# FORECASTING EXERCISE LAPSES IN INDIVIDUALS WITH TYPE 1 DIABETES USING STATE SPACE MODELS

Rajat K. Doshi*[1], Yunyang Li*[1], Garrett I. Ash[1], Jason Liu[1], Laura M. Nally[1], Matthew Stults-Kolehmainen[2], Stuart A. Weinzimer[1], Lisa M. Fucito[1], and Mark B. Gerstein[1]

[1]Yale University
[2]Yale New Haven Hospital

## ABSTRACT

Exercise is important for those with type 1 diabetes (T1D), but T1D poses exercise barriers including blood glucose destabilization and fear of hypoglycemia. To investigate these barriers, we measured the day-to-day variation of the barriers and used them to predict the occurrence of days without exercise (i.e., "exercise lapses"). The study participants were 17 adults with T1D without regular exercise routines. They wore biosensors and completed real-time surveys to track exercise, mood, and sleep during 10 weeks of a flexibly-timed, beginner-level home exercise program. We leverage various machine learning techniques, consisting of logistic regression, random forest, time series transformers, and Mamba, a state-of-art state space model (SSM), for forecasting exercise lapses. We demonstrate that we can achieve 75.55 ± 2.6% accuracy with SSM, an improvement over the top baseline accuracy of 72.06 ± 2.9% achieved by classical ML techniques.

## 1 INTRODUCTION

Regular physical exercise is an important component of type 1 diabetes (T1D) management because its cardiometabolic benefits offset the 8-fold increased cardiovascular disease risk posed by T1D Schofield et al. (2019). The recommended exercise frequency is daily (or as close to daily as possible) to regularize its impact upon diurnal blood glucose patterns Association (2023). Unfortunately, just 18%-33% of people with T1D meet these recommended exercise goals and this rate declines with age McCarthy et al. (2016). Cross-sectional surveys Brazeau et al. (2008); Dubé et al. (2006) indicate that a major barrier to exercise with T1D is the glucose management challenges. No studies have addressed the day-to-day time sequence between these barriers and exercise among people with T1D, even though blood glucose, fear of hypoglycemia, and exercise engagement have been observed to fluctuate daily Martyn-Nemeth et al. (2017).

The emergence of deep learning techniques has revolutionized many fields, including healthcare, by providing powerful tools for behavioral prediction and personalized interventions. Among these advancements, transformers Vaswani et al. (2023) have contributed significantly in time series forecasting task. However, despite their success, the computational demands of transformers limit their applicability in real-time environments, a critical consideration for technologies designed to support individuals with T1D in managing exercise-related glucose dynamics.

In response to these challenges, we explore Mamba Gu & Dao (2023), a novel deep learning architecture based on state space models (SSMs) Gu et al. (2021). The architecture is designed to filter inputs based on their relevance, significantly improving processing speeds - it allows SSM to achieve inference times on average about 5 times faster than traditional transformer models, presenting a significant advancement in the development of real-time glucose management tools.

The main contributions of the paper are:

*These authors contributed equally to this work.

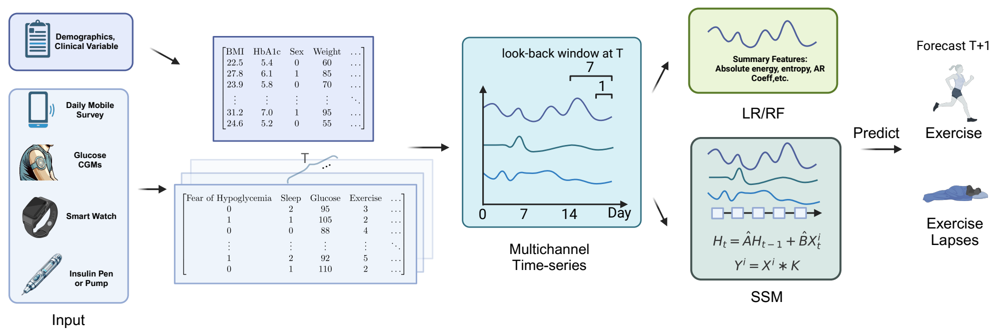


Figure 1: Forecasting Overview: We established a ML workflow for predicting exercise behaviors in T1D individuals. Firstly, data collection and preprocessing involve four main feature types: covariates, survey data, continuous glucose readings, and their temporal interaction with exercise and insulin dosing. Next, feature engineering includes utilizing comprehensive statistics to generate 1000 features for machine learning modeling or a 7-day lookback "rolling-window" technique for sequential modeling. For modeling, we employed a state-space model, which adopts a dual-form of recurrent neural network, and convolutional neural network. Finally, the objective is to predict the exercise behavior on the next day.

- **Evaluation of Mamba's Predictive Performance**: Mamba has predominantly been utilized in the field of NLP. However, this study pioneers the exploration of Mamba's potential in predicting health-related time series data. Furthermore, this research compares Mamba's effectiveness with that of other widely adopted modeling frameworks, providing a benchmark for its performance in this novel application area.
- **Use of ML for T1D Exercise Behavior Prediction**: This study is among the first to utilize deep learning techniques to predict exercise lapses in individuals with type 1 diabetes, addressing the dynamic interplay between daily variations in blood glucose levels, mood, sleep, and their influence on exercise adherence.
- **Insights into Determinants of Exercise Engagement**: Our analysis reveals how specific glucose and survey measurements can potentially indicate an increased or decreased likelihood of exercising on any given day.

## 2 Methods

### 2.1 Prliminary

**Exercise Lapses Forecasting (ELF)**: Given a wearable biosensor time series with channel set $\mathcal{D} = \{1, 2, ..., d\}$, the historical data is represented by $\mathcal{X} = \{X_t^1, ..., X_t^d\}_{t=1}^L$, where $L$ is the look-back window size and $X_t^i$ is the value of the $i$-th channel at the $t$-th time step. Let $e \in \mathcal{D}$ denotes the channel for exercise behavior, and $X^e \in \{0, 1\}$, where 0 indicates exercise lapses and 1 represents active exercise. The goal of ELF is to find a model function $f : \mathcal{X} \to \hat{\mathcal{Y}} := \{\hat{X}_t^e\}_{t=L+1}^{L+T}$ for future $T$ steps.

