2 – FOMAML / ANIL	18.5042	13.3228	0.9577	0.9147
B3 – Semantics-as-input	18.8296	14.2346	0.9560	0.9116
B4 – Reptile	21.2787	16.0367	0.9422	0.8871
FMGP (Semantic FiLM, Single)	18.3731	13.5991	0.9570	0.9159
FMGP (3-Seed Ensemble, our proposed)	18.1862	13.7031	0.9580	0.9176

TABLE VI: Ablation Study of Patient Semantic Information and Encoder Design on the CGMacros Dataset (30-min Forecasting, $K = 5$)

Ablation Variant	RMSE (mg/dL)	SD	Δ RMSE (mg/dL)
w/o Demographic	8.3001	0.3599	-0.1115
w/o Clinical	8.2367	0.3666	-0.0481
w/o CGM Statistics	8.2103	0.3774	-0.0217
w/o Behavioral	8.2109	0.4177	-0.0223
w/o Derived Phenotypes	8.3993	0.5915	-0.2107
MLP replacing PubMedBERT	8.3716	0.4874	-0.1830
MLP replacing PubMedBERT (w/o CGM Statistics)	8.5499	0.5808	-0.3613
Fine-tuned PubMedBERT	8.2581	0.4612	-0.0695
Full FMGP	8.1886	0.3822	baseline

Note: Results are based on 10 repeated random subject-wise splits. “w/o” denotes masking of the corresponding patient-profile category. Δ RMSE is calculated relative to Full FMGP, with negative values indicating increased prediction error.

training burden while maintaining stable patient representations, making it well suited to cold-start personalization.

It should be noted that for concise presentation, Tables V and VI report results on OhioT1DM and CGMacros, respectively, while the other dataset shows consistent trends.

D. Robustness and Computational Efficiency

1) *Support-Set Size Sensitivity Analysis:* As shown in Table VII, when K increases from 5 to 20, the RMSE and standard deviation of FMGP vary only slightly on both CGMacros and OhioT1DM, indicating that FMGP is robust to the number of support samples. More importantly, increasing the support-set size provides only marginal additional improvement, while $K = 5$ already achieves performance comparable to that obtained with larger support sets. This demonstrates that FMGP can capture most of the personalization benefit using only a

TABLE VII: Sensitivity Analysis of FMGP under Different Support-Set Sizes (Repeated Patient-level Split ($n = 50$))

Dataset	Metric	$K = 5$	$K = 10$	$K = 15$	$K = 20$
CGMacros	RMSE (mg/dL)	9.4439	9.4274	9.4289	9.4200
	SD	0.8122	0.8103	0.8171	0.8094
OhioT1DM	RMSE (mg/dL)	17.6388	17.6122	17.5929	17.5806
	SD	1.5820	1.5653	1.5734	1.5728

TABLE VIII: Computational Cost and Deployment Overhead of FMGP

Deployment Indicator	Semantic Representation	Personalized Temporal Model	Complete FMGP
Total Parameters	109.48 M	1.54 M	111.02 M
Trainable Parameters	0	1.54 M	1.54 M
New-Patient Initialization / First-Forecast Latency	20.57 ms	58.73 ms	77.22 ms
Peak Allocated GPU Memory	433.11 MiB	30.66 MiB	449.52 MiB
Recurring Forecast Latency	Cached	5.87 ms/query	5.87 ms/query

Note: Component-wise latencies are measured independently and therefore do not necessarily sum to the end-to-end latency.

small amount of target-patient data, enabling accurate cold-start glucose forecasting without requiring substantial patient-specific data accumulation. Therefore, all main experimental results reported in this study are based on the $K = 5$ few-shot setting.

2) *Computational Efficiency Analysis*: As shown in Table VIII, although the semantic encoder accounts for 109.48M of the 111.02M total parameters, it remains frozen throughout the framework, leaving only 1.54M trainable parameters in FMGP. For a new patient, semantic representation generation takes 20.57 ms, while the complete process of patient initialization, K -shot personalization, and the first prediction takes 77.22 ms. Since the semantic representation is computed only once during initialization and subsequently cached, PubMedBERT does not need to be repeatedly invoked for subsequent predictions, resulting in a recurring inference latency of only 5.87 ms/query. On a Tesla T4 GPU, the complete pipeline requires 449.52 MiB of peak allocated GPU memory.

