Decoding Type 2 Diabetes Progression via Metabolic Hormone Time-Series

Hesam Abouali
Electrical and Computer Engineering
University of Waterloo
hesam.abouali@uwaterloo.ca

Sanjana Srikant
Electrical and Computer Engineering
University of Waterloo
sanjana.srikant@uwaterloo.ca

Nicole G. Barra
Department of Biochemistry and Biomedical Sciences
McMaster University
barrang@mcmaster.ca

Ali Etemad
Electrical and Computer Engineering
Queen's University
ali.etemad@queensu.ca

Jonathan D. Schertzer
Department of Biochemistry and Biomedical Sciences
McMaster University
schertze@mcmaster.ca

Mahla Poudineh
Electrical and Computer Engineering
University of Waterloo
mahla.poudineh@uwaterloo.ca

Abstract

Unraveling the dynamics of diabetes hormones and peptides, including insulin, glucagon, and C-peptide, is important for understanding diabetes progression and for personalized treatment. The current gold standard methods to monitor these biomarkers have a prolonged processing time, and their protocols do not allow real-time data collection. We recently developed the quantum dot–integrated real-time ELISA (QIRT-ELISA), a continuous monitoring platform that quantifies these key biomarkers at minute-by-minute resolution. Here, we leverage QIRT-ELISA's high temporal resolution for real-time, multiplexed monitoring of insulin, glucagon, and C-peptide in animal models of prediabetes and type 2 diabetes, generating the first *in vivo*, hormone time-series dataset. Using the raw QIRT-ELISA time series, we trained and evaluated four classical machine-learning classifiers to classify metabolic phenotypes, a capability not achievable with conventional assays or with unprocessed QIRT-ELISA outputs. Among these classifiers, logistic regression performed the best, achieving 89% accuracy in distinguishing healthy, prediabetic, and diabetic states. QIRT-ELISA's capacity to capture high-resolution, time-series hormone data, along with machine-learning–based classification, has the potential to enable phenotypic characterization and early detection of metabolic dysfunction in diabetes.

1 Introduction

Type 2 diabetes (T2D) is a chronic disease in which the regulatory mechanisms of glucose homeostasis become imbalanced [1, 2]. Insulin, secreted by pancreatic β-cells, regulates blood glucose by promoting cellular glucose uptake, particularly in the liver and skeletal muscle. In addition to insulin, other hormones and peptides coordinate blood-glucose regulation [3, 4]. For example, C-peptide (connecting peptide) is co-secreted with insulin in equimolar amounts but has a longer half-life, making it a robust marker of insulin secretion and β-cell function [5, 6]. Glucagon, another key

39th Conference on Neural Information Processing Systems (NeurIPS 2025) Workshop: Learning from Time Series for Health.

regulator of glucose homeostasis, is produced primarily by pancreatic α-cells; its principal role is to prevent low blood glucose conditions by releasing stored glucose into the bloodstream, thereby counteracting insulin's actions [7, 8]. In T2D, pancreatic β-cell function declines—reducing insulin and C-peptide secretion—while α-cell dysregulation leads to inappropriate glucagon secretion. Thus, understanding how insulin, glucagon, and C-peptide evolve during T2D progression will provide fundamental insight into how factors beyond blood glucose drive disease progression [9, 10, 11, 8].

Blood glucose can be continuously measured using continuous glucose monitoring (CGM) devices and the time-course of changes in blood glucose before and after a meal is well-established [12, 13, 14, 15]. However, compared with glucose, the temporal dynamics of circulating insulin, glucagon, and C-peptide remain largely unmeasured, mainly due to technological limitations. Gold standard methods—such as enzyme-linked immunosorbent assays (ELISAs)—require bulky, specialized instrumentation, extensive sample pre-processing, and hours per run, severely limiting the acquisition of high-frequency time-series data for these hormones and peptides [16]. Additionally, technologies such as CGMs are enzyme-based (e.g., glucose oxidase) [17] and are not readily adaptable to insulin, glucagon, or C-peptide, for which no suitable catalytic enzymes are available.

To address this need, we recently developed QIRT-ELISA [18], a continuous monitoring platform that provides minute-by-minute, multiplexed measurements of insulin and glucagon with high sensitivity directly from whole blood in rat models. The QIRT-ELISA platform employs quantum dot-based fluorescence immunoassay in a multi-module microfluidic system to enable continuous, multiplexed measurements without any sample pre-processing in whole blood in conscious rat models.