**State Space Models for Time Series Representation**: State Space Models (SSMs), exemplified by the Structured State Space Sequence models (S4) and Mamba, are predicated on mapping a time series $X_t^i \in \mathbb{R}$ to an output $Y_t^i \in \mathbb{R}$, facilitated through an intermediary latent state $H_t \in \mathbb{R}^n$. The models operationalize $A \in \mathbb{R}^{n \times n}$ as the state transition matrix, with $B \in \mathbb{R}^{n \times 1}$ and $C \in \mathbb{R}^{1 \times n}$ serving as input and output projection matrices, respectively. The dynamics of the system are given as:

$$H_t' = AH_t + BX_t^i, \quad Y_t^i = CH_t, \tag{1}$$

In adapting these principles to discrete settings, a scaling parameter $\Delta$ is introduced to convert the continuous-time matrices $A$ and $B$ into their discrete-time analogues $\hat{A}$ and $\hat{B}$ through a discretization process which could be written as:

$$\hat{A} = \exp(\Delta A), \quad \hat{B} = A^{-1}(\exp(\Delta A) - I)B \cdot \Delta, \tag{2}$$

where $I$ denotes the identity matrix. The discrete system's dynamics, alongside the mechanism for output computation via global convolution, could be reformulated as:

$$H_t = \hat{A}H_{t-1} + \hat{B}X_t^i, \quad Y^i = X^i * K, \tag{3}$$

where $K = (C\hat{B}, C\hat{A}\hat{B}, \ldots, C\hat{A}^{L-1}\hat{B})$ with $L$ indicating the length of the input sequence, $*$ is the convolution operator and $K \in \mathbb{R}^L$ representing the structured convolutional kernel. The convolutional formulation of SSMs allows for efficient training, while the recurrent formulation enables fast inference. This is particularly beneficial for mobile applications, facilitating online training and real-time inference.

### 2.2 Data Preprocessing/Feature Engineering

To preprocess the data for training, we first applied normalization to scale the biosensor and survey data, ensuring uniformity in value ranges. Approximately 14.3% of data entries, which had missing journal entries or non-wear of CGM, were systematically omitted to maintain dataset integrity. Given the observed class imbalance (41.3% exercise days and 58.7% non-exercise days (i.e., exercise lapses), we applied the Synthetic Minority Over-sampling Technique (SMOTE) Chawla et al. (2002) to up-sample the minority class. For train-test split, we stratified the dataset by participant. We randomly assigned all the time series data of a participant to a fold and then used a 5-fold Cross Validation strategy during training.

We designed four distinct feature sets for analytical purposes. The first set consisted exclusively of individual-level covariates to establish a baseline. The second set expanded upon this by adding both individual-level covariates and raw features, including the previous day's survey data, glucose measurements, exercise activities, and glucose-exercise interactions (Appendix Table 2; left part of Figure 1). The third set leveraged TSFresh library Christ et al. (2018) to extract 5065 features from the raw data based on their temporal and statistical patterns (Appendix Table 2; upper green area in Figure 1). A univariate analysis, with a false discovery rate of 0.05, narrowed these down to 937 significant features. To mitigate the observed high collinearity among these features, Principal Component Analysis (PCA) was employed, identifying approximately 110 principal components representing 95% of the variance. These components, alongside demographic covariates, were incorporated into our models. The fourth feature set applied a rolling window method with a 7-day interval to accurately assess exercise behavior Laeschke et al. (2018).

### 2.3 Modeling and Baselines

For the ELF task, we used a particular implementation of SSMs called Mamba Gu & Dao (2023). Mamba incorporates time-varying parameters into the SSM and proposes a hardware-aware algorithm for effective training and inference. Mamba's impressive scaling performance shows its potential as an alternative to the Transformer in real-time exercise lapses forecasting systems.

For baselines, we implemented a Time Series Transformer Zerveas et al. (2020) customized for time series classification. This architecture commence with an input encoding layer that maps the time-series feature vector to a high-dimensional space. It is then followed by a positional encoder (PE) to account for temporal dynamics. Subsequently, a multi-head self-attention mechanism and a stacked encoder layer are employed for hierarchical feature extraction. Lastly, a classification head is used specifically for ELF.

In addition to the Time Series Transformer, we also utilized traditional machine learning models, namely logistic regression (LR) and random forest (RF), as part of our comparative baseline. These models, although less complex, offer a high degree of interpretability and play a crucial role in clinical applications.

# 3 RESULTS

## 3.1 MODEL PERFORMANCE

Several combinations of feature engineering and modeling were used to assess the effectiveness of forecasting exercise behaviors based on individuals' wearable data, demographics, and mood survey data. Conventional models (LR, RF) were fed with covariates only, covariates plus raw features for either 1 or 7 days back from the day being predicted, or covariates plus TSFresh features. Sequential models (TST, SSM) were fed with the 7-day lookback rolling window. The findings are summarized in Table 1.