Overall, the additional computational overhead introduced by semantic representation is primarily confined to one-time patient initialization and does not impose a persistent burden on online inference. The inference latency of 5.87 ms/query is substantially shorter than the minute-level CGM update interval, indicating that FMGP can meet the computational requirements of real-time glucose forecasting and demonstrating its deployment potential for mobile glucose monitoring.

E. Clinical Reliability and Glycemic-Risk Analysis

1) *Hypoglycemia and Hyperglycemia Prediction*: To evaluate the reliability of FMGP under clinically important glycemic conditions, we categorize glucose levels as hypoglycemia (< 70 mg/dL), target range (70–180 mg/dL), and hyperglycemia (> 180 mg/dL), and evaluate glycemic-state classification performance and range-specific prediction errors. As shown in Table IX, FMGP achieves Macro-F1 scores of 82.44% and 71.69% and balanced accuracies of 81.13% and 70.97% on CGMacros and OhioT1DM, respectively. More importantly, the class-specific recall shows that FMGP effectively identifies clinically relevant glycemic states, with particularly strong performance in the target range on both datasets and in hyperglycemia detection on OhioT1DM. The corresponding F1-scores further demonstrate consistent classification performance across glycemic states. Notably, the

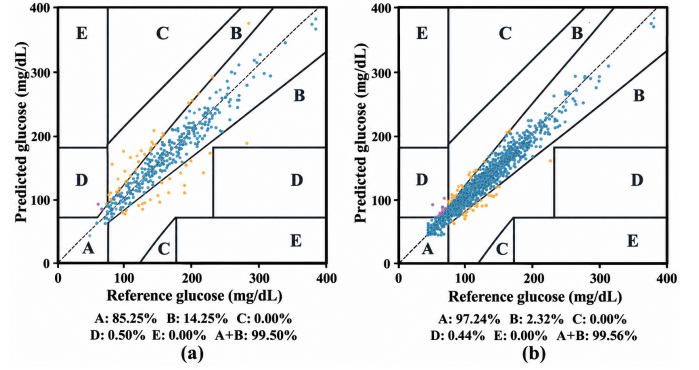


Fig. 2: Clarke Error Grid Result of CGMacros and OhioT1DM (an example result form one of the 50 repeated rounds).

lower hypoglycemia recall and F1-score on OhioT1DM are associated with the limited number of hypoglycemic samples (314), indicating that hypoglycemia detection remains challenging under data-scarce conditions. The RMSEs across the three glycemic ranges are 8.29, 10.14, and 21.38 mg/dL on CGMacros and 18.84, 18.92, and 24.36 mg/dL on OhioT1DM, respectively. Both datasets exhibit higher RMSE in the hyperglycemic range, indicating greater difficulty in continuous-value prediction at high glucose levels. Overall, recall and F1-score characterize the ability to identify different glycemic states, while range-specific RMSE evaluates continuous prediction accuracy across clinically important glucose ranges.

2) *Clarke Error Grid Analysis*: To further assess the clinical risk associated with prediction errors, we perform Clarke Error Grid analysis. Fig. 2 presents an example from the 50 repeated patient-level evaluation rounds, while Table X summarizes the mean proportions of predictions in each zone. On CGMacros and OhioT1DM, 96.288% and 94.190% of predictions fall within Zones A+B, respectively. In particular, the proportions in Zone A reach 93.006% and 85.240%, indicating that most predictions show high clinical agreement. The higher-risk Zones C–E account for only 3.712% and 5.810%, respectively, with no Zone E errors observed on either dataset, further demonstrating the clinical safety of FMGP.

Taken together, the glycemic-state classification, range-specific regression, and Clarke Error Grid results demonstrate the predictive reliability of FMGP under clinically important conditions, including hypoglycemia and hyperglycemia, while the low proportions of predictions in Zones C–E further support its clinical safety. Meanwhile, the lower and more variable hypoglycemia F1 on OhioT1DM indicates that detecting relatively rare hypoglycemic events remains an area for further improvement.

