Integrated Glycemic Phenotypes for Glucose Prediction: A Comparative Study Using Hidden Markov Models and Machine Learning

Arumugam P[1], Sornalatha M E[2], Pradeep Nijanthan P[3]
[1]Professor and Head, Department of Statistics, Manonmaniam Sundaranar University, Abishekapatti, Tirunelveli - 627 012, Tamil Nadu, India
Email: sixfacemsu@msuniv.ac.in
[2]Research Scholar, Department of Statistics, Manonmaniam Sundaranar University, Abishekapatti, Tirunelveli - 627 012, Tamil Nadu, India
Email: swarnalathapramod@gmail.com
[3]Research Scholar, Department of Statistics, Manonmaniam Sundaranar University, Abishekapatti, Tirunelveli - 627 012, Tamil Nadu, India
Email: pradeepnijanthanp@gmail.com
Received: 25th Aug, 2026 | Revised: 1st Sep, 2026 | Accepted: 5th Sep, 2026 | Available Online: 7th Sep, 2026

ABSTRACT

INTRODUCTION: Inter-individual variation of glucose dynamics confounds generic "one-size-fits-all" forecasting models of continuous glucose monitoring data. We developed a phenotype-stratified evaluation framework to characterize a model's performance and its complementarity for short-term glucose prediction across different patient profiles.

METHODS: To analyze the data from CGMs of 44 adults. Principal Component Analysis and Fuzzy C-means clustering methods were applied and resulted in 3 latent glycemic phenotypes. Inside each cluster, we built a 4-state Gaussian hidden Markov model and an autoregressive eXtreme Gradient Boosting model. Models were validated on held-out data at 15, 30, 60, and 120 min horizon by Mean Absolute Error, Root Mean Square Error, and Mean Absolute Relative Difference.

RESULTS: Three clinically meaningful glycemic phenotypes were identified from the PCA and Fuzzy C-Means. XGBoost, for the most part, had the lower average error metrics compared to the HMM. However, a large quantity of complementarity was seen with neither model being always the winner with phenotype and horizon-dependent preferences for HMM, XGBoost, or any similar performances. The HMM offers strong and understandable insights into hidden glycemic states and their transitions, enhancing the predictive capabilities of XGBoost.

CONCLUSIONS: Our framework demonstrates great heterogeneity in the performance of CGM forecasting. The presence of a meaningful complementarity between HMM and XGBoost suggests that adaptive and phenotype-informed model selection methods would prove to be more effective than any of the individual global models and can provide a practical template for personalized management for diabetes.

Keywords: continuous glucose monitoring; Hidden Markov model; XGBoost; glycemic phenotypes; predictive modelling; patient stratification; clinical decision support.

How to cite this article: Arumugam P1 Sornalatha M E2 Pradeep Nijanthan P3. Integrated Glycemic Phenotypes for Glucose Prediction: A Comparative Study Using Hidden Markov Models and Machine Learning. Int J Drug Deliv Technol. 2026;16(80s):1688-1698. DOI: 10.25258/ijddt.16.80s.209.

Introduction

The availability of high-resolution data streams has transformed many aspects of healthcare, and diabetes management is one of the leading examples of high-dimensional data. Continuous glucose monitoring (CGM) systems, together with sensors and smart devices, provide dense, time-resolved glucose data that were not available with traditional self-monitoring of blood glucose or occasional laboratory measurements [2,15]. These data streams enable detailed characterization of glycemic variability and open new possibilities for predicting adverse events before they occur [1,3].

Despite this, translating the rich CGM data into meaningful and actionable clinical knowledge remains challenging. Glucose dynamics are nonlinear, noisy and highly heterogeneous across individuals, and conventional approaches often rely on simple summary measures or linear assumptions that do not fully exploit the temporal structure of

CGM data [4]. More advanced analytical methods, including machine learning, deep learning and functional data analysis, have demonstrated improved performance for short- and long-term glucose prediction [11,12]. However, many such models behave as black boxes, offer limited interpretability and frequently ignore the underlying stochastic and physiological state structure of glucose regulation [16,19].

Hidden Markov models (HMMs) are a natural way to capture latent glycemic states and their transitions over time. In an HMM, observed glucose measurements are viewed as emissions from an unobserved sequence of states that can be interpreted as different metabolic or behavioural conditions[8,21]. By explicitly modelling both the emission process and the state-transition process, HMMs are well suited to represent noisy, nonlinear physiological signals and recurrent temporal patterns [9,10]. Recent studies have begun to use HMMs and related methods to define glucotypes and glycemic phenotypes, moving toward individual risk stratification and personalised treatment strategies [14,25].

However, important gaps remain. Many CGM-based prediction methods focus primarily on point forecasts of future glucose values, without explicitly modelling latent glycemic states that precede clinically relevant events such as hypoglycaemia, sustained hyperglycemia or large excursions from target ranges [5,6]. In addition, patient heterogeneity is often underutilised. Large cohorts are treated as a single, heterogeneous population or personalised only through superficial parameter tuning, rather than by learning phenotype-specific state dynamics [7]. Clinical guidelines increasingly emphasise the use of CGM-derived metrics such as time in range to support patient-centred glycemic management [17,20], but there remains a need for modelling frameworks that link these metrics to underlying latent states and individual trajectories.

The key objectives of this work are:

1. **To derive data-driven glycemic phenotypes** from CGM data using principal component analysis and fuzzy C-means clustering, summarizing inter-individual differences in glycemic exposure, variability, and hypo/hyperglycemia burden.
2. **To build cluster-specific forecasting models** by fitting, within each glycemic phenotype, a four-state Gaussian Hidden Markov Model and an autoregressive XGBoost model for short-term glucose prediction at multiple horizons (15, 30, 60, and 120 minutes).
3. **To perform a rigorous, phenotype-stratified comparison** of HMM and XGBoost performance using mean absolute error, root mean square error, and mean absolute relative difference, thereby quantifying how relative performance depends on prediction horizon and glycemic phenotype.
4. **To characterize model complementarity** by identifying, at the patient and phenotype levels, settings in which HMM, XGBoost, or both provide comparable performance, and to assess whether adaptive, phenotype-informed model selection could outperform any single global model.
5. **To interpret the derived phenotypes and model behaviors clinically**, relating each cluster's glycemic profile (time in range, hypo- and hyperglycemia) to forecasting performance, and discussing implications for personalized decision support in diabetes management.