Table 1: Performance metrics for different models and feature sets

| Feature Set & Model | Accuracy | AUC | Sensitivity | Specificity |
|---|---|---|---|---|
| Covariates Only - LR | $0.6009 \pm 0.037$ | $0.5501 \pm 0.039$ | $0.6422 \pm 0.028$ | $0.4278 \pm 0.016$ |
| Covariates Only - RF | $0.5899 \pm 0.017$ | $0.554 \pm 0.029$ | $0.4504 \pm 0.022$ | $0.6536 \pm 0.046$ |
| Covariates + 1 Day Back - LR | $0.7011 \pm 0.029$ | $0.6624 \pm 0.032$ | $0.5653 \pm 0.043$ | $0.7959 \pm 0.014$ |
| Covariates + 1 Day Back - RF | $0.7116 \pm 0.048$ | $0.6920 \pm 0.030$ | $0.5419 \pm 0.033$ | $\mathbf{0.8323 \pm 0.045}$ |
| Covariates + TSFresh - LR | $0.7206 \pm 0.029$ | $0.7729 \pm 0.045$ | $0.5601 \pm 0.029$ | $\mathbf{0.8323 \pm 0.049}$ |
| Covariates + TSFresh - RF | $0.7176 \pm 0.022$ | $0.7502 \pm 0.041$ | $0.5550 \pm 0.016$ | $0.8286 \pm 0.042$ |
| Covariates + 7 Days Back - LR | $0.6941 \pm 0.042$ | $0.7478 \pm 0.020$ | $0.5468 \pm 0.027$ | $0.7973 \pm 0.028$ |
| Covariates + 7 Days Back - RF | $0.6784 \pm 0.044$ | $0.7282 \pm 0.049$ | $0.5407 \pm 0.044$ | $0.7739 \pm 0.040$ |
| Covariates + 7 Days Back - TST | $0.7440 \pm 0.033$ | $0.7902 \pm 0.027$ | $0.7774 \pm 0.012$ | $0.6946 \pm 0.013$ |
| Covariates + 7 Days Back - Mamba | $\mathbf{0.7549 \pm 0.026}$ | $\mathbf{0.7945 \pm 0.032}$ | $\mathbf{0.7951 \pm 0.023}$ | $0.6995 \pm 0.026$ |

Our analysis reveals that using an SSM was the most effective model for forecasting exercise lapses based on AUC and accuracy. In comparison to the traditional machine learning methods, it had a higher sensitivity. This suggests that SSM predicted a greater proportion of the exercise lapses (i.e., points of behavioral vulnerability) Nahum-Shani et al. (2018). For LR and RF, we saw between a 6% to 10% increase in RF and LR's AUC after leveraging TSFresh's feature extraction functions and then running PCA on the outputs. This implies that, from a data-centric machine learning viewpoint, incorporating features derived from wearable biosensors could effectively bolster the model's predictive performance.

## 3.2 FEATURE IMPORTANCE

To understand clinical implications of the models, we conducted an explainability analysis on RF and LR leveraging the 1-day look-back window feature set. Through SHAP (SHapley Additive exPlanations) analysis Lundberg & Lee (2017), we successfully identified both individual-level covariates and daily factors that predict the likelihood of engaging in exercise activities, as detailed in Appendix Figure 2. Notably, lower Body Mass Index (BMI) and younger age were among the individual-level predictors of increased exercise frequency throughout the study period. Furthermore, daily variables such as improved sleep quality, reduced evening fear of hypoglycemia, and minimized daytime experiences of hyperglycemia were associated with a heightened probability of exercise in the subsequent 1-7 days. Similarly, the LR model also highlighted the significance of both individual-level covariates and daily factors in predicting exercise activities, with a particular focus on the patterns of blood glucose levels in relation to exercise, as indicated in Appendix Table 2.

# 4 DISCUSSION

In conclusion, machine learning models that leverage biosensor, mood, and sleep data were used to predict days of non-exercise among adults with T1D, based on information collected leading up to the day. These time-series deep learning models (SSM, TST) could be the starting algorithm for a just-in-time adaptive intervention mobile app, that could subsequently be trained upon individual users to anticipate their personal days of vulnerability to missing exercise and send encouraging messages on the mornings of those days. Given SSM success in this task and its higher efficiency in memory usage over transformer models, we envision the future of edge computing Hanzelik et al. (2022) to involve state-of-art state space models like Mamba.

## 5 LIMITATION

A limitation of our study is the relatively small amount of available data, with just over 1,000 training instances, which may impact the generalizability and robustness of the machine learning models. Additionally, although there were missing data entries (approximately 14.3% of entries), we assumed missingness at random which may not have been the case. Future studies can address the limitations of our work by incorporating datasets with more datapoints, which would enhance the generalizability and robustness of machine learning models and add power to address variance. It could be similarly useful to address additional variables. For instance, hypoglycemia was assessed both objectively and for psychological sequalae, while hyperglycemia was only assessed objectively. Yet both were predictors of exercise lapses so psychological sequalae of hyperglycemia could be a valuable additional variable even though it has only been minimally addressed in the literature Singh et al. (2014); Polonsky et al. (2021); Lin et al. (2022).

## 6 ACKNOWLEDGMENTS

We acknowledge support from the National Institutes of Health and from the AL Williams Professorship funds. As well, the dataset acquisition and author GIA's time were supported by a Robert E. Leet and Clara Guthrie Patterson Trust Mentored Research Award, Bank of America, N.A., Trustee, American Heart Association Grant 852679 (2021–2024), and the National Institute of Diabetes, Digestive, and Kidney Diseases of the National Institutes of Health under a mentored research scientist development award (K01DK129441). Author LMN's time was supported by the National Institute of Diabetes, Digestive, and Kidney Diseases of the National Institutes of Health under a mentored research scientist development award (K23DK128560). Author LMF's time was supported by the National Institute on Alcohol Abuse and Alcoholism (R01AA030136). The funders played no role in study design, data collection, analysis and interpretation of data, or the writing of this manuscript.

## REFERENCES

GI Ash, LM Nally, M Stults-Kolehmainen, et al. Personalized digital health information to substantiate human-delivered exercise support for adults with type 1 diabetes. *Clin J Sport Med*, 2023. doi: 10.1097/JSM.0000000000001078. Online ahead of print.

American Diabetes Association. Standards of care in diabetes - 2023, 2023.

T Battelino, CM Alexander, SA Amiel, et al. Continuous glucose monitoring and metrics for clinical trials: an international consensus statement. *Lancet Diabetes Endocrinol*, 11(1):42–57, 2023. doi: 10.1016/S2213-8587(22)00319-9.

AS Brazeau, R Rabasa-Lhoret, I Strychar, and H Mircescu. Barriers to physical activity among patients with type 1 diabetes. *Diabetes Care*, 31(11):2108–2109, 2008. doi: 10.2337/dc08-0720. Epub 2008 Aug 8. PMID: 18689694; PMCID: PMC2571055.