V. CONCLUSION

This work presents FMGP, a semantic-guided personalization framework for cold-start continuous glucose forecasting. FMGP constructs patient semantic representations from patient profiles and uses them to generate personalized modulation and initialization parameters before few-shot adaptation, allowing patient-specific prior information to directly guide personalization for previously unseen patients. Simulation on

TABLE IX: Clinical Performance and Range-Specific Regression Analysis of 30-min Glucose Forecasts ($K = 5$)

Dataset	Overall Classification		Recall (%)			F1-score (%)			Range-Specific RMSE (mg/dL)		
	Macro-F1	Balanced Acc.									
			< 70	70–180	> 180	< 70	70–180	> 180	< 70	70–180	> 180
CGMacros	82.56 ± 6.65	80.93 ± 7.36	72.53 ± 14.29	97.35 ± 1.46	72.90 ± 17.81	75.45 ± 11.30	96.63 ± 1.47	75.59 ± 17.00	8.29 ± 1.80	10.14 ± 0.78	21.38 ± 5.44
OhioT1DM	71.69 ± 8.55	70.96 ± 8.63	36.77 ± 26.70	92.03 ± 2.72	84.09 ± 5.54	39.13 ± 26.10	90.65 ± 2.06	85.30 ± 3.50	18.84 ± 9.79	18.92 ± 2.33	24.36 ± 3.27

Note: Results are based on 50 repeated random subject-wise splits. Values reported with $\pm$ represent mean $\pm$ standard deviation.

TABLE X: Mean Clarke Error Grid Zone Distribution on CGMacros and OhioT1DM (30-min Forecasting, $K = 5$)

Dataset	Zone A (%)	Zone B (%)	Zone C (%)	Zone D (%)	Zone E (%)
CGMacros	93.006	3.282	2.254	1.458	0.000
OhioT1DM	85.240	8.950	3.490	2.320	0.000

CGMacros, OhioT1DM, and the independent DiaTrend dataset demonstrate strong and stable performance for 30-minute glucose forecasting. Repeated patient-level evaluations and paired patient-level comparisons further confirm the consistency of the performance improvements, while ablation studies show that multiple categories of patient information and their semantic encoding contribute to personalized forecasting. Support-set sensitivity analysis further indicates that FMGP achieves performance comparable to larger support sets using only $K = 5$ target-patient samples, highlighting its effectiveness under limited patient-specific data. Clinical analyses show that FMGP can effectively identify clinically important glycemic states, including hypoglycemia and hyperglycemia. On CGMacros and OhioT1DM, 96.288% and 94.190% of predictions fall within Clarke Zones A+B, respectively, with no Zone E errors observed on either dataset. In addition, the cached recurring inference latency is only 5.87 ms/query, indicating that semantic personalization introduces no substantial recurring computational burden for online forecasting. Overall, these results demonstrate that patient semantic information can serve as an effective prior for personalization before adaptation, supporting cold-start glucose forecasting under limited patient data. Future work will explore multimodal patient representations, improve prediction of rare high-risk glycemic events, and evaluate the framework on real-world mobile and wearable devices.