The publication, which includes the detailed information about this platform and the immunoassays for detection of insulin and glucagon, can be found in [18]. In the present work, we expanded QIRT-ELISA to enable continuous C-peptide quantification, complementing the existing insulin and glucagon assays, and tested its enhanced temporal resolution for monitoring these hormones/peptides during a fed-state experiment in which conscious rats did not have to be fasted overnight in contrast to traditional glucose tolerance test (GTT) (Fig. 1) [19, 20]. Experiments were performed in three groups of rats: healthy (control), obese (prediabetic), and T2D. We implemented a data-analysis workflow to extract dynamic features from insulin, glucagon, and C-peptide time-series (details available in appendix A.1 and measurements available in appendix A.2). We then applied machine learning techniques, including logistic (LR), random forest (RF), k-nearest neighbors (KNN), support vector machines (SVM) to classify phenotypes in the animal models. Our results show that QIRT-ELISA enables high temporal resolution hormone monitoring, revealing *in vivo* dynamic responses that are otherwise inaccessible with standard methods such ELISA with longer time intervals for measurements (Fig. 1).

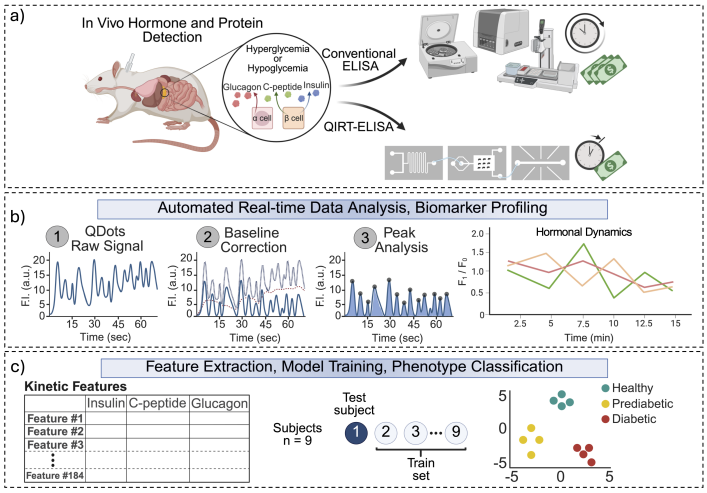

Figure 1: Overview. a) The schematic of experiments performed using conventional ELISA and QIRT-ELISA. b) The fluorescence measurements from QIRT-ELISA are processed, and are used for profiling. c) Multiple features of the hormonal dynamics are extracted and machine learning models are used for phenotype classification of animal models. Created with BioRender.com.

In this work, we present minute-resolution dynamics of insulin, glucagon, and C-peptide in conscious rats under fed-state, plus an interpretable pipeline that maps dynamics to metabolic phenotype. Our main contributions include: (1) A new dataset (first multiplexed *in vivo* diabetes hormone time-series at minute scale, to our knowledge) and (2) interpretable dynamic features aligned with diabetes physiology. *We aim to share our data and results with workshop participants and take steps towards advancing diabetic metabolic phenotyping enabled by QIRT-ELISA to detect and ultimately help*

prevent T2D. Additionally, once adapted to humans, QIRT-ELISA time-series data along with our modeling framework, can inform sex-specific, personalized treatment regimens for both men and women.