Rather than proposing a new algorithm, our contribution is a phenotype-stratified, head-to-head comparison of HMM and XGBoost for CGM forecasting, highlighting when and for whom each model class is preferable. To our knowledge, this is the first study to evaluate HMMs and gradient-boosted trees side by side for CGM-based glucose forecasting across distinct glycemic phenotypes and multiple prediction horizons.

1. Literature Review

High-resolution CGM data provide dense, longitudinal records of glucose dynamics that support a range of analytical tasks, from descriptive characterisation of glycemic control to short- and long-term prediction of future glucose values [1–3,11,12]. In parallel, expert consensus and clinical practice guidelines have established CGM-derived metrics, including time in range and related measures, as core tools for assessing glycemic control and guiding patient-centred diabetes management [15,17,20,22–24]. Early analytical work primarily used simple summary metrics or linear time-series models, which can miss important nonlinearities and temporal dependencies in CGM data [4]. These approaches, while interpretable and easy to implement, may fail to capture abrupt excursions, recurrent patterns and complex daily profiles that are increasingly recognised as clinically relevant.

More recently, machine learning and deep learning approaches have been applied to CGM-based prediction, showing improved accuracy for both near-term and longer-horizon forecasts in various diabetes populations [4–6,11,12]. These methods include autoregressive integrated moving average (ARIMA) and long short-term memory (LSTM) models for glucose prediction [4], as well as a range of nonlinear and ensemble approaches benchmarked for

hypoglycaemia forecasting and general glucose dynamics [6]. Additional work has explored explainable clustering and classification strategies that use CGM data and related signals to predict postprandial events and support insulin dose optimisation, highlighting the potential of data-driven models for personalised decision support [3,11]. Despite these advances, many models operate as black boxes and provide limited insight into the underlying physiological or behavioural states that drive observed glucose trajectories.

Hidden Markov models (HMMs) have emerged as a promising alternative for modelling CGM time series and other biomedical signals in a more interpretable way. In their classical formulation, HMMs posit that the observed series arises from an unobserved sequence of latent states, with separate components governing the emission of observations from states and the transitions between states over time [16]. This framework has been used extensively in disease mapping, progression modelling and other longitudinal health applications [16,19,21,25]. Within diabetes research, HMMs and related regime-switching models have been deployed to describe glucose dynamics, capture recurrent patterns and model transitions between distinct metabolic regimes, often in the context of diabetes or mixed populations [8,9,14,19,21]. For example, regime-switching and multilevel HMMs have been used to represent heterogeneous daily profiles, inter-individual variability and context-dependent changes in glucose control [7,9,14,21,25].

Beyond diabetes, HMM-based approaches have been successfully applied to model cancer recurrence from gene expression trajectories [19], to infer multimorbidity patterns and their progression in elderly populations [25], and to represent complex longitudinal phenotypes in other chronic conditions [16]. These studies underscore the flexibility of HMMs for capturing latent state structures and transitions across diverse biomedical domains. However, relatively few works have integrated HMMs with explicit glycemic phenotyping or with guideline-relevant CGM metrics such as time in range, time below range and time above range [15,22–24]. Similarly, although phenotype-based risk stratification and state-based representations are conceptually aligned with contemporary clinical guidelines [17,18,20], most CGM prediction models still focus on point forecasts of future glucose values rather than on state trajectories that may better reflect underlying pathophysiology and patient heterogeneity [5,6,11,12].

Taken together, the existing literature highlights both the potential and the limitations of current CGM analytics. Advanced machine learning and deep learning models have improved predictive performance, but often at the cost of interpretability and explicit state representation [4–6,11,12]. HMMs and related latent-state models offer a principled way to represent glycemic states and transitions, yet they have not been fully integrated with glycemic phenotyping frameworks or with clinically endorsed CGM metrics [3,7,14–16,20–25]. This motivates the development of approaches that combine phenotype-specific HMMs with modern CGM metrics, with the goal of producing models that are both predictive and clinically interpretable.

2. Methods and Materials

This study was undertaken using secondary continuous glucose monitoring data [13]. obtained from a previously described cohort of 44 people with 12+ continuous glucose monitoring days. Data was averaged to 15-minute periods with missing periods of < 15 min filled in by interpolation (within the same day) or carry-forward/backward filling. Patient-level features of glycemic data and events were extracted from these cleaned and regularly sampled CGM time series. A related approach uses Markov models to infer types of diabetes from diabetes patient data under free-living conditions [14].

2.1. Patient Stratification Glycemic Phenotyping Using PCA and Fuzzy C-Means

To understand how heterogeneity exists in continuous glucose monitoring profiles among patients with diabetes, we then used a multi-step phenotyping approach. This involved characterization of the glycemic profile of each patient by extracting an extensive set of glycemic and event-based features of the patient-level, cleaned, regularly sampled CGM time series. These features carefully captured different aspects of chronic glucose exposure, glycemic variability and burden of hypo- and hyperglycemia. In particular, the following categories of features were developed:

- Time in glycemic ranges: Percentages of time in clinically relevant ranges.
- Circadian Patterns: Features representing the dynamics of glucose at different time points, e.g. mean and standard deviation of nocturnal glucose, as well as separate the mean glucose for daytime and nighttime periods.
- Extremes and Sampling Characteristics: Important Statistical Extremes (max, min), Total number of CGM measurements, etc. Dysglycemia and Treatment-Related Events Measures of hypoglycemic and hyperglycemic burden with measures of count and duration of events, and episode-based measures.

- Insulin and Dietary Event Logs: Counts of various events for insulin administration and dietary events, where available

All features at the patient level that vary continuously were standardized (mean 0, unit variance) to ensure similar contributions for variables of different scales.

