

Original Paper

Continuous Glucose Monitoring Data Compression Using Peak-Nadir Encoding in Diabetes: Method Development and Evaluation

Clara Bender¹, PhD; Line Davidsen^{2,3}, MD; Søren Schou Olesen^{2,3}, MD, PhD; Simon Cichosz¹, BME, PhD

¹Department of Health Science and Technology, Aalborg University, Gistrup, North Denmark, Denmark

²Department of Clinical Medicine, Aalborg University Hospital, Aalborg, North Denmark, Denmark

³Centre for Pancreatic Diseases and Mech-Sense, Department of Gastroenterology and Hepatology, Aalborg University Hospital, Aalborg, Denmark

Corresponding Author:

Simon Cichosz, BME, PhD
Department of Health Science and Technology
Aalborg University
Selma Lagerlöfs Vej 249
Gistrup, North Denmark 9260
Denmark
Phone: 45 99403809
Email: simcich@hst.aau.dk

Abstract

Background: Continuous glucose monitoring (CGM) generates high-frequency time-series data, creating challenges for efficient storage, transmission, and analysis.

Objective: This study aimed to develop and evaluate a CGM-specific compression method that achieves high compression ratios while preserving signal fidelity and clinically relevant glycemic metrics.

Methods: We introduce a content-based encoding approach (PN+) that represents CGM profiles using physiologically salient landmarks: glucose peaks and nadirs and a small set of optimally selected support points. Reconstruction is performed using piecewise cubic Hermite interpolation. PN+ was evaluated against peaks and nadirs only, uniform downsampling, piecewise aggregate approximation, and autoencoder-based compression. Performance was assessed across multiple compression ratios using 2 complementary datasets: 40,000 synthetic CGM profiles and real-world CGM data from a randomized crossover trial (558 days from 30 patients). Performance was evaluated using compression ratio, mean absolute error, and R^2 between original and reconstructed CGM-derived clinical metrics.

Results: At a compression ratio of 13 (22 points per 24-hour profile), PN+ achieved substantially lower reconstruction error than comparator methods (mean absolute error=0.77 vs 2.75-3.45) and consistently higher R^2 values across glycemic metrics. Improvements were most pronounced for excursion-sensitive measures such as mean amplitude of glycemic excursions, where PN+ reduced error by more than 4-fold compared with downsampling, piecewise aggregate approximation, and autoencoders. These performance advantages were preserved in heterogeneous real-world data. Encoding and decoding required less than 0.2 seconds per profile, supporting practical scalability.

Conclusions: PN+ enables robust CGM data compression by explicitly preserving physiologically meaningful glucose dynamics. The method outperforms generic compression techniques in reconstructing clinically relevant metrics while maintaining low computational overhead, making it well suited for large-scale CGM storage, interoperability, and downstream analytics.

JMIR Biomed Eng 2026;11:e91959; doi: [10.2196/91959](https://doi.org/10.2196/91959)

Keywords: compression; continuous glucose monitoring; CGM; encoding; reconstruction; signal; diabetes; data

Introduction

Background

Continuous glucose monitoring (CGM) has become a cornerstone of modern diabetes care, yielding a time series of glucose measurements every few minutes [1-3]. This high-resolution data stream has revolutionized both clinical management and research, enabling advanced evaluation of therapies, classification of patient subgroups, data-driven prediction of glucose dynamics, and discovery of glycaemic patterns via machine learning [4-9]. For example, the adoption of CGM in type 1 diabetes has surged in recent years, reflecting its importance for decision support and personalized treatment [10]. However, the richness of CGM data and other biomedical data also poses challenges [11]: large numbers of daily readings per patient produce very large datasets that must be stored, shared, and processed efficiently.

Storing and interoperating on such high-frequency data can be prohibitive. As Jacobsson et al [11] state, modern health care systems face an “ever-increasing need for retrieving, storing, and managing the large amount of biomedical signal data generated,” and common data exchange standards such as the Health Level Seven Fast Healthcare Interoperability Resources are inefficient when used as a long-term storage format. Naively saving CGM streams as text-based Fast Healthcare Interoperability Resources records or CSV files incurs high overhead and wastes storage. These inefficiencies hinder scalable CGM data integration and slow down analysis. Together, these challenges highlight the need to explore compression or compact encoding methods that reduce data size while preserving critical clinical information [11].

Efficient encoding and compression are also key to secure data sharing and interoperability. In emerging health care architectures, technologies such as blockchain and the Internet of Things require lean data formats to function at scale [12,13]. For instance, blockchain-based platforms for diabetes care have been proposed to improve data sharing and patient-centric control [14], but the limited throughput of distributed ledgers means that raw CGM time series cannot be stored on-chain without preprocessing. By compressing and encoding glucose traces into concise, standardized representations, data can be embedded in interoperable payloads or off-chain repositories and exchanged securely between hospitals and patients.

At the same time, CGM analysis increasingly relies on artificial intelligence and deep learning. Neural networks for glucose forecasting and pattern recognition often require well-structured, fixed-dimensional input and benefit from reduced noise. Textual encoding of CGM traces has been shown to greatly aid such analytics. For example, Igbe et al [15] demonstrated that mapping CGM profiles to reduced-alphabet “strings” could enable powerful search and classification, as well as support predictive modeling, anomaly detection, and even generative artificial intelligence applications. In clinical practice, alphabetic encoding could facilitate data indexing, reduce search time for glycaemic

patterns, and reduce computational complexity by leveraging text-based data structures and machine learning pipelines. Such encoded biomedical data can be fed directly into deep learning models to improve the speed and accuracy of analysis [16,17]. Several time-series data representations have been proposed for analysis, such as perceptually important points, which reduce dimensionality by dynamically retaining only the critical data points, such as extreme peaks or rapid fluctuations, that govern the primary trajectory of a sequence [18]. Another technique is piecewise aggregate approximation (PAA), which discretizes a raw time series into equal-length segments and replaces each with its arithmetic mean [19], whereas symbolic aggregate approximation further extends this by mapping those aggregated numerical values into a discrete alphabet [20]. Beyond these time-domain heuristics, frequency-domain transformations have proven effective for electrocardiogram (ECG) signals [21]. In addition, methods such as the use of convolutional autoencoders have been applied to ECG signals. This approach leverages the feature extraction power of deep learning to achieve significant data reduction while using a residual mechanism to specifically address and correct the reconstruction errors [22].

Problem Statement

Despite the rapid growth in the use of CGM, compression algorithms specifically designed for CGM data remain relatively underexplored. Most existing approaches rely on generic signal compression techniques that do not explicitly account for the clinical characteristics of glucose dynamics. CGM exhibits less temporal redundancy in contrast to other biological signals such as ECGs, which makes compression challenging. In this study, we aimed to propose a compression approach that extends the concept of perceptually important points by focusing on peak-nadir structures and strategic support points within CGM signals. The underlying hypothesis was that CGM data can be compressed at high compression ratios while still enabling accurate reconstruction of the signal, preserving clinically relevant information and maintaining low error in commonly used CGM-derived clinical metrics.

Contributions

This work makes 3 primary contributions. First, we introduce a CGM-specific, content-based compression method (PN+) that encodes glucose profiles using physiologically salient landmarks—peaks, nadirs, and a small set of optimally selected support points—enabling high compression while preserving clinically meaningful signal dynamics. Second, we provide a comprehensive quantitative evaluation of PN+ across multiple compression ratios by comparing it with uniform downsampling, PAA, and autoencoder-based compression using both synthetic and real-world CGM datasets and demonstrate better fidelity for excursion-sensitive consensus CGM metrics. Third, we show that high-fidelity reconstruction of CGM signals is achievable from a sparse landmark representation with minimal computational overhead.