NV Chawla, LO Hall, and WP Kegelmeyer. Smote: Synthetic minority over-sampling technique, 2002.

M Christ, N Braun, J Neuffer, and AW Kempa-Liehr. Time series feature extraction on basis of scalable hypothesis tests (tsfresh – a python package), 2018.

MC Dubé, P Valois, D Prud'homme, SJ Weisnagel, and C Lavoie. Physical activity barriers in diabetes: development and validation of a new scale. *Diabetes Research and Clinical Practice*, 72(1):20–27, 2006. doi: 10.1016/j.diabres.2005.08.008. Epub 2005 Oct 26. PMID: 16256239.

A Gu and T Dao. Mamba: Linear-time sequence modeling with selective state spaces, 2023.

A Gu, K Goel, and C Ré. Efficiently modeling long sequences with structured state spaces. *arXiv preprint arXiv:2111.00396*, 2021.

PP Hanzelik, A Kummer, and J Abonyi. Edge-computing and machine-learning-based framework for software sensor development. *Sensors*, 22:4268, 2022. doi: 10.3390/s22114268.

L Laeschke, A Steinbrecher, S Jeran, et al. Variability and reliability study of overall physical activity and activity intensity levels using 24 h-accelerometry-assessed data. *BMC Public Health*, 18:530, 2018. doi: 10.1186/s12889-018-5415-8.

YK Lin, CR Richardson, I Dobrin, et al. Beliefs around hypoglycemia and their impacts on hypoglycemia outcomes in individuals with type 1 diabetes and high risks for hypoglycemia despite using advanced diabetes technologies. *Diabetes Care*, 45(3):520–528, 2022. doi: 10.2337/dc21-1285.

SM Lundberg and S Lee. A unified approach to interpreting model predictions. *Advances in neural information processing systems*, 30, 2017.

P Martyn-Nemeth, L Quinn, S Penckofer, C Park, V Hofer, and L Burke. Fear of hypoglycemia: Influence on glycemic variability and self-management behavior in young adults with type 1 diabetes, 2017. Epub 2017 Jan 20. PMID: 28143733; PMCID: PMC5350014.

MM McCarthy, R Whittemore, and M Grey. Physical activity in adults with type 1 diabetes. *Diabetes Educator*, 42(1):108–115, 2016. doi: 10.1177/0145721715620021. Epub 2015 Dec 8. PMID: 26655391; PMCID: PMC6713193.

TH Monk, CF Reynolds, DJ Kupfer, et al. The pittsburgh sleep diary. *Journal of Sleep Research*, 3: 111–120, 1994.

I Nahum-Shani, SN Smith, BJ Spring, LM Collins, K Witkiewitz, A Tewari, and SA Murphy. Just-in-time adaptive interventions (jitais) in mobile health: Key components and design principles for ongoing health behavior support. *Annals of Behavioral Medicine*, 52(6):446–462, May 18 2018. doi: 10.1007/s12160-016-9830-8. PMID: 27663578; PMCID: PMC5364076.

WH Polonsky, AL Fortmann, D Price, and L Fisher. ”hyperglycemia aversiveness”: Investigating an overlooked problem among adults with type 1 diabetes. *J Diabetes Complications*, 35(6):107925, 2021. doi: 10.1016/j.jdiacomp.2021.107925.

J Schofield, J Ho, and H Soran. Cardiovascular risk in type 1 diabetes mellitus, 2019.

H Singh, L Gonder-Frederick, and Schmidt. Assessing hyperglycemia avoidance in people with type 1 diabetes. *Diabetes Manag*, 4(3):263–271, 2014.

S Tonstad, P Herring, J Lee, and JD Johnson. Two physical activity measures: Paffenbarger physical activity questionnaire versus aerobics center longitudinal study as predictors of adult-onset type 2 diabetes in a follow-up study. *American Journal of Health Promotion : AJHP*, 32(4):1070–1077, 2018. doi: 10.1177/0890117117725282.

A Vaswani, N Shazeer, N Parmar, J Uszkoreit, L Jones, AN Gomez, L Kaiser, and I Polosukhin. Attention is all you need, 2023.

G Zerveas, S Jayaraman, D Patel, A Bhamidipaty, and C Eickhoff. A transformer-based framework for multivariate time series representation learning, 2020.

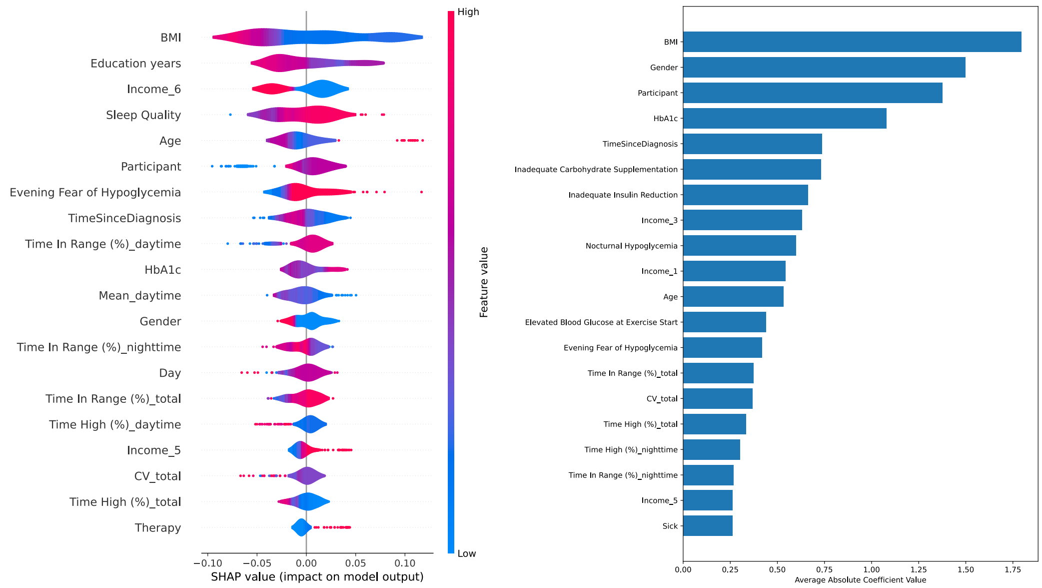


Figure 2: On the left, RF feature importance (fold 1) expressed as a Shapley Plot. On the right, LR feature importance (average of 5 folds) expressed as absolute value of the coefficient.