2 Problem Formulation

Datasets: Let $\mathcal{P} = \{\textit{healthy}, \textit{prediabetic}, \textit{diabetic}\}$ represent our set of metabolic phenotypes corresponding to healthy, prediabetic, and diabetic rat models. We have a total $n = 9$ of experimental subjects, with $n_p = 3$ subjects per phenotype $p \in \mathcal{P}$. Subjects are assigned a phenotype label $y_k \in \mathcal{P}$. For the subjects indexing we have: $k \in \mathcal{K} = \{1, 2, 3, \cdots, 9\}$, where $k = \{1, 2, 3\}$ correspond to healthy rats, $k = \{4, 5, 6\}$ to prediabetic rats, and $k = \{7, 8, 9\}$ to diabetic rats. Each subject k, undergoes a fed-state experiment for a duration of $T_{exp} = 15$ minutes with spectral measurements of 1 minute, and averaged at regular time intervals $\Delta t = 2.5$ minutes for data analysis, yielding $m = 6$ temporal observations per subject (6-fold improvement over conventional ELISA 15-minute sampling intervals). So, we have a temporal domain of $\mathcal{T} = \{t_1, t_2, \cdots, t_6\} = \{2.5, 5, 7.5, 10, 12.5, 15\}$ minutes. QIRT-ELISA measures three biomarker at each timepoint $t_i \in \mathcal{T}$. For the biomarker vector, we will have $x_{i,k} = \left[x_{i,k}^{(ins)}, x_{i,k}^{(cpep)}, x_{i,k}^{(gluc)}\right]^T \in \mathbb{R}^3$, with a complete temporal profile of $X_k = [x_{1,k}, x_{2,k}, \cdots, x_{6,k}] \in \mathbb{R}^{3\times 6}$. In each fed-state experiment, we have two glucose measurements, $G_{start}^k, G_{end}^k \in \mathbb{R}^+$, and $\overline{G}_k = (G_{start}^k + G_{end}^k)/2$. Thus, the complete data set will be $\mathcal{D} = \{(X_k, y_k, \overline{G}_k)\}_{k=1}^9$. Similar to glucose, for each biomarker, we have a conventional ELISA cross-validation at the beginning and end of the experiment.

Classification goal: Learn mapping $h : \mathbb{R}^f \to \mathcal{P}$, with a training set of $\mathcal{D}_{train} = \{(\phi_k, y_k)\}_{k=1}^9$, with a feature vector of $\phi_k = \phi\left(X_k, \overline{G}_k\right) = \left[\phi_1^k, \phi_2^k, \cdots, \phi_f^k\right]^T$.

3 Proposed Solution

Feature Engineering: The raw time-series data of QIRT-ELISA, despite having high dimensions, do not provide useful information for physiology study. To address this challenge, we adopted a physiologically-informed feature engineering pipeline. To transform raw dynamic temporal measurements into physiologically interpretable metrics, ϕ_k, we followed the established procedures available in previous literature [12, 13, 14, 15], and adopted a comprehensive feature extraction operator (details available in appendix A.1). Given the number of extracted features (184 per subject) and sample size of $n = 9$, we face constraints in the methodological choices for data analysis due to the curse of dimensionality. To resolve this issue, we applied a multi-stage feature selection with the aim of retaining top 15 features: $\mathbb{R}^{184} \overset{variance}{\to} \mathbb{R}^{d_1} \overset{correlation}{\to} \mathbb{R}^{d_2} \overset{F-test}{\to} \mathbb{R}^{15}$. First, we filtered the features with minimal variation across the subject. For each feature $j \in \{1, 2, \cdots, 184\}$: $\sigma_j^2 = \frac{1}{n-1}\sum_{k=1}^9 \left(\phi_j^k - \overline{\phi}_j\right)^2$, where $\overline{\phi}_j = \frac{1}{9}\sum_{k=1}^9 \phi_j^k$. We set the variance threshold $\tau_{var} = 0.01$, and filtered the features whose contribution is negligible $\mathcal{S}_{var} = \left\{j : \sigma_j^2 > \tau_{var}\right\}$. Next, we deployed correlation-based redundancy removal by eliminating features that provide duplicate information. We set the correlation threshold $\tau_{corr} = 0.95$. We created the Pearson correlation matrix: $R = [r_{ij}] \in \mathbb{R}^{d_1 \times d_1}$, and identified the redundant feature pairs (i, j), $\mathcal{P}_{redundant} = \{(i, j) : |r_{ij}| > \tau_{corr}, i < j\}$, and removed the second feature $\mathcal{R} = j : \exists\, i < j \;\; such \;\; that \;\; (i, j) \in \mathcal{P}_{redundant}$. For the remaining features, we applied univariate statistical feature selection to identify a compact subset such that $|\mathcal{S}| = 15$. First, we conducted an F-test (using ANOVA) to identify features that individually have discriminative power. For each feature $j \in \{1, 2, \cdots, 15\}$, we tested the null hypothesis $H_0^{(j)} : \mu_1^{(j)} = \mu_2^{(j)} = \mu_3^{(j)}$, where $\mu_p^{(j)}$ represents the population mean of feature j per phenotype $p \in \mathcal{P}$. F-statistics: $F_j = \frac{SSB_j(k-1)}{SSW_j/(N-k)}$ where $SSB_j = \sum_{p=1}^3 n_p \left(\overline{x}_{p,j} - \overline{x}_j\right)^2$, and $SSW_j = \sum_{p=1}^3 \sum_{k\in\mathcal{G}_p} \left(\phi_j^k - \overline{x}_{p,j}\right)^2$ for each subject index $\mathcal{G}_p$ for phenotypes, P-values: $p_j = P(F_{2,6} \geq F_j)$, and effect size: $\eta_j^2 = SSB_j/(SSB_j + SSW_j)$ were calculated, and based on the F-statistic ranking we achieved the final features $j_1, j_2, \cdots, j_{15}$ such that $F_{j1} \geq F_{j2} \geq \cdots \geq F_{j15}$.