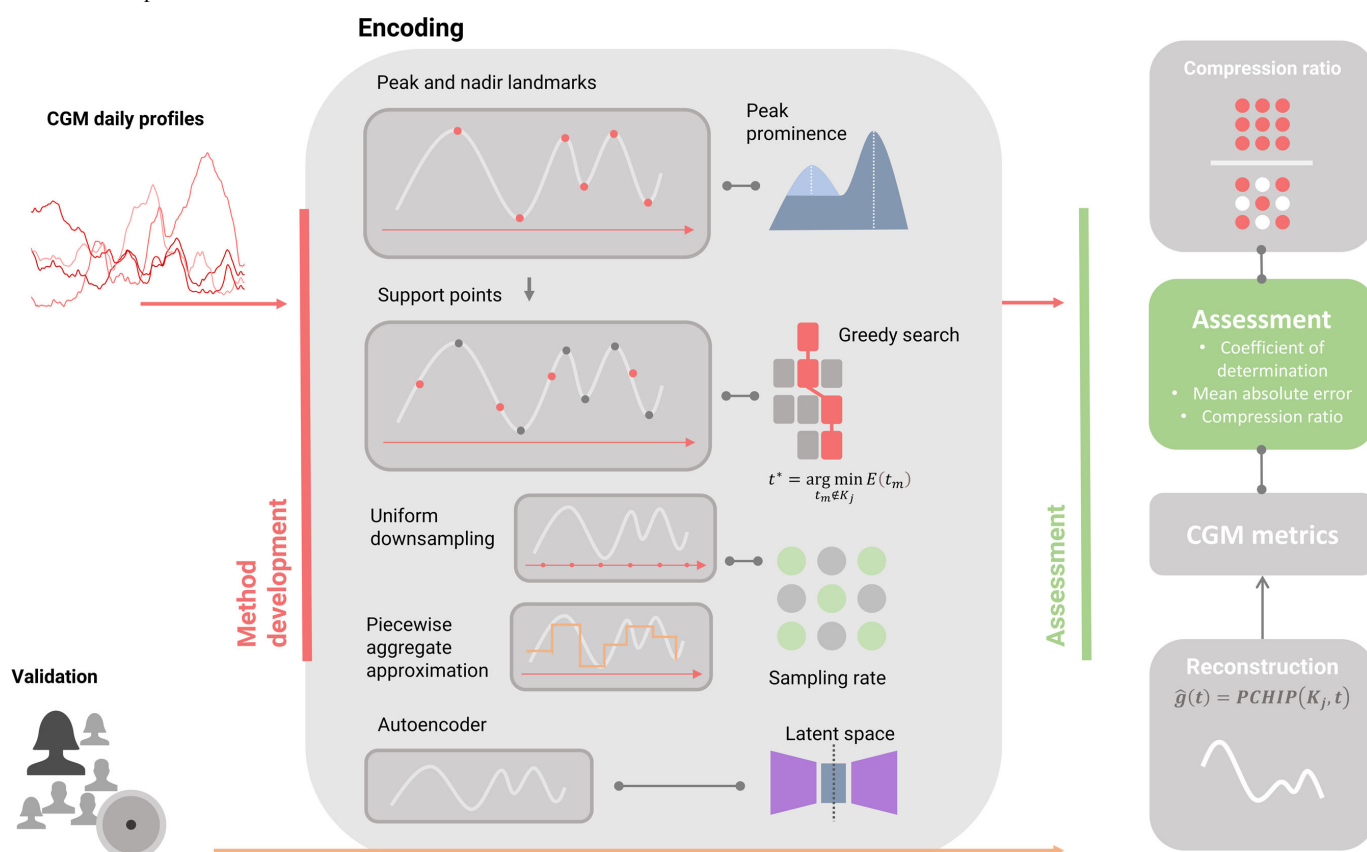
Methods

Overview

We used 2 complementary CGM datasets—synthetic data generated via a conditional generative adversarial network (CGAN) [23,24] and real-world CGM measurements from a randomized crossover trial [25]—to develop and validate 2 encoding methods based on peak and nadir information and on peaks and nadirs with support points and compare

them with uniform downsampling, autoencoder, and PAA approaches. The encoding reduced each profile to a small set of values, and reconstruction of the CGM signal used interpolation. We compared compression ratios, reconstruction error (mean absolute error; MAE), and established CGM glycemic metrics between original and decoded signals. Statistical analyses assessed whether encoding preserved clinically relevant glucose features. An overview of the methodology is illustrated in Figure 1.

Figure 1. Overview of the methodological framework. Continuous glucose monitoring (CGM) daily profiles were encoded using four approaches: (1) peak and nadir landmarks identified by peak prominence, (2) support points selected through a greedy search procedure, (3) uniform downsampling at a predefined sampling rate, and (4) autoencoder-based compression. The encoded signals were subsequently reconstructed using piecewise cubic Hermite interpolating polynomial (PCHIP). Reconstruction performance was evaluated based on the coefficient of determination, mean absolute error, and compression ratio.



Ethical Considerations

All procedures were performed in compliance with relevant laws and institutional guidelines and have been approved by the appropriate institutional committee. For the validation data, the North Denmark Region Committee on Health Research Ethics approved the protocol (N-20210064), and the study adhered to the Declaration of Helsinki and good clinical practice. Written informed consent was obtained before enrollment.

Data Sources

Discovery Data

We used a publicly available synthetic CGM dataset [23,24] generated by a CGAN trained on real CGM time series from healthy individuals and patients with type 1 diabetes across 4

hemoglobin A_{1c} groups (<6.5%, 6.5% to <7%, 7% to <8%, and ≥8%). In brief, the CGAN architecture was based on a conditional on-class label (hemoglobin A_{1c} group), with generator and discriminator networks following the formulation by Mirza and Osindero [26]. The dataset included 40,000 profiles of 24 hours each with a sample frequency of 288 measurements per day.

Clinical Validation Data

For validation, we used CGM baseline recordings from a randomized, open-label, crossover trial in patients with chronic pancreatitis and insulin-treated diabetes (N=30; mean age 64.4, SD 8.8 years; 75.9% male), each undergoing 20 days of CGM (Dexcom G6; sampling rate: 5-minute intervals) for the baseline [25]. The dataset consisted of 558 days, each treated as an independent observation in the

analysis. The patient cohort was characterized by high inter- and intraglycemic variability, making it a relevant population for evaluating compression methods in CGM data with low temporal redundancy [27,28].

Encoding and Decoding Algorithms

We compared 5 encoding approaches for each CGM profile. For each signal landmark (peak, nadir, or support point), we recorded its time stamp and glucose value.

Peaks and Nadirs Only

This method involved identifying local maxima (“peaks”) and minima (“nadirs”) by sign changes in the discrete derivative of the glucose time series subject to a minimum prominence threshold to avoid noise.

Peaks, Nadirs, and Support Points (PN+)

This method extended the peak and nadir approach by incorporating additional “support” points between peaks and nadirs. These points were selected using a greedy search strategy to better capture the overall shape of the glucose trajectory.

Uniform Downsampling

As a baseline comparison, we applied uniform downsampling by removing every x th sample to simulate a reduced sampling rate.

Autoencoders

A deep neural network autoencoder architecture [29] was implemented as a comparison approach using an encoder-decoder framework. The network was trained for 800 epochs with L2 and sparsity regularization. Five-fold cross-validation was used to minimize overfitting. The latent space dimensionality was chosen to achieve compression ratios comparable to those of the other methods.

PAA Method

The CGM time series was partitioned into a predefined number of contiguous, nonoverlapping segments of approximately equal length. Within each segment, glucose values were replaced with their mean, yielding a compressed representation that preserved the overall temporal trend while reducing high-frequency variability. The number of segments was selected to achieve a compression ratio comparable to those of the other methods, and reconstruction was performed by assigning each segment mean back to its corresponding time interval.