The standardized feature data were then subjected to Principal Component Analysis, yielding a lower-dimensional representation that captures the main axes of glycemic variation, and multicollinearity was dealt with. The top three key factors were chosen in accordance with explained variance and clinical interpretability in that PC1 included the chronic glycemic exposure and hyperglycemic burden; PC2 included the hypoglycemic burden; and PC3 included the glycemic excursions and variability. These components have developed a clinically relevant orthogonal coordinate system of downstream phenotyping.

Then, the principal component scores were clustered using Fuzzy C-Means to determine latent glycemic phenotypes. The FCM algorithm was optimized to three clusters, which was in line with valid methodological past and cluster validation measures. FCM allows partial memberships and irregular profiles by giving fuzzy membership probabilities to each patient in each cluster; to do descriptive analyses, the most common phenotype was given as the maximum membership. The combined PCA-FCM pipeline has been embraced to interpret inter-patient heterogeneity, which produces interpretable glycemic phenotypes to stratify forecasting performance evaluations.

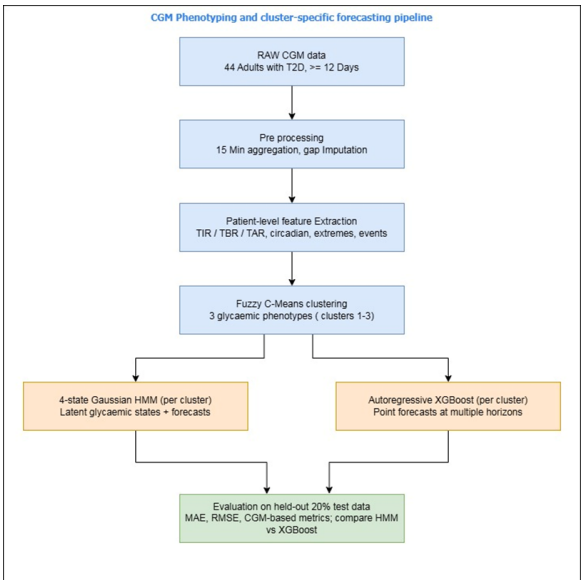

Figure 1. Overview of the continuous glucose monitoring (CGM) phenotyping and forecasting pipeline.

As outlined in Figure 1, after cluster assignment, we trained two model families per cluster: a four-state Gaussian HMM and an autoregressive XGBoost model.

2.2. Cluster-Specific Forecasting Model Development

A univariate Gaussian Hidden Markov Model was estimated on the CGM time series at 15 minutes per cluster. The HMM had a dual role that it offered a generative model of latent glycemic regimes that are interpretable, and also offered a structured predictive process.

In each cluster, the observed CGM sequence O_t was assumed to be generated by an unobserved, finite-state Markov chain S_t. We specified four latent states (N=4) for the HMM within each cluster, chosen to balance model parsimony with the capacity to represent qualitatively distinct glycemic regimes.

Explanation of terms:

A: This is the 4×4 transition probability matrix.

a_{ij}: An element of the transition matrix, representing the probability of moving from state $S_{t-1}=i$ to state $S_t=j$ at time t.

$P(S_t=j|S_{t-1}=i)$: The conditional probability of being in state j at time t, given that the system was in state i at time t-1.

$b_j(O_t)$: The emission probability for state j , representing the likelihood of observing glucose value O_t when the system is in state j.

$P(O_t|S_t=j)$: The conditional probability of observing O_t given that the system is in state j.

$N(\mu_j, \sigma_j^2)$: Denotes a Normal distribution for state j, with μ_j as the mean glucose level and σ_j^2 as the variance of glucose levels within that state

Each cluster state had an initial state distribution, transition matrix, and state-specific Gaussian parameters, which were estimated on the 80% training segment using the expectation-maximization algorithm. A four-state latent specification had a balanced statistical stability and interpretability. One-step-ahead and multi-step-ahead forecasts were obtained by propagating smoothed state probabilities and projecting onto emission means, interpreting forecasts as expected state-conditional mean glucose levels.

An autoregressive eXtreme Gradient Boosting model was independently fitted for each cluster to replace the comparator as a flexible and data-driven comparator. For each patient, the 15-minute CGM series was converted to examples of supervised learning with a sliding window method. Predictor vectors consisted of past CGM values, which may be supplemented with features derived from them. The prediction targets were the future value of glucose at certain horizons. Separate XGBoost trees were trained for each combination of cluster and horizon based on the squared error loss function, hyperparameter optimization and early stopping.

These models gave point forecasts and therefore served as a robust, non-parametric benchmark.

In sum, the cluster-specific forecasting framework included two distinct model families: a four-state univariate Gaussian HMM, emphasized for its interpretability, and an autoregressive XGBoost model, which was valued for its predictive flexibility. Both were independently fitted per cluster using an identical time train-test protocol. All subsequent forecast accuracy and complementarity comparisons were made using the held-out 20% test segments.

2.3. Model Evaluation and Performance Metrics

Forecasting models were evaluated at four prediction horizons (15, 30, 60, 120 minutes), corresponding to 1-, 2-, 4-, and 8-step-ahead forecasts for 15-minute CGM data.

Forecast accuracy was quantified using three error measures:

Mean Absolute Error: $MAE=\frac{1}{n}\sum_{i=1}^{n}|y_i-\hat{y}_i|$

Root Mean Square Error: $RMSE=\sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i-\hat{y}_i)^2}$

Mean Absolute Relative Difference: $MARD=\frac{100}{n}\sum_{i=1}^{n}\frac{|y_i-\hat{y}_i|}{|y_i|}$

All metrics were computed on the held-out 20% test data for every model, within each glycemic cluster and prediction horizon, establishing a consistent basis for comparative assessment.

2.4. Complementarity Assessment

Beyond aggregate performance metrics, we quantified the complementarity between the Hidden Markov Model and the autoregressive eXtreme Gradient Boosting model at two levels: point-wise and patient-level. This aimed to demonstrate their distinct strengths for superior overall forecasting.