Phenotype Classification: For the final data set, $\mathcal{D}_{final} = \{(\phi_k^{(final)}, y_k)\}_{k=1}^9$, where $\phi_k^{(final)} \in \mathbb{R}^{15}$, we employed four classic models for classification: logistic regression (LR), random forest

(RF), k-nearest neighbors (KNN), support vector machines (SVM). Given the small sample size of $n = 9$, we adopted Leave-One-Out Cross-Validation (LOO-CV) framework for training $\mathcal{D}^k_{train} = \{(\phi_i^{(final)}, y_i : i \neq k)\}$ and testing $\mathcal{D}^k_{test} = \{(\phi_i^{(final)}, y_k)\}$. The predicted label for the held-out subject is: $\widehat{y}_k = \arg\max_{c\in\{0,1,2\}} P(y = c|\phi_k^{\text{final}}; \widehat{\Theta}^{(k)})$, where $\widehat{\Theta}^{(k)}$ is the model-specific parameters trained on $\mathcal{D}^k_{train}$. We calculated the overal accuracy: $ACC_{LOO} = \frac{1}{9}\sum_{k=1}^{9} \mathbb{I}[\widehat{y}_k = y_k]$ as the model performance metric.

Classifier Justification: Out of the four implemented models, LR and SVM has the best cross-validation accuracy. However, LR in prediction confidence: for classifier h, given input x_i and class $j \in \{healthy, prediabetic, diabetic\}$: $C_i = \underset{j\in0,1,2}{max}\, P(y_i = j|x_i, h)$ outperformed SVM significantly.

Classification Significance: To test the significance of the classification, we deployed permutation testing, with the null hypothesis (H_0) that the top 15 selected features are not truly discriminative. With $B = 1000$ iterations of random permutations of phenotype labels, and using the same top 15 features and LOOCV, we calculated the statistical significance: $p - value = (1 + \sum_{b=1}^{B} \mathbb{I}[ACC_{LOO,b} \geq ACC_{LOO,true}])/(B + 1)$.

4 Results and Discussion

We monitored the dynamics of insulin, glucagon, and C-peptide using QIRT-ELISA. Blood samples were directly withdrawn for 15 minutes from rats using a peristaltic pump from indwelling catheter surgically placed in the jugular veins and injected into QIRT-ELISA for real-time biomarker measurement. The rats were conscious and under normal physical activity. Please see appendix A.4 for full experimental details and appendix Fig. 4 **a-c** for time-series QIRT-ELISA data. We extracted features from the time-series data of these targets, and ranked them (Fig. 2 **a**). Cohen's d analysis quantifies the size of the difference between groups, and a larger value of Cohen's d indicates a larger difference between groups (Fig. 2 **b**). Next, we deployed different classifiers including LR, RF, KNN, SVM (100% training accuracy). First, we investigated the cross-validation accuracy of each model for the top 15 ranked metrics. LR and SVM, both showed a better cross-validation accuracy (89%, appendix Table 1). For the final model selection, we performed prediction confidence. Across all animals, LR showed a significantly better classification certainty (appendix Fig. 5 **a-b**). Next, for all the deployed models, we classified the animal models based on the top two metrics: as the best model, LR showed classifying boundaries for: the glucagon mean rise rate (GMRS, [a.u./min]) and the insulin mean rise time (IMRT, [min]). Higher GMRS values indicate a faster release of glucagon from the α-cells. The IMRT is the average time that it takes for insulin to rise from a valley point (the minimum point before the peak) to the peak point. For analysis here, we statistically transformed both metrics by robust scaling (scaled to the median of all rats). Obese rats exhibit strongly negative values, consistent with α-cell hyper-compensation in prediabetic rodents [21]. T2D rats show consistently positive values, indicating impaired glucose sensitivity and hyperglucagonemia, a hallmark of T2D [22, 23, 24]. For insulin, lean rats maintain sustained rise times; Obese/T2D rats show shortened times (In Fig. 3, negative values indicate values less than population median), reflecting β-cell exhaustion during metabolic challenges. Using only these two metrics, we established decision boundaries that accurately separated lean (healthy), obese (prediabetic), and T2D phenotypes with 89% accuracy (Fig. 3 **a-d**).