Signal Reconstruction

Decoded profiles were reconstructed using piecewise cubic Hermite interpolating polynomial (PCHIP) [30] using the identified landmarks, yielding a time series equivalent to the original 5-minute sampling rate. This simple method preserves key excursion patterns while minimizing computational load.

Peak Detection in Glucose Time Series

To identify the significant excursions in glucose levels, we used a peak detection algorithm. Glucose values were treated as a 1D time series $g(t)$, where t represents the corresponding time stamps of glucose measurements.

Local maxima were identified with a minimum peak prominence threshold of 15 mg/dL to ensure physiological relevance and reduce the influence of minor fluctuations and noise. In this context, prominence quantifies how much a peak stands out due to its height and separation from neighboring valleys. Specifically, the prominence of each peak was computed as the vertical distance between the peak and the lowest point in the signal, separating it from a higher neighboring peak.

The detection criterion can be formally described as that each candidate peak at time t_i was retained only if the following applied:

$$g(t_i) - \max(\min_{t_k < t_i} g_k^\downarrow, \min_{t_k > t_i} g_k^\downarrow) \geq 15 \text{ mg/dL}$$

In this equation, $g_k^\downarrow$ represents local valleys in the signal. This approach yielded a set of clinically relevant glucose peaks. For nadir detections, the same procedure was repeated with an inverted signal.

Including Additional Support Points

To efficiently approximate CGM profiles using the limited number of representative landmarks, we implemented a greedy optimization strategy for selecting support points. The goal was to identify a sparse subset of additional points that allowed for the accurate reconstruction of the original signal using PCHIP.

Given an initial subset of k known landmark points $K_0 = \{(t_{i1}, g_{i1}), \dots, (t_{ik}, g_{ik})\}$ where $i_j \in \{1, \dots, N\}$, the goal was to iteratively augment this set to a total of n points by selecting additional time points that minimized reconstruction error. The upper bound for n was constrained by the desired compression ratio.

At each iteration j , the algorithm evaluated all remaining candidate time points $t_m \notin K_j$ and selected the one that, when added, minimized the L2 norm of the reconstruction error:

$$E_t(t_m) = \|g(t) - \hat{g}(K_j, t_m)\|_2$$

In this equation, $g(K_j; t_m)$ is the interpolated glucose signal using PCHIP over the updated key set $K_{j+1} \cup \{(t_m, g(t_m))\}$. The optimal new point was selected as follows:

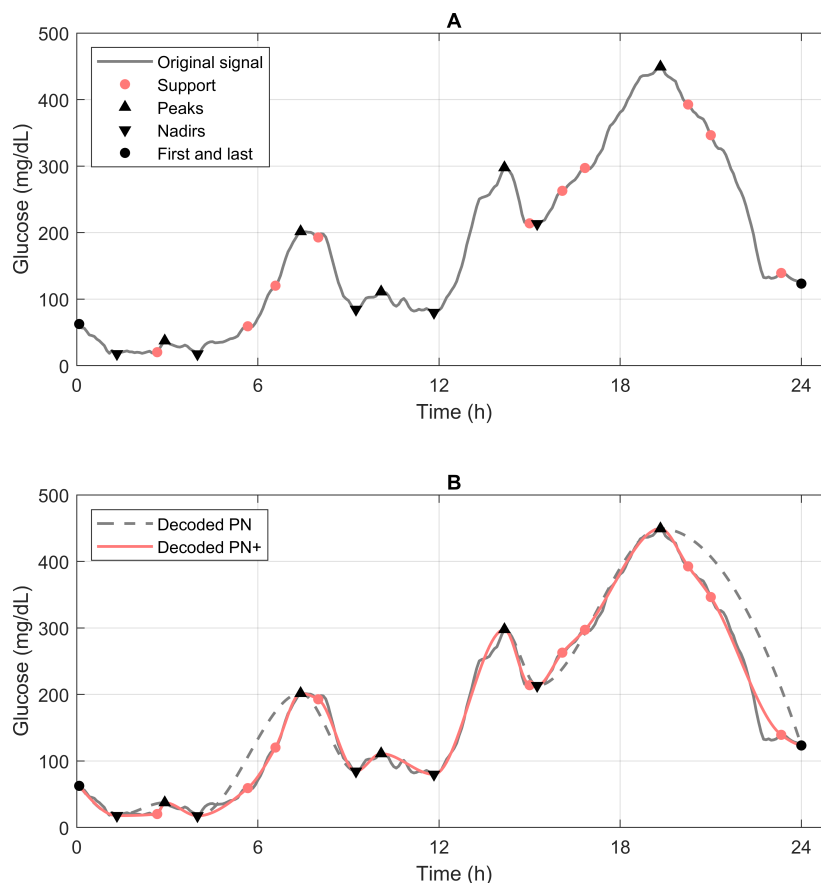
$$t^* = \underset{t_m \notin K_j}{\operatorname{argmin}} E(t_m)$$

This process was repeated until $|K_j|$ reached the maximum number of desired landmark points. The final reconstructed signal $g(t)$ was generated via $g(t) = \text{PCHIP}(K_j, t)$.

Figure 2 illustrates the selected landmark points extracted from the CGM profiles, including both peak and nadir points as well as the additional support points used in the extended PN+ approach. The corresponding reconstructed signals for both methods are shown. Notably, the inclusion of

support points in the PN+ approach improved the alignment of the reconstructed signal with the original glucose profile, particularly in regions between peaks and nadirs where the rate of change was nonuniform.

Figure 2. (A) An example of a 24-hour continuous glucose monitoring profile (original signal) with the peak, nadir, and support points identified by the algorithm and (B) the reconstructed signals using only the peak and nadir points (PN) and using the method with additional support points (PN+).



Assessments

To assess the compression of the approaches (peaks and nadirs only, PN+, downsampling, autoencoders, and PAA), we calculated the compression ratio between the uncompressed and compressed profiles:

$$\text{Compression ratio} = \frac{\text{uncompressed}}{\text{compressed}}$$

Furthermore, we calculated the MAE between the original and decoded glucose profile metrics:

$$\text{MAE} = \frac{1}{N} \sum_{i=1}^n |\text{metric}(P_i)_{\text{original}} - \text{metric}(P_i)_{\text{reconstructed}}|$$

To assess whether the encoding profiles preserved clinically relevant glucose features, we computed standard consensus CGM metrics and glycemic variability metrics [31-33] on both original and decoded profiles. The MAE and coefficient of determination (R^2) were calculated for each metric between

the original and decoded profiles. The standard consensus CGM metrics included were mean glucose, SD of glucose, coefficient of variation, time in range (70-180 mg/dL), time in tight range (70-140 mg/dL), time below range (TBR; <70 mg/dL), TBR level 1 (54-70 mg/dL), TBR level 2 (<54 mg/dL), time above range (TAR; >180 mg/dL), TAR level 1 (180-250 mg/dL), TAR level 2 (>250 mg/dL), and mean amplitude of glycemic excursions (MAGE) [34].

Analyses were performed in MATLAB (vR2021b; MathWorks Inc) on a laptop equipped with an 11th-generation Intel Core i7-11850H (2.50 GHz), 32 GB of RAM, and an NVIDIA T1200 graphics processing unit. Standard toolboxes for signal processing and deep learning were used, and all analyses were performed on this configuration to ensure the reproducibility of the results. For computing CGM metrics, we used an open-source tool, Quantification of Continuous Glucose Monitoring, designed for CGM data analysis using the MATLAB environment [35].