# APPENDIX

## BIOMETRIC OBSERVATION DATA CAPTURE:

The trial intervention (NCT04204733)Ash et al. (2023) provided participants with exercise videos and biometric-based coaching, both supplied by GlucoseZone.com (Fitscript LLC, New Haven, CT). Recruitment (December 2019 – August 2020) used social media, clinic registry, advertisements, and previous trial participants. The inclusion criteria were: 18-65 years old with ≥6mo diagnosis of T1D or other absolute insulin deficiency diabetes, inadequate baseline exercise patterns (<3 exercise sessions/wk) Tonstad et al. (2018), English literacy, current user of a smartphone and continuous glucose monitor (CGM). Exclusion criteria were chronic disease or physical disability requiring exercise adjustments outside the scope of the digital exercise guidance platform.

Observational data capture included: **(1) Exercise**. Apple Watch 3 (Cupertino, CA) heart rate ≥50% age-predicted maximum while viewing exercise video (subdivided as moderate if heart rate 50%-79% maximum or resistance exercise in beginner-intermediate difficulty category assigned by producer [GlucoseZone.com, New Haven, CT], vigorous if heart rate ≥80% or resistance exercise in advanced category). Most (99%) of the bouts amounted to ≥30 metabolic equivalent (MET)-minutes, meaning they required as much or more energy as 10 minutes of brisk walking. **(2) Blood glucose.** Participants used their own CGM device which was the Dexcom (San Diego, CA) G6 (40%), G5 (10%), G4 (5%), Medtronic (Dublin, Ireland) 670G (35%), and Abbott (Chicago, IL) Freestyle Libre (10%). Values were measured every 5min for the Dexcom and Medtronic devices, and every 15min for the Abbott. They were captured and processed by Dexcom Clarity, Medtronic CareLink, or Abbott Libreview Pro respectively, and for each day 24hr summary metrics were recorded including mean, variability (coefficient of variation), % of time hyperglycemic (¿180 mg/dL), % of time in clinical target range (70-180 mg/dL), and % of time hypoglycemic (¡70 mg/dL) Battelino et al. (2023). Completeness was 90% (SD 9%) bypca participant. **(3) Ecological momentary** (i.e., real-time) surveys. Qualtrics (Provo, UT) smartphone survey each morning at self-expected waketime (response delay ≤3hr allowed) prompted participant to report previous night sleep quality (1-10 scale) Monk et al. (1994) and overnight fear of hypoglycemia (1-5 scale) Martyn-Nemeth et al. (2017), current fear of hypoglycemia for the coming day (1-5 scale) Martyn-Nemeth et al. (2017), and illness (yes/no).

Table 2: Descriptive statistics of variables used for features. Expressed as mean±standard deviaition if normally distributed, median (25th, 75th %'ile) if skewed, frequency if binary.

| **Metric** | **Definition** | **Day-Level Participant-days (k=1,020)** | **Participant-Level (i.e., weighted by participant) Participants (n=17)** | **Standard Clinical Recommendation** |
|---|---|---|---|---|
| Blood glucose patterns, 24hr | | | | |
| Variability | 24hr Coefficient of variation (%) | 26.3 (21.5, 31.6) | 26.9 (21.7, 32.9) | $\leq 36\%$ |
| Time in hyperglycemia | 24hr time >180 mg/dL (%) | 23.9 (10.7, 43.1) | 26.6 (11.0, 49.5) | $\leq 25\%$ |
| Time in target range | 24hr time 70-180 mg/dL (%) | 72.2 (49.17, 85.9) | 74.4 (50.2, 87.0) | $\geq 70\%$ |
| Time in hypoglycemia | 24hr time <70 mg/dL (%) | 0.0 (0,15.3) | 0.0 (0, 2.2) | $\leq 4\%$ |
| Mean | 24hr mean (mg/dL) | 145 (126, 174) | 148 (127, 183) | $\leq 154$ |
| Blood glucose patterns, around exercise | | Total occurrences (% of the 421 exercise days) | Total participants with at least one occurrence (% of the 17 participants) | |
| #1: Inadequate Carbohydrate Supplementation | Blood glucose <70 mg/dL at the start of exercise OR (blood glucose 70-100 mg/dL at the start of exercise AND <70 mg/dL during or within 1hr after exercise) | 22 (5.2%) | 5 (29.4%) | To avoid |
| #2: Nocturnal Hypoglycemia | Blood glucose <70 mg/dL for 30min of consecutive nocturnal readings after day with exercise | 17 (4.0%) | 5 (29.4%) | To avoid |
| #3: Inadequate Insulin Reduction | Insulin bolus (observed on pump or Bluetooth pen) <120min before start of exercise AND blood glucose <70 mg/dL during or within 1hr after exercise | 14 (3.3%) | 6 (35.3%) | To avoid |
| #4: Elevated Blood Glucose at Exercise Start | Momentary surveys (i.e., ecological momentary assessment) | 7 (1.7%) | 3 (17.6%) | To avoid |
| Survey Values | | | | |
| Sick Day | Yes or No | 71 (16.9%) | N/A | |
| Morning Fear of Hypoglycemia | 1 – 5 Likert scale | 154 (36.6%) cases with values $> 1$ | 1.0 (1.0, 2.1) | N/A |
| Evening Fear of Hypoglycemia | 1 – 5 Likert scale | 141 (33.5%) cases with values $> 1$ | 1.0 (1.0, 2.1) | N/A |
| Sleep Quality | 1–10 Likert scale | $7.0 \pm 2.0$ | $7.2 \pm 1.7$ | N/A |
| Demographics | | | | |
| Gender | | | 55% female | N/A |
| Age (years) | | | $42.3 \pm 15.0$ | N/A |
| T1D duration (years) | | | $20.5 \pm 15.3$ | N/A |
| T1D therapy modality | | | 85% continuous subcutaneous insulin infusion pump | N/A |
| | | | 15% multiple daily injections | N/A |
| BMI (kg/m$^2$) | | | $29.5 \pm 5.1$ | 18.5 to 24.9 |
| HbA1c (%) | | | $7.2 \pm 1.1$ | <7.0 |
| Systolic blood pressure (mmHg) | | | $123 \pm 16$ | <120 |
| Diastolic blood pressure (mmHg) | | | $77 \pm 9$ | <80 |
| Race/Ethnicity | | | 95% non-Hispanic white | N/A |
| Income annual | | | 15% <\$50,000 | N/A |
| | | | 35% \$50,000 - \$79,999 | N/A |
| | | | 50% ≥\$80,000 | N/A |