For point-wise comparison, paired predictions for HMM and XGBoost were obtained for each glycemic cluster and prediction horizon. Absolute prediction errors (e_i^{HMM}, e_i^{XGB}) were computed for each test time point i. Points were classified as "HMM better" ($e_i^{HMM}+\delta<e_i^{XGB}$), "XGBoost better" ($e_i^{XGB}+\delta<e_i^{HMM}$), or "Similar performance" ($|e_i^{HMM}-e_i^{XGB}| \leq \delta$), using a small tolerance δ (e.g., 2-5 mg/dL). This analysis revealed that neither model consistently dominated, with substantial temporal regions where each outperformed the other, supporting complementarity.

Patient-level preference was assessed by computing patient-specific Mean Absolute Errors ($MAE_{p,h}^{HMM}$ and $MAE_{p,h}^{XGB}$) for each patient and prediction horizon. A small positive threshold (e.g., 1-2 mg/dL) on the MAE difference defined patient-level categories: "XGB-friendly" ($\Delta MAE_{p,h}>\epsilon$), "HMM-friendly" ($\Delta MAE_{p,h}<-\epsilon$), or "Similar" ($-\epsilon \leq \Delta MAE_{p,h} \leq \epsilon$). Consistent representation across clusters and horizons demonstrated that optimal modeling approaches could vary even within the same phenotype.

Collectively, findings from both analyses underscore a robust principle of model complementarity: no single global model is uniformly optimal. The HMM excels in capturing latent state dynamics and providing interpretability, while XGBoost offers superior predictive flexibility. This strong evidence suggests that adaptive or personalized model selection strategies, informed by patient phenotype and clinical context, would yield more effective forecasting outcomes than relying on any solitary modeling paradigm.

3. Results

3.1. Cohort characteristics

A total of 44 adults with type 2 diabetes were included in the CGM phenotyping analysis.

Baseline clinical characteristics are summarized in Table 1.

Table 1. Baseline clinical characteristics of the continuous glucose monitoring cohort

Characteristic	Value
Number of participants	44
Female, n (%)	18 (40.9)
Male, n (%)	26 (59.1)
Age, years	63.3 ± 12.3
BMI, kg/m^2	23.6 ± 2.4
Duration of diabetes, years	9.2 ± 9.3
HbA1c, mmol/mol[a]	62.1 ± 19.5
Glycated albumin,%[b]	19.5 ± 6.8

[a]HbA1c available for 36 participants.

[b] Glycated albumin available for 24 participants.

Values are mean ± standard deviation unless otherwise indicated

3.2. Baseline glycemic markers and CGM-derived metrics

In this cohort, mean HbA1c was 62.1 ± 19.5 mmol/mol (n = 36) and mean glycated albumin was 19.5 ± 6.8% (n = 24). Both long-term laboratory markers were strongly associated with CGM-derived glycemic burden. Higher HbA1c correlated with higher mean CGM glucose (Pearson r = 0.69, p = 3.1 × 10−6) and lower percentage time in range 70–180 mg/dL (r = −0.75, p = 1.2 × 10−7). Glycated albumin showed a similar pattern, correlating positively with mean CGM glucose (r = 0.64, p = 7.1 × 10−4) and inversely with time in range (r = −0.67, p = 3.8 × 10−4). These findings support the consistency between conventional intermediate-term glycemic markers and CGM-derived metrics in this population and provide a clinical context for the CGM-based phenotypes and latent state analyses described below.

3.3. Glycemic Phenotypes

The initial step in characterizing glycemic heterogeneity involved dimensionality reduction via Principal Component Analysis followed by Fuzzy C-

Means clustering. The resulting three distinct glycemic phenotypes are visually represented and characterized as follows:

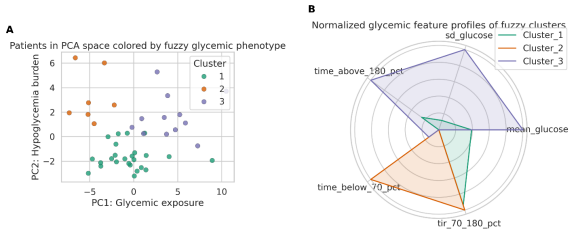

Figure 2: Characterization of Glycemic Phenotypes via PCA and Fuzzy C-Means Clustering.
(A) Patients in PCA space colored by fuzzy glycemic phenotype. (B) Normalizedglycemic feature profiles of fuzzy clusters.

Figure 2A illustrates patient distribution in the PC1-PC2 plane, with coloring indicating maximum-membership cluster assignment. This shows the three clusters occupying distinct, though partially overlapping, regions. Cluster 1 exhibits moderately low PC1 and PC2 scores, indicating moderate glycemic exposure with minimal hypoglycemia. Cluster 2 is characterized by lower PC1 values but higher PC2, suggesting lower mean glucose but increased hypoglycemia burden. In contrast, Cluster 3 is situated towards higher PC1 and moderately elevated PC2, signifying higher mean glucose, greater time above 180 mg/dL, and some hypoglycemia. This spatial separation visually confirms effective capture of glycemic exposure and hypoglycemia burden by the fuzzy clusters.

Figure 2B further characterizes these phenotypes by displaying normalized profiles of key glycemic measures (mean glucose, glucose standard deviation, time in range, time above range, and time below range) for each cluster's centroid. This highlights clear and clinically interpretable differences:

- Cluster 1: Shows low mean glucose and variability, high time in range, and minimal hypo- or hyperglycemia.
- Cluster 2: Characterized by lower mean glucose but a pronounced increase in time below 70 mg/dL.
- Cluster 3 (Hyperglycemic –High-Variability Phenotype): Exhibits elevated mean glucose, higher glycemic variability, and a large proportion of time above 180 mg/dL.

Collectively, Figure 2 demonstrates that the three fuzzy C-means clusters are well-separated in the principal component space and readily interpretable on the original continuous glucose monitoring scale, fulfilling the objective of defining statistically coherent and clinically meaningful glycemic phenotypes.