Results

Discovery Data

As shown in Figure 3, the methods were compared across various settings—downsampling rate, latent space dimensions, peak prominence threshold, and the allowed number of total support points—to evaluate performance under different compression rates. Among the methods, PN+ consistently achieved the highest average R^2 , whereas downsampling yielded the lowest performance at each comparable compression ratio. Furthermore, the impact of varying peak prominence thresholds on the peaks and nadirs only and PN+ approaches is illustrated in Figure 4. As expected, PN+ resulted in lower compression at each threshold level due to the inclusion of additional support points. However, when comparing the methods at equivalent compression ratios, PN+ exhibited substantially lower error than the peaks and nadirs only method. This suggests that PN+ is a more effective approach for capturing clinically relevant physiological signal patterns. The encoding and decoding combined took an average of 0.13 (SD 0.05) seconds per 24-hour profile on a PC laptop (11th-generation Intel Core i7-11850H at 2.50 GHz, 32 GB of RAM).

A direct comparison of CGM-derived metrics from the reconstructed signals and the original signals is shown in Table 1 using a peak prominence threshold of 15 mg/dL and a minimum compression ratio of 13 (corresponding to 22/288, 7.6% of the original landmark points). At this compression level, PN+ consistently achieved higher R^2 values across most CGM-derived metrics than the downsampling, autoencoder, and PAA methods. More notably, the MAE for each metric was substantially lower with PN+, with a reduction factor of 3.6 to 4.5 (Table 1). This improvement was especially pronounced for MAGE, which captures large glucose excursions—an area where the downsampling and autoencoder methods introduced significant estimation errors, as illustrated in Figure 5 through a Bland-Altman analysis plot.

Overall, PN+ demonstrated strong alignment between metrics derived from reconstructed and original signals. However, the estimation of TBR level 2, representing rare low-glucose episodes, remained sensitive to the encoding and decoding process due to the sparse occurrence of such events over a typical day.

Figure 3. Compression ratio plotted against the average mean absolute error (MAE) values for the 4 methods across all glycemic metrics derived from the reconstructed continuous glucose monitoring signals. PAA: piecewise aggregate approximation; PN: peaks and nadirs only; PN+: peaks and nadirs and support points.

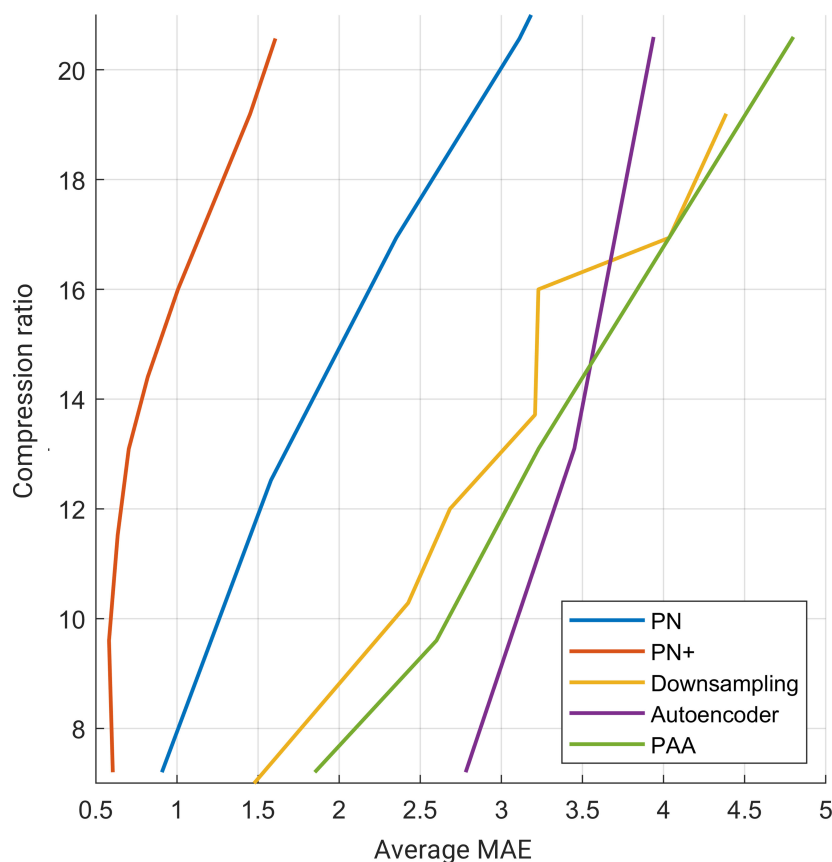


Figure 4. Impact of varying peak prominence thresholds on the mean absolute error (MAE) of the peaks and nadirs only (PN) and peaks and nadirs with support points (PN+) approaches.

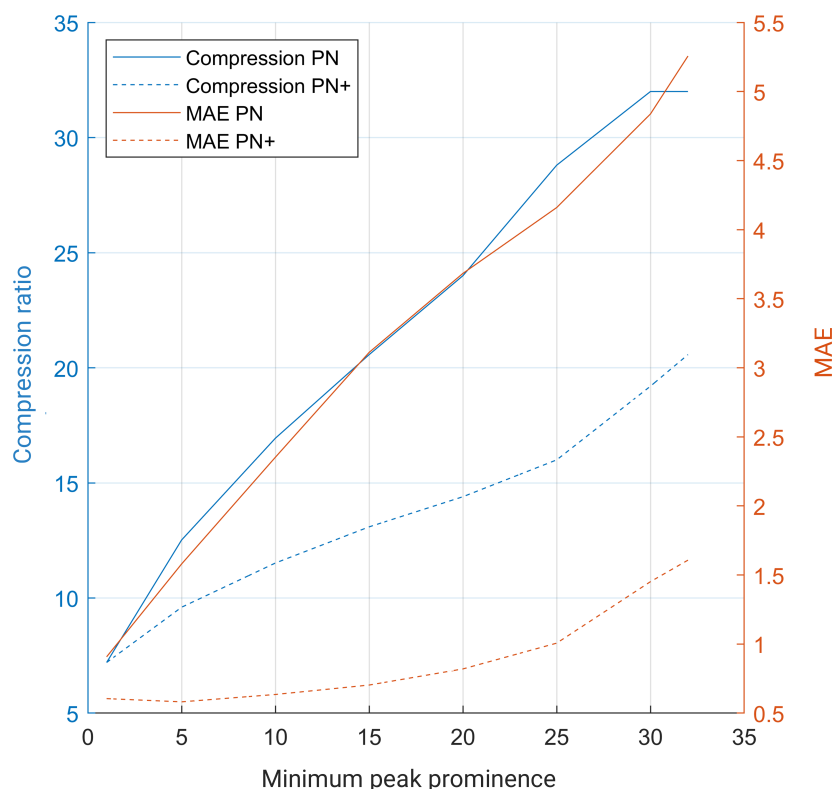


Table 1. Overall performance and compression ratio of peaks and nadirs only (PN) encoding, PN encoding with support points (PN+), down-sampling, autoencoders, and piecewise aggregate approximation (PAA) for each continuous glucose monitoring metric in the discovery dataset (n=40,000).