REFERENCES

- [1] M. Yang *et al.* Application of artificial intelligence to measure and predict patient values and preferences: a scoping review. *npj Digital Medicine*, 8(1):769, 2025.
- [2] C. Xie *et al.* Sage: Subject-adaptive graph-based modeling with decision-level calibration for stress and cognitive workload monitoring. *IEEE Journal of Biomedical and Health Informatics*, 2026.
- [3] J. Kesavadev *et al.* Optimizing continuous glucose monitoring adoption in india: From current challenges to future solutions. *Diabetes Therapy*, pp. 1–16, 2026.
- [4] W. Guo *et al.* Denoising of photoplethysmograms for non-invasive blood glucose estimation via slant transform based bit plane method. In *2023 IEEE International Conference on Industrial Technology (ICIT)*, pp. 1–4. IEEE, 2023.
- [5] A. Aththanagoda *et al.* Precision and personalization: How large language models redefining diagnostic accuracy in personalized medicine—a systematic literature review. *IEEE Journal of Biomedical and Health Informatics*, 2025.
- [6] Z. Li and H. Zhang. Parallel multi-scale convolution based prototypical network for few-shot ecg beats classification. In *2022 IEEE-EMBS International Conference on Biomedical and Health Informatics (BHI)*, pp. 1–4. IEEE, 2022.
- [7] H. Ying *et al.* Domain adaptation and generalization using foundation models in healthcare imaging. Available at SSRN 5345726, 2025.
- [8] F. Ahmad and R. AlGhamdi. Personalized multi-agent reinforcement learning framework for adaptive chronic disease therapy management. *Scientific Reports*, 2026.
- [9] Y. Kuang and M. Xie. Drugdoctor: enhancing drug recommendation in cold-start scenario via visit-level representation learning and training. *Briefings in Bioinformatics*, 25(6):bbae464, 2024.
- [10] M. I. Hossain *et al.* Explainable ai for medical data: current methods, limitations, and future directions. *ACM Computing Surveys*, 57(6):1–46, 2025.
- [11] Y. Li *et al.* A comparative study of pretrained language models for long clinical text. *Journal of the American Medical Informatics Association*, 30(2):340–347, 2023.
- [12] Z. Lotfi *et al.* Artificial intelligence-driven approaches for prediction, management, and complication risk in type 2 diabetes: A systematic review. *InfoScience Trends*, 2(6):1–17, 2025.
- [13] T. Zhu *et al.* Population-specific glucose prediction in diabetes care with transformer-based deep learning on the edge. *IEEE transactions on biomedical circuits and systems*, 18(2):236–246, 2024.
- [14] S. Gowriswari and S. Brindha. A hybrid temporal convolutional and long short-term memory network for accurate and interpretable glycemia prediction. *Journal of Electrical Engineering & Technology*, pp. 1–14, 2026.
- [15] C. Piao *et al.* Privacy preserved blood glucose level cross-prediction: an asynchronous decentralized federated learning approach. *IEEE Journal of Biomedical and Health Informatics*, 2025.
- [16] M. H. Lim *et al.* A deep learning framework for virtual continuous glucose monitoring and glucose prediction based on life-log data. *Scientific Reports*, 15(1):16290, 2025.
- [17] V. P. Dwivedi *et al.* Representation learning of structured data for medical foundation models. *arXiv preprint arXiv:2410.13351*, 2024.
- [18] A. Das *et al.* Cgmacros: a pilot scientific dataset for personalized nutrition and diet monitoring. *Scientific Data*, 12(1):1557, 2025.
- [19] C. Marling and R. Bunescu. The ohio1dm dataset for blood glucose level prediction: Update 2020. In *CEUR workshop proceedings*, volume 2675, pp. 71, 2020.
- [20] T. Prioleau *et al.* Glucose-ml: A collection of longitudinal diabetes datasets for development of robust ai solutions. *arXiv preprint arXiv:2507.14077*, 2025.
- [21] Z. Wang and C. Toumazou. Aegis-es: Hazard-adaptive conformal event-sparse inference for glucose monitoring. In *2026 IEEE International Symposium on Circuits and Systems (ISCAS)*, pp. 4563–4567. IEEE, 2026.
- [22] R. Giancotti *et al.* Forecasting glucose values for patients with type 1 diabetes using heart rate data. *Computer Methods and Programs in Biomedicine*, 257:108438, 2024.
- [23] K. Li *et al.* Glunet: A deep learning framework for accurate glucose forecasting. *IEEE journal of biomedical and health informatics*, 24(2):414–423, 2019.
- [24] K. Moon *et al.* Personalized blood glucose prediction in type 1 diabetes using meta-learning with bidirectional long short term memory-transformer hybrid model. *Scientific Reports*, 15(1):30636, 2025.
- [25] A. Mahmoudi *et al.* Diabllm: an llm-based framework for blood glucose prediction in type 1 diabetes. *IEEE Journal of Biomedical and Health Informatics*, 2026.
- [26] T. Prioleau *et al.* Diatrend: A dataset from advanced diabetes technology to enable development of novel analytic solutions. *Scientific Data*, 10(1):556, 2023.
- [27] B. Zhu and C. Connor. Sugarnet: Personalized blood glucose forecast with a fourier kolmogorov–arnold network. *Expert Systems with Applications*, 280:127476, 2025.
- [28] S. J. Chu *et al.* Blood glucose level prediction in type 1 diabetes using machine learning. *arXiv preprint arXiv:2502.00065*, 2025.
- [29] B. Lu *et al.* Deep learning for blood glucose prediction: Reproducibility challenges and factors affecting differential performance. *PLOS Digital Health*, 5(9):e0001633, 2026.