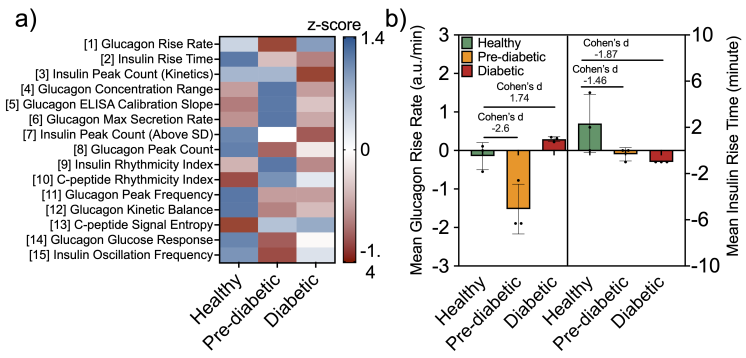

Figure 2: Feature ranking. a) Top 15 ranked features. b) Size effect of top two features among the rat models.

5 Conclusion and Future Steps

In this work, we captured minute-resolution, multiplexed time-series of insulin, glucagon, and C-peptide in conscious rats (healthy, prediabetic, and T2D) using the QIRT-ELISA platform. From these data, we derived physiologically informed dynamics (e.g., glucagon mean rise rate, insulin rise time). Logistic regression achieved the best performance on n=9 animals, and we showed that as few as two features yield clear decision boundaries.

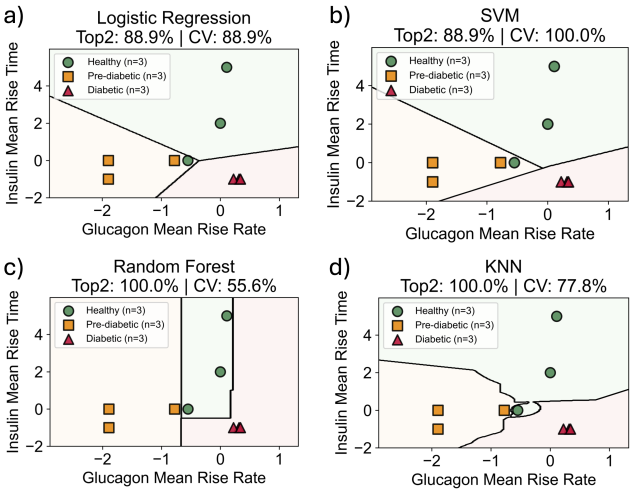

Figure 3: Decision boundaries. a-d) Performance of classifiers on the top two dynamic metrics.

This study has limitations we will address next. First, we evaluated only a small set of classical models; we will broaden this to more sophisticated models to increase the classification accuracy. Second, the animal cohort size is small (n=9, three per group); while we tried to tackle the limits of statistical power by the LOO-CV approach and maximize data utilization, we will expand the study to a larger cohort (n=60 animals per phenotype) to strengthen statistical power and generalizability of the temporal biomarkers. To enable benchmarking and community feedback, we have publicly released the code and *in vivo* data, with the goal of accelerating QIRT-ELISA–enabled diabetes research and catalyzing TS4H work on diabetes time-series.

Acknowledgments and Disclosure of Funding

This research was supported by the Canadian Institutes of Health Research, and the Natural Sciences and Engineering Research Council of Canada.

Code and Data Availability

The code and experimental in vivo data supporting this work are publicly available at: `https://github.com/hesam-abouali/qirt-elisa-fed-state-analysis/`

References

[1] Hector I. Ortega, Miriam S. Udler, Anna L. Gloyn, and Seth A. Sharp. Diabetes mellitus polygenic risk scores: Heterogeneity and clinical translation. *Nature Reviews Endocrinology*, 21(9):530–545, September 2025.