	PN ^a		PN+ ^b		Downsampling ^b		Autoencoder ^b		PAA ^b	
	R^2	MAE ^c	R^2	MAE	R^2	MAE	R^2	MAE	R^2	MAE
Mean glucose (mg/dL)	0.98	5.24	1	0.64	1	0.61	1	1.98	1	0.14
CV ^d (ratio)	0.97	2.01	1	0.50	0.99	2.02	0.97	2.08	1	1.42
SD of glucose (mg/dL)	0.98	3.3	1	0.65	1	2.91	0.99	2.84	1	1.99
TIR ^e (%)	0.97	3.36	1	1.14	0.99	2.12	0.97	3.83	0.99	2.43
TITR ^f (%)	0.97	3.64	1	1.36	0.99	2.36	0.98	4.45	0.99	2.68
TBR ^g (%)	0.90	1.52	0.98	0.66	0.97	1.1	0.87	2.5	0.98	1.15
TBR1 ^h (%)	0.75	1.36	0.92	0.82	0.89	1.11	0.64	1.95	0.92	1.32
TBR2 ⁱ (%)	0.90	0.66	0.98	0.32	0.96	0.47	0.75	1.42	0.98	0.5
TAR ^j (%)	0.97	2.79	1	0.70	0.99	1.38	0.99	1.63	0.99	1.83
TAR1 ^k (%)	0.92	2.76	0.99	0.89	0.97	1.87	0.96	2.14	0.98	2.34
TAR2 ^l (%)	0.96	1.89	1	0.40	0.99	0.98	0.99	1.03	0.99	1.25
MAGE ^m (mg/dL)	0.97	3.71	0.99	1.1	0.90	16.13	0.89	15.55	0.92	21.76
Overall mean	0.94	2.69	0.99	0.77	0.98	2.75	0.92	3.45	0.96	3.23

^aCompression ratio: 20.6.

^bCompression ratio: 13.1.

^cMAE: mean absolute error.

^dCV: cross-validation.

^eTIR: time in range.

^fTITR: time in tight range.

^gTBR: time below range.

^hTBR1: TBR level 1.

ⁱTBR2: TBR level 2.

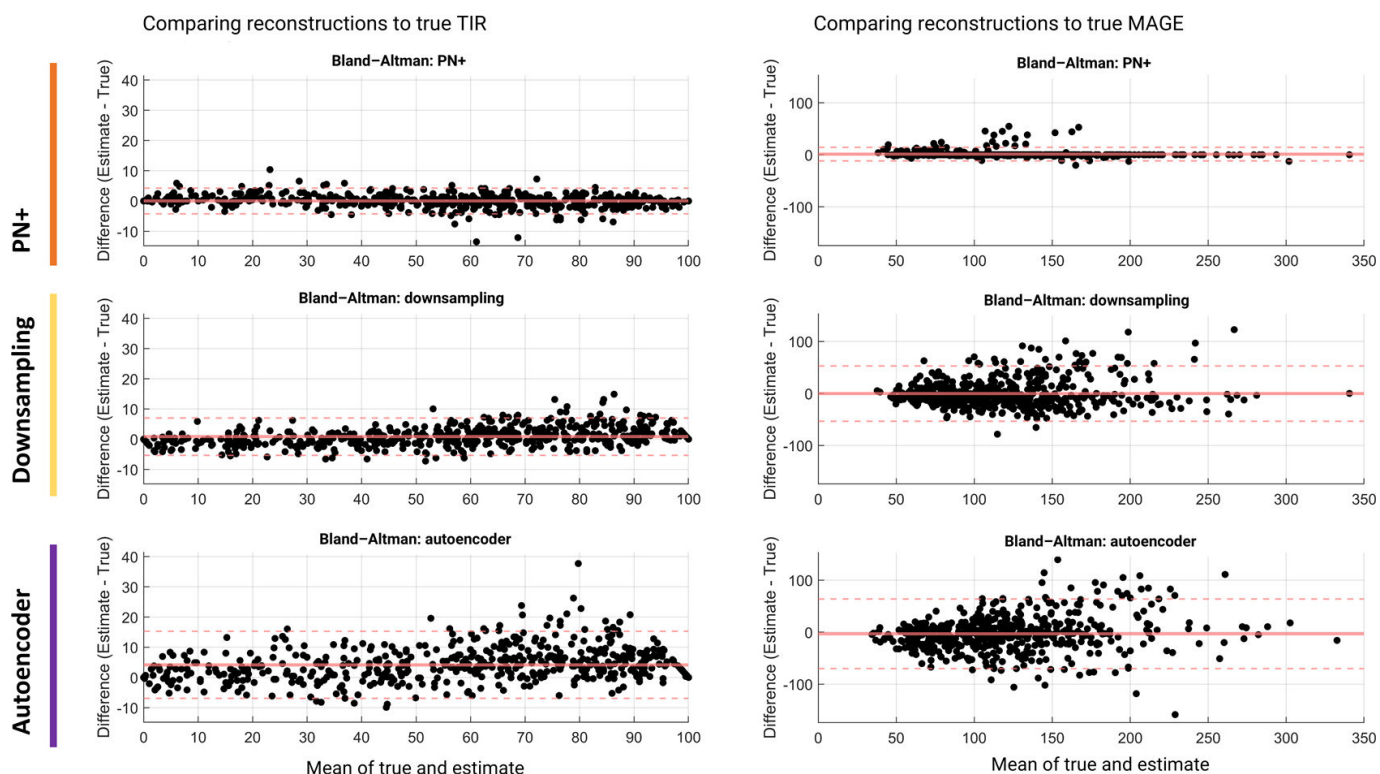
^jTAR: time above range.

^kTAR1: TAR level 1.

^lTAR2: TAR level 2.

^mMAGE: mean amplitude of glycemic excursions.

Figure 5. Bland-Altman analysis plots comparing the methods (peaks, nadirs, and support points [PN+], downsampling, and autoencoders) for reconstruction error estimates of clinical metrics (time in range [TIR] and mean amplitude of glycemic excursions [MAGE]) on clinical validation data.



Clinical Validation Data

As shown in Table 2, the validation dataset yielded overall performance trends similar to those observed in the discovery dataset. However, the R^2 and MAE values were slightly worse across all methods, likely reflecting the increased complexity and variability inherent in real-world patient data. Notably, the validation cohort included individuals with diabetes associated with chronic pancreatitis, a diabetes

subtype known for markedly impaired glucose regulation and high glycemic variability due to deficiencies in insulin and glucagon secretion along with variable insulin sensitivity [27]. Despite this added complexity, PN+ continued to outperform the downsampling, autoencoder, and PAA methods at equivalent compression ratios, achieving the lowest error and highest R^2 among the compared methods.

Table 2. Overall performance and compression ratio of peaks and nadirs only (PN) encoding, PN encoding with support points (PN+), downsampling, autoencoders, and piecewise aggregate approximation (PAA) for each continuous glucose monitoring metric in the clinical validation dataset.

	PN ^a		PN+ ^b		Downsampling ^b		Autoencoder ^b		PAA ^b	
	R^2	MAE ^c	R^2	MAE	R^2	MAE	R^2	MAE	R^2	MAE
Mean glucose (mg/dL)	0.96	8.13	1	0.99	1	1.4	0.99	5.78	1	0.17
CV ^d (ratio)	0.87	2.68	0.99	0.69	0.97	1.68	0.93	2.24	1	1.25
SD of glucose (mg/dL)	0.89	5.16	0.99	1.15	0.98	2.9	0.95	4.38	1	2.15
TIR ^e (%)	0.93	6.26	0.99	1.43	0.99	2.32	0.98	5.16	1	2.4
TTIR ^f (%)	0.89	5.84	0.99	1.63	0.98	2.34	0.97	5.34	0.99	2.3
TBR ^g (%)	0.75	0.95	0.95	0.33	0.90	0.49	0.8	1.02	0.96	0.45
TBR1 ^h (%)	0.60	0.80	0.92	0.31	0.91	0.39	0.79	0.81	0.94	0.47
TBR2 ⁱ (%)	0.75	0.41	0.84	0.16	0.56	0.16	0.16	0.21	0.82	0.14
TAR ^j (%)	0.94	5.65	1	1.24	0.99	2.01	0.99	4.24	1	2.2
TAR1 ^k (%)	0.79	5.83	0.98	1.57	0.96	2.89	0.92	5.45	0.98	2.88
TAR2 ^l (%)	0.95	3.56	1	0.81	0.99	1.5	0.99	2.6	1	1.56
MAGE ^m (mg/dL)	0.90	5.79	0.98	1.86	0.72	20.2	0.79	24.32	0.81	23.41
Overall mean	0.85	4.25	0.97	1.01	0.91	3.19	0.78	5.12	0.92	3.28

^aCompression ratio: 24.1.