3.4. Cluster-Specific HMM Latent State Characterization

For each identified glycemic cluster, a 4-state Gaussian Hidden Markov Model was fitted to characterize distinct latent glycemic regimes. Characteristics of Latent Glycemic States Identified by the Hidden Markov Model in Cluster 1, Cluster2 and Cluster3.The four-state Gaussian HMM yielded clinically interpretable latent glycemic states within each phenotype cluster

Table 2: Latent glycemic states identified by the four-state Gaussian HMM within each glycemic phenotype.

State	Mean (mg/dL)	SD (mg/dL)	CV (%)	Q1	Median	Q3	% <70	% 70--180	% >180	N (Points)
Cluster 1: well-controlled phenotype										
S1	92.5	10.3	1110.0	86.4	93.6	100.8	2.3	97.70		8878.0
S2	119.2	8.2	690.0	113.4	118.8	126.0	0.0	100.0	0.0	9158.0
S0	148.7	11.1	750.0	140.4	147.6	156.6	0.0	99.90	0.10	9027.0
S3	200.6	28.2	1410.0	180.0	194.4	214.2	0.0	27.10	72.90	6075.0
Cluster 2: hypoglycemia-prone phenotype										
3	72.0	9.6	1340.0	68.4	73.8	79.2	29.60	70.40	0.0	2196.0
0	92.5	6.4	690.0	86.9	91.8	97.2	0.0	100	0.0	2826.0
2	115.6	9.0	780.0	108.0	115.2	122.4	0.0	100	0.0	2606.0
1	158.8	23.9	1500.0	140.4	153.0	169.2	0.0	83.70	16.30	1604.0
Cluster 3: hyperglycemic–high-variability phenotype										
3	95.8	14.9	1550	86.4	97.2	108	5.7	94.30	0	3929
0	136	13.2	970	124.2	136.8	147.6	0	100	0	4000
2	183.9	17	930	169.2	183.6	196.2	0	45.90	54.10	4507
1	257.7	36	1400	230	250.2	277.2	0	0	100	3164

In the well-controlled cluster, states were dominated by normoglycemic and in-range periods with limited hyperglycemia. In the hypoglycaemia-prone cluster, one state concentrated substantial time below 70 mg/dL, whereas others remained largely within range. In the hyperglycemic–high-variability cluster, two states captured persistent and severe hyperglycemia

(up to 100% of time > 180 mg/dL), highlighting a markedly different risk profile despite using the same four-state HMM structure.

Table 3: Dwell times (minutes) per HMM state and cluster

State	Median	Q1	Q3	Number of Runs
Cluster 1: well-controlled phenotype				
S0	75	30	135	1329
S1	135	75	270	660
S2	60	30	120	1348
S3	105	60	153	652
Cluster 2: hypoglycemia-prone phenotype				
S0	75	45	135	402
S1	105	71	150	216
S2	60	30	120	442
S3	120	60	225	206
Cluster 3: hyperglycemic–high-variability phenotype				
S0	75	45	135	551
S1	105	75	206	290
S2	75	45	150	576
S3	120	75	225	346

From Table[3], Dwell times highlight distinct temporal dynamics across phenotypes.

- Cluster 1 (well-controlled): Typically 1-2 hours, with S1 (median 135 min) indicating longer sustained dysglycemic states despite overall good control.
- Cluster 2 (hypoglycemia -prone): Transient normoglycemic states (S0/S2 median 60-75 min) contrasted with more persistent S1/S3 (median 105-120 min) during unstable periods.
- Cluster 3 (hyperglycemic –high-variability): Also 1-2 hours, with prolonged S3 (median 120 min) consistent with sustained hyperglycemia or instability

Figure 3:HMM State Transition Graphs for Glycemic Phenotypes.(A) Cluster 1 HMM state transition graph (well-controlled phenotype). (B) Cluster 2 HMM state transition graph (hypoglycemia-prone phenotype). (C) Cluster 3 HMM state transition graph(hyperglycemic–high-variability phenotype).

The dynamic transitions between latent glycemic states are visualized in Figure 3, where each node represents a state, and arrows indicate transition probabilities. Arrow thickness and numerical labels signify the likelihood of moving between regimes, including self-loops reflecting state persistence.

Cluster 1 (well-controlled phenotype): As shown in Figure[3]A, this cluster exhibits prominent self-loops for all states, indicating high persistence in the current glycemic state. Euglycemic states (S0 and S1) demonstrate particularly strong stability. The graph also illustrates dynamic, albeit less frequent, pathways towards states of mild hyperglycemia, such as transitioning from S0 to S2.

Cluster 2 (hypoglycemia -prone phenotype): Figure[3]B highlights critical transitions related to lower glucose levels. While self-loops are strong, persistence within the hypoglycemia-prone State S3 is notable. The visual representation clearly illustrates the fluctuating glycemic control and the likelihood of entering or remaining in vulnerable states within this population.

Cluster 3 (hyperglycemic –high-variability phenotype): In Figure[3]C, the strong self-loop for State S1 (severe hyperglycemia) underscores its high persistence once entered. The graph also visually reinforces the dynamic drift towards elevated glucose levels, showing probabilities of moving from more controlled states (e.g., S0) towards hyperglycemic states, characteristic of this phenotype's persistent and recurring hyperglycemia

.

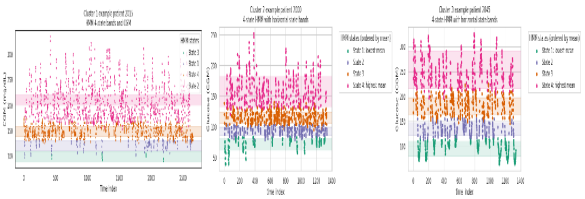

Figure 4. Cluster-specific HMM latent state dynamics. (A) Cluster 1 example patient; (B) Cluster 2; (C) Cluster 3. CGM trajectories are colored by inferred 4-state HMM latent states, with horizontal bands indicating typical glucose ranges for each state.

Figure[4] shows the example CGM trajectories with overlaid HMM states for one representative patient from each latent glycemic phenotype. Each panel shows CGM measurements colored by the inferred four-state Gaussian HMM state, visually demonstrating how patients within each phenotype move between latent glycemic states over time.