§Three participants (15%) were excluded (insulin pump download error, manual insulin logging with unreasonable timestamps).
§From remainder, 161 (14%) of person-days were excluded for missing survey or CGM data leaving 1,020 person-days for analysis.

**Covariance With TSFresh:**

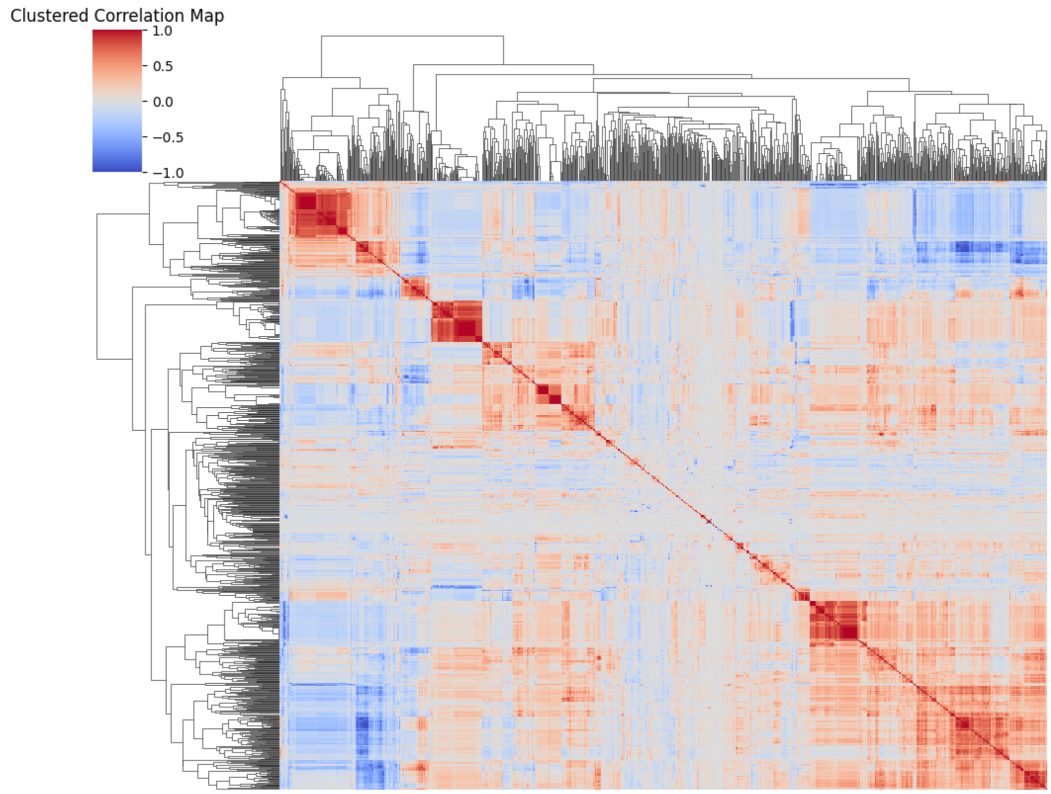


Figure 3: Heat map with clustering of TSFresh features. The presence of several distinct clusters indicates some highly collinear features, supporting the decision to do feature reduction.

**PCA**

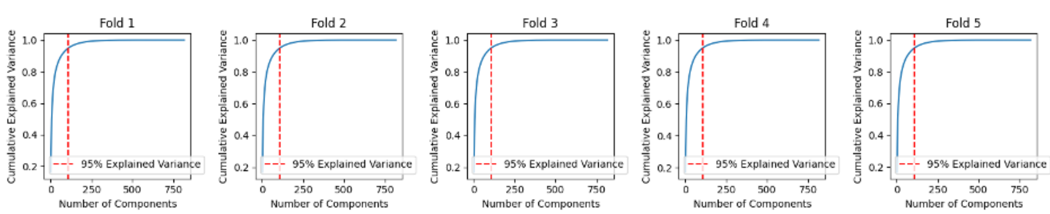


Figure 4: We used PCA to reduce collinearity among the initial 1000 features generated from TSFresh and we found that 109, 108, 107, 106, and 108 components were needed to capture 95% of the variance, respectively.