^bCompression ratio: 13.1.

^cMAE: mean absolute error.

^dCV: cross-validation.

^eTIR: time in range.

^fTITR: time in tight range.

^gTBR: time below range.

^hTBR1: TBR level 1.

ⁱTBR2: TBR level 2.

^jTAR: time above range.

^kTAR1: TAR level 1.

^lTAR2: TAR level 2.

^mMAGE: mean amplitude of glycemic excursions.

Discussion

Principal Findings

The proposed novel encoding algorithm PN+ demonstrated advantages for CGM data compression by explicitly selecting physiologically meaningful points. By anchoring the compressed signal on glucose peaks and nadirs (and additional support points via a greedy strategy), PN+ yielded substantially lower reconstruction error and higher explained variance (R^2) across CGM-derived metrics than the uniform downsampling, PAA, and autoencoder methods at equivalent compression ratios. The method consistently preserved clinical metrics across both the discovery and clinical validation datasets. Performance improvements were observed across multiple glycemic measures, and the methods demonstrated resilience to variations in signal patterns, including postprandial excursions. In particular, PN+ better preserved MAGE, a gold-standard CGM metric for assessing large glycemic excursions [36]. Because MAGE and similar variability indexes depend on accurately capturing excursions, PN+'s targeted point selection translates directly to better metric fidelity. Furthermore, reconstructing the compressed signal using PCHIP maintains the original shape and smoothness of glucose trends [35], avoiding the overshoot or oscillation that might arise with simpler interpolation. Overall, the proposed algorithm was able to reconstruct the CGM signals, preserving key clinically relevant features of the curves from a substantially decompressed version. While the deep learning-based comparison method, the autoencoder, enables nonlinear encoding and decoding of the CGM signal and represents a more advanced data-driven compression strategy, it did not translate into improved signal reconstruction with respect to clinically derived metrics in our experiment. The implemented autoencoders were constrained by the latent space dimensionality to achieve compression ratios comparable to those of the other evaluated methods. The downsampling approach performed comparably in preserving clinical metrics that were not sensitive to variability; however, it was less effective in capturing dynamic metrics compared with PN+.

To date, relatively few studies have addressed the compression or encoding of CGM profiles. However, Kovatchev et al [37], using a multistep machine learning procedure, showed how it is possible to reconstruct a virtual CGM profile from the original sparse data (7-point blood glucose profiles) while preserving clinically relevant

information [38]. Notably, Igbe and Kovatchev [15] recently introduced a novel approach that encodes daily CGM profiles into symbolic representations—referred to as CGM strings and CGM texts—that preserve key clinical metrics while achieving data compression. Their analysis demonstrated that a 9-character encoding corresponding to an approximate compression ratio of 10:1 could retain information on cross-validation with an R^2 of up to 0.93. The method presented in this study builds on and extends these findings by enabling the preservation of a broader range of CGM-derived metrics across multiple compression levels. Moreover, our approach offers the additional advantage of reconstructing the original signal with high precision, supporting both clinical interpretability and downstream analytical use.

The proposed approach is fundamentally a content-based compression strategy and could be applied to other periodic or episodic biomedical signals. For example, ECG waveforms hinge on distinct QRS complex peaks for heartbeats, and compression methods that preserve those peaks can maintain diagnostic integrity. Indeed, prior work has shown that a lossy ECG compression scheme preserved the main features of the ECG morphology (notably, the QRS complexes) despite a high compression ratio of 4.5 [39]. Similarly, photoplethysmographic waveforms depend on capturing the systolic pulse peaks; a PN+—like method could retain these pulse peaks as key support points. In both ECG and photoplethysmography contexts, the idea is the same as for CGM: allocate sampling “budget” to physiologically salient events (heartbeats or pulses) and interpolate the rest. Thus, PN+ is broadly relevant to any biosignal where preserving the height and timing of key peaks and nadirs is critical for clinical or functional interpretation. Future work should explore the potential applicability of the proposed methodology to other biological signals. Moreover, future work should also investigate the clinical impact on decision-making based on reconstructed signal data.

Strengths and Limitations

This study was based on a large synthetic dataset of CGM profiles derived from a heterogeneous population of individuals with diabetes. Furthermore, the proposed methods were externally validated using data from a distinct patient group with diabetes secondary to chronic pancreatitis. This could support the generalizability of the results to a broader population of individuals with diabetes. However,

the proposed PN+ method has some limitations typical of lossy, feature-based compression approaches. One concern is rare event sensitivity: very short or abrupt excursions (such as sudden severe hypoglycemia events) might not always produce easily identified peaks or may occur between selected support points. In practice, we observed that extremely brief dips can be slightly underrepresented by PN+ unless thresholds are tuned to catch them. Relatedly, the performance gain of PN+ depends on the presence of signal structure. If a CGM trace is unusually flat or noisy (eg, due to sensor artifacts), the greedy peak-nadir selection may yield fewer benefits, and overall reconstruction error for all methods can increase. Indeed, in our heterogeneous validation cohort, all compression methods saw modestly reduced accuracy, although PN+ remained superior; this suggests that

real-world variability poses challenges even to structured approaches. Finally, PN+ currently uses fixed rules (greedy selection and threshold-based peak and nadir identification), so it may require adaptation (eg, dynamic thresholds or rule changes) when applied to data with very different characteristics.

Conclusions

The proposed PN+ method produces a compact CGM representation that retains critical glycemic dynamics while discarding redundant portions of the signal. The CGM signal can be reconstructed with high precision from the encoding representation. The application of the method to other biomedical signals needs further investigation.

Funding

The authors declared no financial support was received for this work.

Data Availability

The datasets generated or analyzed during this study are not publicly available due to sensitive personal information but are available from the corresponding author on reasonable request.

Conflicts of Interest

SC has received research funding from i-SENS, holds shares in Novo Nordisk A/S, and has received consultancy fees from Roche Diagnostics and Medicus Engineering. All other authors confirm they have no conflicts to declare.