Cluster 1 (well-controlled): The representative patient spends most time in mid-range states with few excursions, indicating stable glycemiccontrol(Figure[4]A).

Cluster 2 (hypoglycemia-prone): The example patient frequently transitions into the lowest state, showing recurrent exposure to low glucose values (Figure[4]B).

Cluster 3 (hyperglycemic –high-variability): This patient spends extended periods in the highest states with large excursions, consistent with sustained

hyperglycemia and high glycemic variability (Figure[4]C).

3.5. Comparative Forecasting Performance

To rigorously assess the short-term forecasting capabilities, we compared the performance of the HMM (using state-mean predictions) and the autoregressive XGBoost model across the three identified glycemic clusters and four prediction horizons (15, 30, 60, and 120 minutes). Forecast accuracy was measured in terms of Mean Absolute Error, Root Mean Square Error, both in mg/dL and Mean Absolute Relative Difference, in percentage.

Table 5: Comparative forecasting performance of HMM (state-mean) and autoregressive XGBoost across clusters and prediction horizons. MAE and RMSE are reported in mg/dL, MARD in percent.

Cluster	Horizon	HMM (State-mean)			XGBoost (AR3)		
		MAE (mg/DL)	RMSE (mg/DL)	MARD (%)	MAE (mg/DL)	RMSE (mg/DL)	MARD (%)
Cluster 1	15 Min	9.92	12.75	9	8.74	12.53	7
	30 Min	12.57	17.19	11	12.51	17.70	11
	60 Min	17.76	25.37	15	18.73	24.99	16
	120 Min	24.89	34.89	21	25.43	31.27	23
Cluster 2	15 Min	11.45	17.81	10	9.64	15.01	9
	30 Min	14.00	21.18	12	13.58	20.91	12
	60 Min	19.30	28.57	17	19.05	28.18	17
	120 Min	27.49	38.65	26	24.10	33.92	22
Cluster 3	15 Min	15.39	19.46	12	12.44	17.14	9
	30 Min	19.39	25.70	14	17.99	24.74	13
	60 Min	28.38	39.17	20	27.74	37.11	20
	120 Min	44.04	59.58	31	40.78	51.87	31

As seen in detail in Table 5, XGBoost tended to produce lower average MAE, RMSE and MARD than the HMM, especially for several horizons and clusters. For example, in Cluster 1 at 15 minutes, XGBoost achieves MAE of 8.74 mg/dL as compared to 9.92 mg/dL for HMM. However, it is also notable that neither model was uniformly dominant, with some cases showing very similar performance or even slight advantages for HMM, e.g. MAE for Cluster 1 at 120 minutes (HMM: 24.89 mg/dL vs. XGBoost: 25.43 mg/dL). This first comparison forms the basis for a more in-depth analysis of the complementarity of the models.

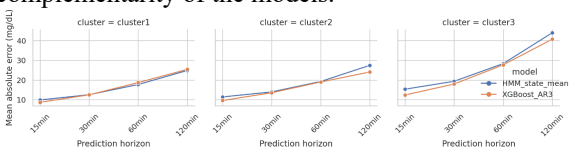

Figure 5: Mean absolute error (MAE) for short-term glucose forecasts across glycemic phenotypes and prediction horizons. Each panel corresponds to one glycemic phenotype (cluster 1–3). Within each phenotype, MAE is shown for the four prediction horizons (15, 30, 60, and 120 minutes) and for two forecasting models: the four-state Gaussian Hidden Markov Model (HMM) and the autoregressive XGBoost model

Figure 5 shows the mean absolute error (MAE) for 15, 30, 60, and 120-minute prediction horizons within each glycemic phenotype, comparing the four-state HMM and the autoregressive XGBoost model. In all three phenotypes, forecast error increases with horizon for both models, with XGBoost generally achieving lower MAE, particularly at longer horizons and in the hyperglycemic/high-variability phenotype. In the well-controlled phenotype, the difference between models is small at the 15-minute horizon, consistent with the similar short-term dynamics captured by both approaches.

3.6. Model Complementarity

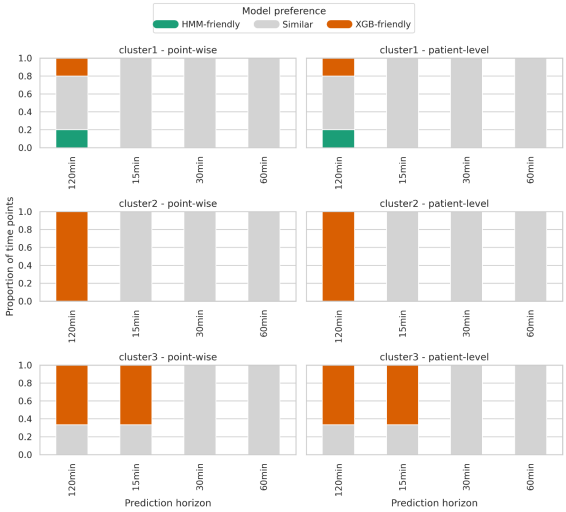

Figure 6: Model Complementarity: Point-wise and Patient-Level Preferences AcrossGlycemic Phenotypes and Prediction Horizons. (A) Cluster 1 - Point-wise complementarity.(B) Cluster 1 - Patient-level preference.(C) Cluster 2 - Point-wise

complementarity. (D)Cluster 2 - Patient-level preference. (E) Cluster 3 - Point-wise omplementarity.(F) Cluster 3 - Patient-level preference

Beyond average performance, we centered our analysis on model complementarity as shown in Figure 6 and represented by the proportion of time points and patients favouring HMM, XGBoost or showing a similar performance across a set of clusters and horizons.

At shorter horizons (15-60min), "Similar" performance was dominating across all clusters, which suggests similar accuracy of models. However, at the 120-minute horizon, clear preferences were demonstrated:

Cluster 1: Demonstrated a significant preference for HMM (around 20% point wise and patient level) with a smaller proportion of favouringXGBoost, indicating the better performance of HMM for some well controlled patterns

Cluster 2: Showed a high preference for XGBoost (almost 100% at 120 minutes) meaning its superiority in this hypoglycemia prone phenotype.