[2] Yan Zheng, Sylvia H. Ley, and Frank B. Hu. Global aetiology and epidemiology of type 2 diabetes mellitus and its complications. *Nature Reviews Endocrinology*, 14(2):88–98, February 2018.

[3] Georgia Xourafa, Melis Korbmacher, and Michael Roden. Inter-organ crosstalk during development and progression of type 2 diabetes mellitus. *Nature Reviews Endocrinology*, 20(1):27–49, January 2024.

[4] Thomas A. Lutz. Mammalian models of diabetes mellitus, with a focus on type 2 diabetes mellitus. *Nature Reviews Endocrinology*, 19(6):350–360, June 2023.

[5] Danijela Tatovic, Parth Narendran, and Colin M. Dayan. A perspective on treating type 1 diabetes mellitus before insulin is needed. *Nature Reviews Endocrinology*, March 2023.

[6] Francisca L. Henriques, Irina Buckle, and Josephine M. Forbes. Type 1 diabetes mellitus prevention: Present and future. *Nature Reviews Endocrinology*, June 2025.

[7] Sofie Hædersdal, Andreas Andersen, Filip K. Knop, and Tina Vilsbøll. Revisiting the role of glucagon in health, diabetes mellitus and other metabolic diseases. *Nature Reviews Endocrinology*, 19(6):321–335, June 2023.

[8] Jesper Gromada, Pauline Chabosseau, and Guy A. Rutter. The α-cell in diabetes mellitus. *Nature Reviews Endocrinology*, 14(12):694–704, December 2018.

[9] Jinsook Son and Domenico Accili. Reversing pancreatic β-cell dedifferentiation in the treatment of type 2 diabetes. *Experimental & Molecular Medicine*, 55(8):1652–1658, August 2023.

[10] Peng Wang, Nathalie M. Fiaschi-Taesch, Rupangi C. Vasavada, Donald K. Scott, Adolfo García-Ocaña, and Andrew F. Stewart. Diabetes mellitus—advances and challenges in human β-cell proliferation. *Nature Reviews Endocrinology*, 11(4):201–212, April 2015.

[11] Décio L. Eizirik, Lorenzo Pasquali, and Miriam Cnop. Pancreatic β-cells in type 1 and type 2 diabetes mellitus: Different pathways to failure. *Nature Reviews Endocrinology*, 16(7):349–362, July 2020.

[12] Hikaru Sugimoto, Ken-ichi Hironaka, Tomoaki Nakamura, Tomoko Yamada, Hiroshi Miura, Natsu Otowa-Suematsu, Masashi Fujii, Yushi Hirota, Kazuhiko Sakaguchi, Wataru Ogawa, and Shinya Kuroda. Improved detection of decreased glucose handling capacities via continuous glucose monitoring-derived indices. *Communications Medicine*, 5(1):103, April 2025.

[13] Ahmed A. Metwally, Dalia Perelman, Heyjun Park, Yue Wu, Alokkumar Jha, Seth Sharp, Alessandra Celli, Ekrem Ayhan, Fahim Abbasi, Anna L. Gloyn, Tracey McLaughlin, and Michael P. Snyder. Prediction of metabolic subphenotypes of type 2 diabetes via continuous glucose monitoring and machine learning. *Nature Biomedical Engineering*, 9(8):1222–1239, December 2024.

[14] Eslam Montaser, Leon S Farhy, and Boris P Kovatchev. Novel Detection and Progression Markers for Diabetes Based on Continuous Glucose Monitoring Data Dynamics. *The Journal of Clinical Endocrinology & Metabolism*, 110(1):254–262, December 2024.

[15] Boris P. Kovatchev. Metrics for glycaemic control — from HbA1c to continuous glucose monitoring. *Nature Reviews Endocrinology*, 13(7):425–436, July 2017.

[16] Hamid Aghamohammadi, Kathryn E. Thomas, Sanjana Srikant, Jason Deglint, Alexander Wong, and Mahla Poudineh. A competitive, bead-based assay combined with microfluidics for multiplexed toxin detection. *Lab on a Chip*, 23(14):3245–3257, 2023.

[17] Lucy Johnston, Gonglei Wang, Kunhui Hu, Chungen Qian, and Guozhen Liu. Advances in biosensors for continuous glucose monitoring towards wearables. *Frontiers in Bioengineering and Biotechnology*, Volume 9 - 2021, 2021.