References

1. Battelino T, Lalic N, Hussain S, et al. The use of continuous glucose monitoring in people living with obesity, intermediate hyperglycemia or type 2 diabetes. *Diabetes Res Clin Pract.* May 2025;223:112111. [doi: [10.1016/j.diabres.2025.112111](https://doi.org/10.1016/j.diabres.2025.112111)] [Medline: [40118193](https://pubmed.ncbi.nlm.nih.gov/40118193/)]
2. Klonoff DC, Ahn D, Drincic A. Continuous glucose monitoring: a review of the technology and clinical use. *Diabetes Res Clin Pract.* Nov 2017;133:178-192. [doi: [10.1016/j.diabres.2017.08.005](https://doi.org/10.1016/j.diabres.2017.08.005)] [Medline: [28965029](https://pubmed.ncbi.nlm.nih.gov/28965029/)]
3. Bender C, Vestergaard P, Cichosz SL. The history, evolution and future of continuous glucose monitoring (CGM). *Diabetology.* 2025;6(3):17. [doi: [10.3390/diabetology6030017](https://doi.org/10.3390/diabetology6030017)]
4. Woldaregay AZ, Årsand E, Walderhaug S, et al. Data-driven modeling and prediction of blood glucose dynamics: machine learning applications in type 1 diabetes. *Artif Intell Med.* Jul 2019;98:109-134. [doi: [10.1016/j.artmed.2019.07.007](https://doi.org/10.1016/j.artmed.2019.07.007)] [Medline: [31383477](https://pubmed.ncbi.nlm.nih.gov/31383477/)]
5. Kahkoska AR, Adair LA, Aiello AE, et al. Identification of clinically relevant dysglycemia phenotypes based on continuous glucose monitoring data from youth with type 1 diabetes and elevated hemoglobin A1c. *Pediatr Diabetes.* Aug 2019;20(5):556-566. [doi: [10.1111/vedi.12856](https://doi.org/10.1111/vedi.12856)] [Medline: [30972889](https://pubmed.ncbi.nlm.nih.gov/30972889/)]
6. Ayers AT, Ho CN, Kerr D, et al. Artificial intelligence to diagnose complications of diabetes. *J Diabetes Sci Technol.* Jan 2025;19(1):246-264. [doi: [10.1177/19322968241287773](https://doi.org/10.1177/19322968241287773)] [Medline: [39578435](https://pubmed.ncbi.nlm.nih.gov/39578435/)]
7. Cichosz SL, Kronborg T, Laugesen E, et al. From stability to variability: classification of healthy individuals, prediabetes, and type 2 diabetes using glycemic variability indices from continuous glucose monitoring data. *Diabetes Technol Ther.* Jan 2025;27(1):34-44. [doi: [10.1089/dia.2024.0226](https://doi.org/10.1089/dia.2024.0226)] [Medline: [39115921](https://pubmed.ncbi.nlm.nih.gov/39115921/)]
8. Lebech Cichosz S, Hasselstrøm Jensen M, Schou Olesen S. Development and validation of a machine learning model to predict weekly risk of hypoglycemia in patients with type 1 diabetes based on continuous glucose monitoring. *Diabetes Technol Ther.* Jul 2024;26(7):457-466. [doi: [10.1089/dia.2023.0532](https://doi.org/10.1089/dia.2023.0532)] [Medline: [38215207](https://pubmed.ncbi.nlm.nih.gov/38215207/)]
9. Lebech Cichosz S, Bender C. Development of machine learning models for the identification of elevated ketone bodies during hyperglycemia in patients with type 1 diabetes. *Diabetes Technol Ther.* Jun 2024;26(6):403-410. [doi: [10.1089/dia.2023.0531](https://doi.org/10.1089/dia.2023.0531)] [Medline: [38456910](https://pubmed.ncbi.nlm.nih.gov/38456910/)]
10. Lacy ME, Lee KE, Atac O, et al. Patterns and trends in continuous glucose monitoring utilization among commercially insured individuals with type 1 diabetes: 2010-2013 to 2016-2019. *Clin Diabetes.* 2024;42(3):388-397. [doi: [10.2337/cd23-0051](https://doi.org/10.2337/cd23-0051)] [Medline: [39015169](https://pubmed.ncbi.nlm.nih.gov/39015169/)]
11. Jacobsson M, Seoane F, Abtahi F. The role of compression in large scale data transfer and storage of typical biomedical signals at hospitals. *Health Informatics J.* 2023;29(4):14604582231213846. [doi: [10.1177/14604582231213846](https://doi.org/10.1177/14604582231213846)] [Medline: [38063181](https://pubmed.ncbi.nlm.nih.gov/38063181/)]

12. Lal B, Gravina R, Spagnolo F, Corsonello P. Compressed sensing approach for physiological signals: a review. *IEEE Sensors J.* 2023;23(6):5513-5534. [doi: [10.1109/JSEN.2023.3243390](https://doi.org/10.1109/JSEN.2023.3243390)]
13. Nassra I, Capella JV. Data compression techniques in IoT-enabled wireless body sensor networks: a systematic literature review and research trends for QoS improvement. *Internet Things.* Oct 2023;23:100806. [doi: [10.1016/j.iot.2023.100806](https://doi.org/10.1016/j.iot.2023.100806)]
14. Lebech Cichosz S, Stausholm MN, Kronborg T, Vestergaard P, Hejlesen O. How to use blockchain for diabetes health care data and access management: an operational concept. *J Diabetes Sci Technol.* Mar 2019;13(2):248-253. [doi: [10.1177/1932296818790281](https://doi.org/10.1177/1932296818790281)] [Medline: [30047789](https://pubmed.ncbi.nlm.nih.gov/30047789/)]
15. Igbe T, Kovatchev B. Finding optimal alphabet for encoding daily continuous glucose monitoring time series into compressed text. *J Diabetes Sci Technol.* Mar 20, 2025;19322968251323913. [doi: [10.1177/19322968251323913](https://doi.org/10.1177/19322968251323913)] [Medline: [40109176](https://pubmed.ncbi.nlm.nih.gov/40109176/)]
16. Hua J, Chu B, Zou J, Jia J. ECG signal classification in wearable devices based on compressed domain. *PLoS One.* 2023;18(4):e0284008. [doi: [10.1371/journal.pone.0284008](https://doi.org/10.1371/journal.pone.0284008)] [Medline: [37014879](https://pubmed.ncbi.nlm.nih.gov/37014879/)]
17. Wu Z, Guo C. Deep learning and electrocardiography: systematic review of current techniques in cardiovascular disease diagnosis and management. *Biomed Eng Online.* Feb 23, 2025;24(1):23. [doi: [10.1186/s12938-025-01349-w](https://doi.org/10.1186/s12938-025-01349-w)] [Medline: [39988715](https://pubmed.ncbi.nlm.nih.gov/39988715/)]
18. Chung FL, Fu TC, Luk WP, Ng VT. Flexible time series pattern matching based on perceptually important points. Presented at: Workshop on Learning from Temporal and Spatial Data in International Joint Conference on Artificial Intelligence; Aug 6, 2001; Seattle, WA. URL: <https://research.polyu.edu.hk/en/publications/flexible-time-series-pattern-matching-based-on-perceptually-impor/> [Accessed 2026-06-30]
19. Siyou Fotso VS, Mephu Nguifo E, Vaslin P. Grasp heuristic for time series compression with piecewise aggregate approximation. *RAIRO Oper Res.* 2019;53:243-259. [doi: [10.1051/ro/2018089](https://doi.org/10.1051/ro/2018089)]
20. Wilson SJ. Data representation for time series data mining: time domain approaches. *WIREs Comput Stat.* 2017;9(1):e1392. [doi: [10.1002/wics.1392](https://doi.org/10.1002/wics.1392)]
21. Maalej A, Ben-Romdhane M, Tlili M, Rivet F, Dallet D, Rebai C. On the wavelet-based compressibility of continuous-time sampled ECG signal for e-health applications. *Measurement (Lond).* Nov 2020;164:108031. [doi: [10.1016/j.measurement.2020.108031](https://doi.org/10.1016/j.measurement.2020.108031)] [Medline: [32565605](https://pubmed.ncbi.nlm.nih.gov/32565605/)]
22. Shi J, Wang F, Qin M, et al. New ECG compression method for portable ECG monitoring system merged with binary convolutional auto-encoder and residual error compensation. *Biosensors (Basel).* Jul 14, 2022;12(7):524. [doi: [10.3390/bios12070524](https://doi.org/10.3390/bios12070524)] [Medline: [35884327](https://pubmed.ncbi.nlm.nih.gov/35884327/)]
23. Cichosz S, Xylander AA. Synthetic continuous glucose monitoring (CGM) signals. *Mendeley Data.* 2021. URL: <https://data.mendeley.com/datasets/chd8hx65r4/1> [Accessed 2026-06-26]
24. Lebech Cichosz S, Xylander AA. A conditional generative adversarial network for synthesis of continuous glucose monitoring signals. *J Diabetes Sci Technol.* Sep 2022;16(5):1220-1223. [doi: [10.1177/19322968211014255](https://doi.org/10.1177/19322968211014255)] [Medline: [34056935](https://pubmed.ncbi.nlm.nih.gov/34056935/)]
25. Davidsen L, Cichosz SL, Stæhr PB, et al. Efficacy and safety of continuous glucose monitoring on glycaemic control in patients with chronic pancreatitis and insulin-treated diabetes: a randomised, open-label, crossover trial. *Diabetes Obes Metab.* Jun 2025;27(6):3379-3388. [doi: [10.1111/dom.16356](https://doi.org/10.1111/dom.16356)] [Medline: [40099620](https://pubmed.ncbi.nlm.nih.gov/40099620/)]
26. Mirza M, Osindero S. Conditional generative adversarial nets. *arXiv.* Preprint posted online on Nov 6, 2014. [doi: [10.48550/arXiv.1411.1784](https://doi.org/10.48550/arXiv.1411.1784)]
27. Davidsen L, Jensen MH, Kronborg T, et al. Increased glycaemic variability in patients with chronic pancreatitis and diabetes compared to type 2 diabetes. *J Diabetes Sci Technol.* Jul 18, 2025;19322968251356239. [doi: [10.1177/19322968251356239](https://doi.org/10.1177/19322968251356239)] [Medline: [40682262](https://pubmed.ncbi.nlm.nih.gov/40682262/)]
28. Davidsen L, Lebech Cichosz S, Knoph CS, et al. Unravelling the heterogeneity of diabetes in chronic pancreatitis: alpha and beta cell dysfunction and association with glycaemic control. *Diabet Med.* Aug 2025;42(8):e70080. [doi: [10.1111/dme.70080](https://doi.org/10.1111/dme.70080)] [Medline: [40449988](https://pubmed.ncbi.nlm.nih.gov/40449988/)]
29. Valencia D, Mercier PP, Alimohammad A. Efficient in vivo neural signal compression using an autoencoder-based neural network. *IEEE Trans Biomed Circuits Syst.* Jun 2024;18(3):691-701. [doi: [10.1109/TBCAS.2024.3359994](https://doi.org/10.1109/TBCAS.2024.3359994)] [Medline: [38285576](https://pubmed.ncbi.nlm.nih.gov/38285576/)]
30. Rabbath CA, Corriveau D. A comparison of Piecewise Cubic Hermite Interpolating Polynomials, cubic splines and piecewise linear functions for the approximation of projectile aerodynamics. *Def Technol.* Oct 2019;15(5):741-757. [doi: [10.1016/j.dt.2019.07.016](https://doi.org/10.1016/j.dt.2019.07.016)]
31. Danne T, Nimri R, Battelino T, et al. International consensus on use of continuous glucose monitoring. *Diabetes Care.* Dec 2017;40(12):1631-1640. [doi: [10.2337/dc17-1600](https://doi.org/10.2337/dc17-1600)] [Medline: [29162583](https://pubmed.ncbi.nlm.nih.gov/29162583/)]