Cluster 3: Also favouredXGBoost (60-70% at 120 minutes), highlighting its robustness for high glucose and variability.

Crucially, consistency between point-wise and patient-level preferences reinforced these findings, underscoring a strong phenotype- and horizon-dependent model utility. This evidence strongly advocates for adaptive or personalized model selection strategies over a single global model.

4. Discussion

This discussion highlights three key aspects of the proposed phenotyping and forecasting framework.

First, we deliberately evaluated the four-state HMM and the XGBoost models on *real* CGM data rather than on simulated trajectories. This choice ensures that both models are exposed to the same artefacts, noise, and missingness patterns that occur in practical use, including irregular sampling, sensor dropouts, and unobserved behavioural changes. A direct consequence is that any imperfections of the underlying data are necessarily inherited by the trained models and by the reported performance metrics. In this sense, the comparison between HMM and XGBoost across the well-controlled, hypoglycemia-prone, and hyperglycemic/high-variability phenotypes should be interpreted as an assessment of how robust each model class is to realistic data quality, rather than an idealised comparison under perfectly observed, noise-free conditions.

Second, despite these imperfections, the inferred HMM structures indicate that the model is able to self-correct to some extent and to regularise away implausible trajectories. Examination of the state-specific emission distributions and transition matrices shows that clinically unlikely transitions—such as abrupt jumps between very low and very high latent glycemic states that would not arise under typical within-patient dynamics—are assigned essentially zero transition probability. Likewise, the learned transition matrices discourage improbable dwell-time patterns. This behaviour is observed across all three glycemic phenotypes and suggests that, even when trained on noisy and partially incomplete CGM data, the HMM tends to favour clinically plausible latent state sequences. This regularisation contributes to stabilising short-horizon forecasts, particularly in the well-controlled phenotype where the HMM and XGBoost often exhibit similar predictive performance.

Finally, our analysis focuses on prediction on an aligned CGM time grid, assuming that the timing of observed points is known, but without explicitly modelling the mechanism that generates missing and irregular observations. In other words, we address only one side of a two-fold problem: given an observed CGM sequence up to a certain horizon, we compare HMM and XGBoost for forecasting future glucose values, but we do not attempt to infer how and why measurements may be missing between two observed points. A logical next step is to extend the framework to explicitly represent and predict missingness, for example, by identifying likely “missing data slots” in future trajectories, quantifying the probability and impact of unobserved glycemic events, and assessing how such events influence model preference across phenotypes and horizons. Incorporating a missingness-aware component into the phenotype-specific HMM/XGBoost comparison developed here would improve the robustness of the conclusions and increase the clinical relevance of the framework, especially in real-world settings where CGM data are often incomplete and irregular.

5. Conclusion

In this study, we integrated glycemic phenotyping, Hidden Markov modeling, and machine learning to investigate short-term glucose prediction in adults with type 2 diabetes. Using CGM-derived features, we identified distinct phenotypes that captured differences in glycemic level, variability, and time in range, and within each phenotype, we fitted a four-state HMM to obtain clinically interpretable latent glycemic regimes. We then compared HMM-based and XGBoost-based predictors across phenotypes, prediction horizons, and patients.

Our findings show that prediction performance is selective rather than uniform. The four-state HMM provided a stable and interpretable representation of glycemic dynamics and performed well at shorter

horizons and in more stable phenotypes. XGBoost tended to outperform in longer horizons and in phenotypes with higher variability and hypoglycemia risk. In the hyperglycemic, highly variable phenotype, neither model was consistently superior, and patient-level preferences mirrored pointwise differences, underscoring the context-dependent nature of forecasting performance.

These results suggest that relying on a single "best" model may be suboptimal for heterogeneous T2DM populations. Instead, model choice and model combination should be informed by glycemic phenotype, prediction horizon, and individual patient characteristics. The proposed framework offers a step toward adaptive, phenotype-aware decision support systems that integrate interpretable latent-state structure with flexible machine learning. Future work should validate these findings in larger and more diverse cohorts, incorporate additional clinical and behavioural covariates, and assess the impact of phenotype-tailored forecasting on treatment decisions and clinical outcomes.

Declaration Of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could influence the work reported in this article.

Acknowledgement

The First author gratefully acknowledges the financial support of Manonmaniam Sundaranar University for the award of M.G. Ramachandran Centenary Fellowship. (MSU/R/RES/R2).

References

[1] Gecili E, Huang R, Khoury J, King E, Altaye M, Bowers K, Szczesniak RD. Functional data analysis and prediction tools for continuous glucose-monitoring studies. *J Clin Transl Sci*. 2020;5. doi:10.1017/cts.2020.545

[2] Aliffi GE, Nastasi G, Romano V, Pitocco D, Rizzi A, Moore EJ, Di Gaetano A. A system of ODEs for representing trends of CGM signals. *J Math Ind*. 2024;14. doi:10.1186/s13362-024-00161-w.

[3] Rehman NU, Contreras I, Beneyto A, Vehí J. Explainable cluster-based learning for prediction of postprandial glycemic events and insulin dose optimization in type 1 diabetes. *PLOS Digit Health*. 2025;4. doi:10.1371/journal.pdig.0000996

[4] Krishnamoorthy U, Karthika V, Mathumitha MK, Panchal H, Jatti VKS, Kumar A. Learned prediction of cholesterol and glucose using ARIMA and LSTM models – a comparison. *Results Control Optim*. 2023;14:100362. doi:10.1016/j.rico.2023.100362

[5] Rahman SA, Mahadi M, Yuliana D, Susilo YKB, Ariffin AE, Amgain K. Artificial intelligence-driven innovations in diabetes care and monitoring. *bioRxiv*. 2025. doi:10.1101/2025.06.02.25328795