**Time Series Transformer: Confusion Matrix**

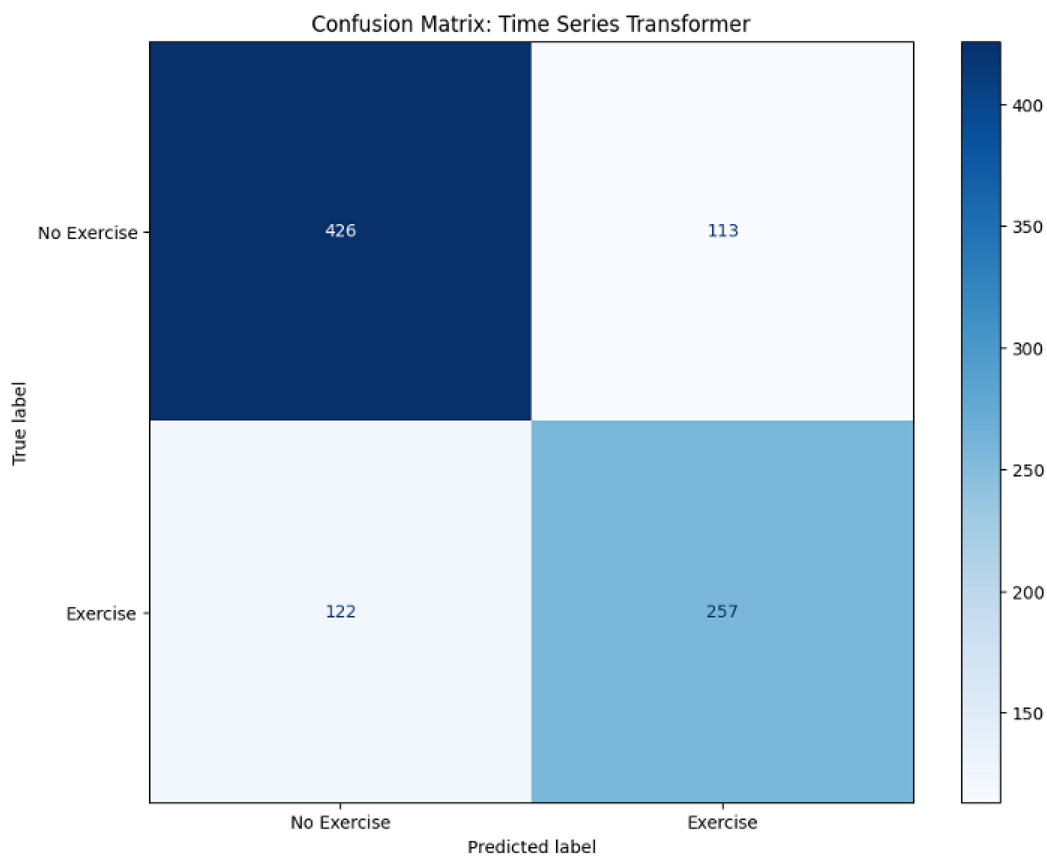


Figure 5: Confusion matrix of time series transformer.

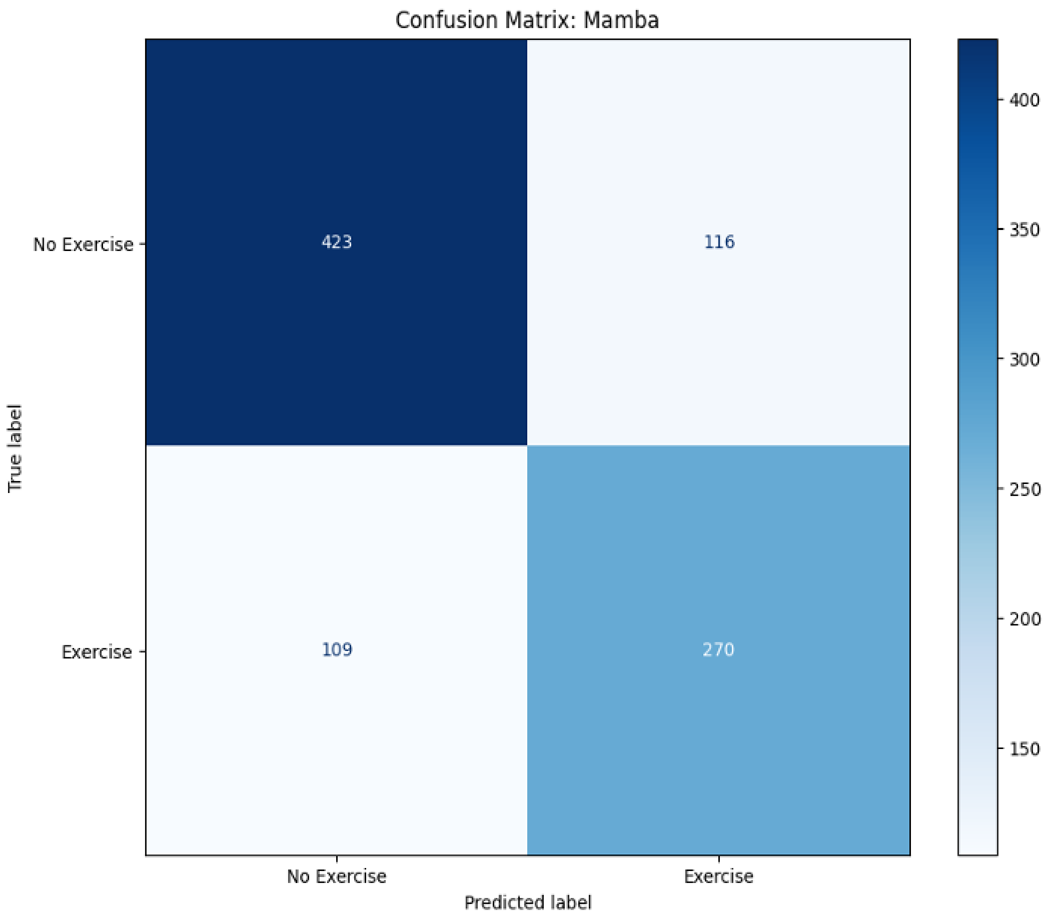


Figure 6: Confusion matrix of mamba.