32. Battelino T, Alexander CM, Amiel SA, et al. Continuous glucose monitoring and metrics for clinical trials: an international consensus statement. *Lancet Diabetes Endocrinol*. Jan 2023;11(1):42-57. [doi: [10.1016/S2213-8587\(22\)00319-9](https://doi.org/10.1016/S2213-8587(22)00319-9)] [Medline: [36493795](https://pubmed.ncbi.nlm.nih.gov/36493795/)]
33. Battelino T, Danne T, Bergenstal RM, et al. Clinical targets for continuous glucose monitoring data interpretation: recommendations from the international consensus on time in range. *Diabetes Care*. Aug 2019;42(8):1593-1603. [doi: [10.2337/dci19-0028](https://doi.org/10.2337/dci19-0028)] [Medline: [31177185](https://pubmed.ncbi.nlm.nih.gov/31177185/)]
34. Service FJ, Molnar GD, Rosevear JW, Ackerman E, Gatewood LC, Taylor WF. Mean amplitude of glycemic excursions, a measure of diabetic instability. *Diabetes*. Sep 1970;19(9):644-655. [doi: [10.2337/diab.19.9.644](https://doi.org/10.2337/diab.19.9.644)] [Medline: [5469118](https://pubmed.ncbi.nlm.nih.gov/5469118/)]
35. Cichosz S, Hangaard S, Kronborg T, Vestergaard P, Jensen MH. From data to insights: a tool for comprehensive Quantification of Continuous Glucose Monitoring (QoCGM). *PeerJ*. 2025;13:e19501. [doi: [10.7717/peerj.19501](https://doi.org/10.7717/peerj.19501)] [Medline: [40511383](https://pubmed.ncbi.nlm.nih.gov/40511383/)]
36. Vergès B, Pignol E, Rouland A, et al. Glycemic variability assessment with a 14-day continuous glucose monitoring system: when and how long to measure MAGE (mean amplitude of glucose excursion) for optimal reliability? *J Diabetes Sci Technol*. Jul 2022;16(4):982-987. [doi: [10.1177/1932296821992060](https://doi.org/10.1177/1932296821992060)] [Medline: [33567877](https://pubmed.ncbi.nlm.nih.gov/33567877/)]
37. Kovatchev BP, Lobo B, Fabris C, et al. The virtual DCCT: adding continuous glucose monitoring to a landmark clinical trial for prediction of microvascular complications. *Diabetes Technol Ther*. Mar 2025;27(3):209-216. [doi: [10.1089/dia.2024.0404](https://doi.org/10.1089/dia.2024.0404)] [Medline: [39772614](https://pubmed.ncbi.nlm.nih.gov/39772614/)]
38. Lobo B, Kanapka L, Kovatchev BP, Kollman C, Beck RW. The association of time-in-range and time-in-tight-range with retinopathy progression in the virtual diabetes control and complications trial continuous glucose monitoring dataset. *Diabetes Technol Ther*. Jul 2025;27(7):558-561. [doi: [10.1089/dia.2025.0033](https://doi.org/10.1089/dia.2025.0033)] [Medline: [39989301](https://pubmed.ncbi.nlm.nih.gov/39989301/)]
39. Elgendi M, Mohamed A, Ward R. Efficient ECG compression and QRS detection for e-health applications. *Sci Rep*. Mar 28, 2017;7(1):459. [doi: [10.1038/s41598-017-00540-x](https://doi.org/10.1038/s41598-017-00540-x)] [Medline: [28352071](https://pubmed.ncbi.nlm.nih.gov/28352071/)]

Abbreviations

CGAN: conditional generative adversarial network
CGM: continuous glucose monitoring
ECG: electrocardiogram
MAE: mean absolute error
MAGE: mean amplitude of glycemic excursions
PAA: piecewise aggregate approximation
PCHIP: piecewise cubic Hermite interpolating polynomial
TAR: time above range
TBR: time below range

Edited by Ariel Teles; peer-reviewed by Chenghao DUAN, Rajiv Ranjan; submitted 22.Jan.2026; final revised version received 14.Apr.2026; accepted 01.May.2026; published 20.Jul.2026

Please cite as:

Bender C, Davidsen L, Olesen SS, Cichosz S

Continuous Glucose Monitoring Data Compression Using Peak-Nadir Encoding in Diabetes: Method Development and Evaluation

JMIR Biomed Eng 2026;11:e91959

URL: <https://biomedeng.jmir.org/2026/1/e91959>

doi: [10.2196/91959](https://doi.org/10.2196/91959)

© Clara Bender, Line Davidsen, Søren Schou Olesen, Simon Cichosz. Originally published in JMIR Biomedical Engineering (<http://biomedeng.jmir.org>), 20.Jul.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Biomedical Engineering, is properly cited. The complete bibliographic information, a link to the original publication on <https://biomedeng.jmir.org>, as well as this copyright and license information must be included.