[6] Prendin F, De Favero S, Vettoretti M, Sparacino G, Facchinetti A. Forecasting of glucose levels and hypoglycemic events: head-to-head comparison of linear and nonlinear data-driven algorithms based on continuous glucose monitoring data only. *Sensors*. 2021;21:1647. doi:10.3390/s21051647

[7] Rao PT, Narendra S. Two layered hidden Markov model for studying type 2 diabetes. *Int J Syst Assur Eng Manag*. 2024. doi:10.1007/s13198-024-02491-9

[8] Montaser E, Díez J-L, Rossetti P, Rashid M, Çınar A, Bondía J. Seasonal local models for glucose prediction in type 1 diabetes. *IEEE J Biomed Health Inform*. 2019;24:2064–2072. doi:10.1109/JBHI.2019.2956704

[9] Frandes M, Timar B, Timar R, Lungeanu D. Chaotic time series prediction for glucose dynamics in type 1 diabetes mellitus using regime-switching models. *Sci Rep*. 2017;7. doi:10.1038/s41598-017-06478-4

[10] Huard B, Kirkham G. Mathematical modelling of glucose dynamics. *Curr OpinEndocrMetab Res*. 2022;25:100379. doi:10.1016/j.coemr.2022.100379

[11] Giancotti R, Bosoni P, Vizza P, Tradigo G, Gnasso A, Guzzi PH, Bellazzi R, Irace C, Veltri P. Forecasting glucose values for patients with type 1 diabetes using heart rate data. *Comput Methods Programs Biomed*. 2024;257:108438. doi:10.1016/j.cmpb.2024.108438

[12] Ying LP, Yin OX, Quan OW, Jain N, Mayuren J, Pandey M, Gorain B, Candasamy M. Continuous glucose monitoring data for artificial intelligence-based predictive glycemic event: a potential aspect for diabetic care. *Int J Diabetes Dev Ctries*. 2024;45:272–287. doi:10.1007/s13410-024-01349-x

[13] Zhu J. Diabetes datasets – ShanghaiT1DM and ShanghaiT2DM. *figshare* dataset. 2022. doi:10.6084/m9.figshare.20444397.v3.

[14] Carvalho DF, Kaymak U, van Gorp P, van Riel NAW. A Markov model for inferring event types on diabetes patients data. *Healthc Anal*. 2022;2:100024. doi:10.1016/j.health.2022.100024

[15] Elido V, Aguilera E, Cardona-Hernandez R, Díaz-Soto G, Villar NGP de, César MJP, Ampudia-Blasco FJ. Expert recommendations for using time-in-range and other continuous glucose monitoring metrics to achieve patient-centeredglycemic control in people with

diabetes. *J Diabetes Sci Technol*. 2022;17(5):1326–1339. doi:10.1177/19322968221088601

[16] Green PJ, Richardson S. Hidden Markov models and disease mapping. *J Am Stat Assoc*. 2002;97(460):1055–1070. doi:10.1198/016214502388618870

[17] Holt RIG, DeVries JH, Hess-Fischl A, Hirsch IB, Kirkman MS, Klupa T, Ludwig B, Nørgaard K, Pettus J, Renard É, Skyler JS, Snoek FJ, Weinstock RS, Peters AL. The management of type 1 diabetes in adults: a consensus report by the American Diabetes Association (ADA) and the European Association for the Study of Diabetes (EASD). *Diabetes Care*. 2021;44(11):2589–2625. doi:10.2337/dci21-0043

[18] Lee J. Diagnosis and management of pediatric type 1 diabetes mellitus. *J Korean Med Assoc*. 2021;64(6):425–432. doi:10.5124/jkma.2021.64.6.425

[19] Momenzadeh M, Sehhati M, Rabbani H. Using hidden Markov model to predict recurrence of breast cancer based on sequential patterns in gene expression profiles. *J Biomed Inform*. 2020;111:103570. doi:10.1016/j.jbi.2020.103570

[20] Moon JS, Kang S, Choi JH, Lee KA, Moon JH, Chon S, Kim DJ, Kim HJ, Seo JA, Kim MK, Lim JH, Song Y, Yang Y, Kim JH, Lee Y, Noh J, Hur KY, Park JS, Rhee SY, Lee B. 2023 clinical practice guidelines for diabetes management in Korea: full version recommendation of the Korean Diabetes Association. *Diabetes Metab J*. 2024;48(4):546–596. doi:10.4093/dmj.2024.0249

[21] Padi TR, Narendra S. Multilevel hidden Markov models for studying type-1 diabetes. *JP J Biostat*. 2024;24(1):161–184. doi:10.17654/0973514324011

[22] Pleus S, Eichenlaub M, Dabla PK, Diem P, Boija EE, Fokkert M, Hinzmann R, Jendle J, Klonoff DC, Lu J, Makris K, Mohan V, Nichols JH, Pemberton J, Selvin E, Slingerland RJ, Thomas A, Tran NK, Witthauer L, Freckmann G. Clinical assessment and acceptance criteria for continuous glucose monitoring (CGM) system performance: a proposed guideline by the IFCC Working Group on CGM. *Clin Chim Acta*. 2025;580:120728. doi:10.1016/j.cca.2025.120728

[23] Scharf M, Fériz K, Yepes CA, Contreras Á, Lequi L, Sanhueza D, Krakauer M, Márquez E, Ré M, Mehta R. Consensus statement on standardizing CGM evaluation metrics in Latin America: an expert approach. *DiabetolMetab Syndr*. 2025;17(1):291. doi:10.1186/s13098-025-01851-0

[24] Timmons JG, Boyle JG, Petrie JR. Time in range as a research outcome measure. *Diabetes Spectr*. 2021;34(2):133–139. doi:10.2337/ds20-0097

[25] Violán C, Fernández-Bertolín S, Guisado-Clavero M, Foguet-Boreu Q, Valderas JM, Manzano JV, Roso-Llorach A, Cabrera-Bean M. Five-year trajectories of multimorbidity patterns in an elderly Mediterranean population using hidden Markov models. *Sci Rep*. 2020;10(1):6872. doi:10.1038/s41598-020-73231-9.