[18] Hesam Abouali, Sanjana Srikant, Md Fahim Al Fattah, Nicole G. Barra, Darryl Chan, Dayan Ban, Jonathan D. Schertzer, and Mahla Poudineh. A Bead-Based Quantum Dot Immunoassay Integrated with Multi-Module Microfluidics Enables Real-Time Multiplexed Detection of Blood Insulin and Glucagon. *Advanced Science*, page 2412185, April 2025.

[19] Benton J. Anderson, Anne M. Curtis, Annie Jen, James A. Thomson, Dennis O. Clegg, Peng Jiang, Joshua J. Coon, Katherine A. Overmyer, and Huishi Toh. Plasma metabolomics supports non-fasted sampling for metabolic profiling across a spectrum of glucose tolerance in the Nile rat model for type 2 diabetes. *Lab Animal*, 52(11):269–277, November 2023.

[20] Torsten P., Benedikt A. Aulinger, Eric P. Smith, Deborah L. Drazen, Yve Ulrich-Lai, Randy J. Seeley, Stephen C. Woods, and David A. D'Alessio. Meal feeding improves oral glucose tolerance in male rats and causes adaptations in postprandial islet hormone secretion that are independent of plasma incretins or glycemia. *American Journal of Physiology-Endocrinology and Metabolism*, 307(9):E784–E792, November 2014.

[21] Beatriz Merino, Paloma Alonso-Magdalena, Mónica Lluesma, Patricia Ñeco, Alejandro Gonzalez, Laura Marroquí, Marta García-Arévalo, Angel Nadal, and Ivan Quesada. Pancreatic alpha-cells from female mice undergo morphofunctional changes during compensatory adaptations of the endocrine pancreas to diet-induced obesity. *Scientific Reports*, 5(1):11622, June 2015.

[22] Xi Chen, Enrique Maldonado, Ralph A DeFronzo, and Devjit Tripathy. Impaired Suppression of Glucagon in Obese Subjects Parallels Decline in Insulin Sensitivity and Beta-Cell Function. *The Journal of Clinical Endocrinology & Metabolism*, 106(5):1398–1409, April 2021.

[23] Muhmmad Omar-Hmeadi, Per-Eric Lund, Nikhil R. Gandasi, Anders Tengholm, and Sebastian Barg. Paracrine control of α-cell glucagon exocytosis is compromised in human type-2 diabetes. *Nature Communications*, 11(1):1896, April 2020.

[24] Antonia Ruiz-Pino, Arianna Goncalves-Ramírez, Margarita Jiménez-Palomares, Beatriz Merino, Manuel Castellano-Muñoz, Jean F. Vettorazzi, Alex Rafacho, Laura Marroquí, Ángel Nadal, Paloma Alonso-Magdalena, Germán Perdomo, Irene Cózar-Castellano, and Ivan Quesada. Hyperglucagonemia and glucagon hypersecretion in early type 2 diabetes result from multifaceted dysregulation of pancreatic mouse α-cells. *Pflügers Archiv - European Journal of Physiology*, 477(2):207–221, February 2025.

A Technical Appendices and Supplementary Material

A.1 Feature Extraction

We studied the profile of each biomarker according to the previous studies that have analyzed the time-series data from CGM devices [12, 13, 14, 15]. For instance, area under the curves (AUCs) of each hormone/peptide, including postive, negative, and total AUCs, coefficient variations %, glucose sensitivity, etc. are calculated similar to the measurements from CGM devices. Since we are measuring new hormones/peptide that CGM devices do not provide any information about them, new metrics are introduced to fully capture the behavior of biomarker profiles for different phenotypes. Moreover, these features help us understand physiological mechanism happening *in vivo* in the animal models such that they will not be visible to us with comparison of raw data measurements. Some of the important definitions are listed below.

Definitions: Let $x_{j,k}^{(j)}$ represent the normalized signal for animal k, biomarker $j \in \{insulin, glucagon, C-peptide\}$ at timepoint $i \in \{1,2,3,4,5,6\}$.