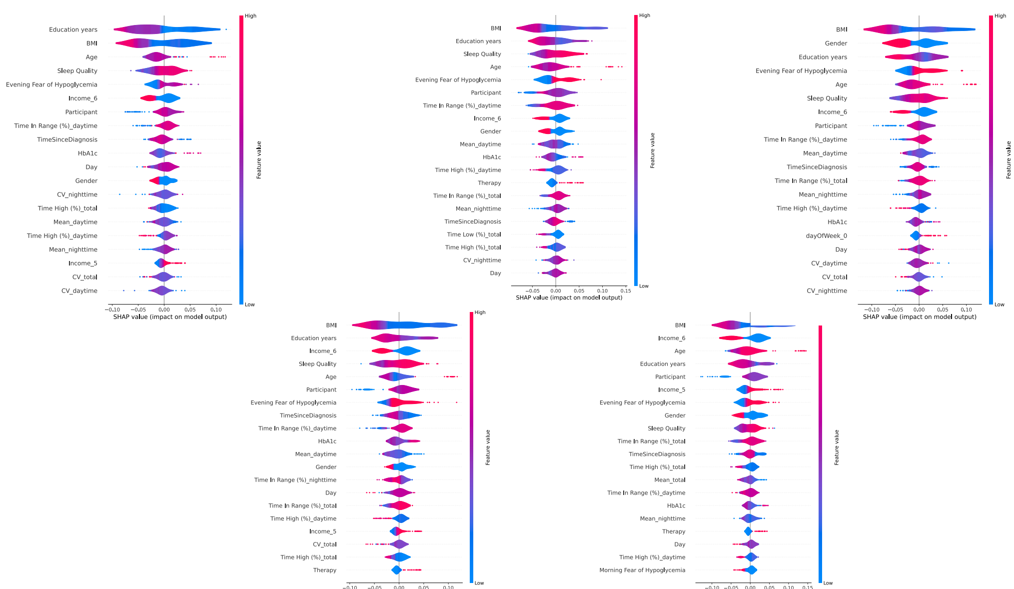

Figure 7: Random forest Shap Plots across each of the 5 folds

Table 3: A comprehensive collection of automatically extracted time series features from TSFresh that encompass statistical, energy-based, distributional, and structural characteristics.

| Class | Feature | Description |
|---|---|---|
| Basic Statistics | mean(x)<br>std(x)<br>max(x)<br>min(x) | Mean of x<br>Standard deviation of x<br>Highest value in x<br>Lowest value in x |
| Distribution | variance(x)<br>skewness(x)<br>kurtosis(x)<br>median(x)<br>absolute_maximum(x)<br>absolute_sum_of_changes(x) | Variance of x<br>Skewness of x<br>Kurtosis of x<br>Median of x<br>Highest absolute value in x<br>Sum of absolute value of consecutive changes in x |
| Energy | abs_energy(x)<br>sum_values(x)<br>root_mean_square(x)<br>sum_of_reoccurring_data_points(x)<br>sum_of_reoccurring_values(x)<br>percentage_of_reoccurring_values(x) | Absolute energy (sum of squared values)<br>Sum of values in x<br>Root mean square (rms) of x<br>Sum of data points present more than once<br>Sum of values present more than once<br>Percentage of values occurring more than once |
| Time Series Characteristics | length(x)<br>has_duplicate(x)<br>has_duplicate_max(x)<br>has_duplicate_min(x)<br>variation_coefficient(x)<br>symmetry_looking(x) | Length of x<br>Checks for duplicate values in x<br>Checks if the maximum value is observed more than once<br>Checks if the minimum value is observed more than once<br>Variation coefficient (std. error / mean) of x<br>Indicates if the distribution of x is symmetric |
| Peaks and Trends | number_peaks(x, n)<br>linear_trend(x, param)<br>linear_trend_timewise(x, param)<br>longest_strike_above_mean(x) | Number of peaks with support n<br>Linear trend over the entire series<br>Linear trend over time<br>Length of consecutive subsequence above mean |
| Entropy and Complexity | permutation_entropy(x, tau, dimension)<br>binned_entropy(x, max_bins)<br>lempel_ziv_complexity(x, bins)<br>fourier_entropy(x, bins)<br>sample_entropy(x) | Permutation entropy<br>Binned entropy<br>Lempel-Ziv complexity<br>Fourier entropy<br>Sample entropy |

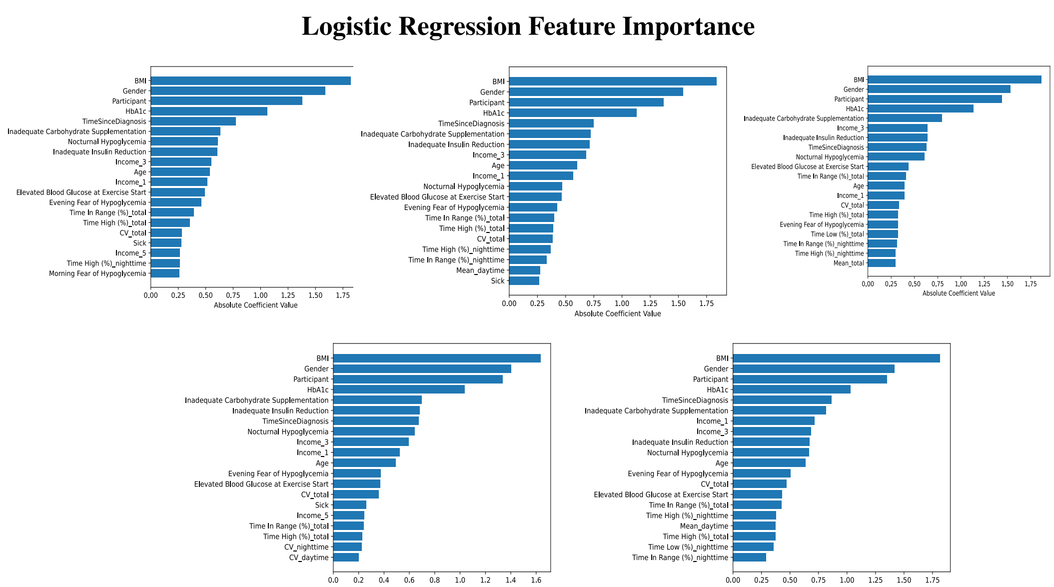


Figure 8: Logistic regression feature importance across each of the 5 folds. Average of the 5 folds shown Figure 2.