Normalization:

$F_0^{j,k} = \frac{1}{6}\sum_{i=1}^{6} F_i^{j,k}$ and $x_{i,k}^{(j)} = \frac{F_i^{j,k}}{F_0^{j,k}}$

Category 1: Basic Device Measurements (15 metrics, 5 per biomarker)

1. Mean raw signal ([a.u.]) = $\mu_{raw}^{j,k} = \frac{1}{6}\sum_{i=1}^{6} F_i^{j,k}$
2. Range Raw Signal (a.u.) = $R_{raw}^{j,k} = \max_i(F_i^{j,k}) - \min_i(F_i^{j,k})$
3. Coefficient of variation (%) = $CV^{j,k} = \frac{\sigma\left(x_{:,k}^{(j)}\right)}{\mu\left(x_{:,k}^{(j)}\right)} \times 100\%$

   where $\sigma\left(x_{:,k}^{(j)}\right) = \sqrt{\frac{1}{5}\sum_{i=1}^{6}\left(x_{i,k}^{(j)} - \mu\left(x_{:,k}^{(j)}\right)\right)^2}$
4. Maximum normalized value = $x_{max}^{j,k} = \underset{i}{max}\left(x_{i,k}^{(j)}\right)$
5. Minimum normalized value = $x_{max}^{j,k} = \underset{i}{min}\left(x_{i,k}^{(j)}\right)$

Category 2: Glucose-Adjusted Parameters (21 metrics)

Glucose Normalization Parameters

- Average Glucose Concentration: $\bar{G}_k = \frac{G_{start,k}+G_{end,k}}{2}$
- Glucose Change: $\Delta G_k = G_{end,k} - G_{start,k}$

Efficiency Metrics (7 per biomarker)

1. Glucose-Corrected Baseline (normalized to 5.6 mM): $\text{corrected}_{5.6}^{j,k} = \mu(\mathbf{x}_{:,k}^{(j)}) \times \frac{5.6}{\bar{G}_k}$
2. Glucose Efficiency Ratio: $\eta_{glucose}^{j,k} = \frac{\mu(\mathbf{x}_{:,k}^{(j)})}{\bar{G}_k}$
3. Glucose Sensitivity: $\text{sens}_{glucose}^{j,k} = \frac{\text{range}(\mathbf{x}_{:,k}^{(j)})}{\bar{G}_k}$

4. Glucose Responsiveness: $\text{resp}_{glucose}^{j,k} = \frac{|\Delta G_k|}{\sigma(\mathbf{x}_{:,k}^{(j)})} \times \text{sign}(\Delta G_k \times \text{corr}(\mathbf{x}_{:,k}^{(j)}, \mathbf{G}_k))$
5. Glucose Correlation: $\rho_{glucose}^{j,k} = \text{corr}(\mathbf{x}_{:,k}^{(j)}, [\bar{G}_k, \bar{G}_k, ..., \bar{G}_k])$
6. Glucose Peak Efficiency: $\eta_{peak}^{j,k} = \frac{\max_i(x_{i,k}^{(j)}) - 1.0}{\bar{G}_k}$
7. Glucose-Adjusted Range: $R_{glucose}^{j,k} = \frac{x_{max}^{j,k} - x_{min}^{j,k}}{\bar{G}_k}$

Category 3: Peak Dynamics Analysis (21 metrics)

Peak Detection Algorithm:

- Threshold Definition: $\tau_{peak}^{j,k} = \mu(\mathbf{x}_{:,k}^{(j)}) + 0.5\sigma(\mathbf{x}_{:,k}^{(j)})$
- Peak set: $P^{j,k} = \{i : x_{i,k}^{(j)} > \tau_{peak}^{j,k} \text{ and } x_{i,k}^{(j)} > x_{i-1,k}^{(j)} \text{ and } x_{i,k}^{(j)} > x_{i+1,k}^{(j)}\}$

Peak Metrics (7 per biomarker)

1. Peak Count Above Standard Deviation: $N_{peaks}^{j,k} = |P^{j,k}|$
2. Peak Frequency (peaks per minute): $f_{peaks}^{j,k} = \frac{N_{peaks}^{j,k}}{T_{exp}/60}$
3. Peak Dynamic Range Ratio: $R_{dynamic}^{j,k} = \frac{\max_{i \in P^{j,k}}(x_{i,k}^{(j)})}{\min_{i \in T^{j,k}}(x_{i,k}^{(j)})}$
4. Peak Mean Width (temporal extent): $W_{peak}^{j,k} = \frac{1}{|P^{j,k}|} \sum_{i \in P^{j,k}} \Delta t_{width,i}$ where $\Delta t_{width,i}$ is the full-width-half-maximum around peak i
5. Peak Total Excursion: $E_{total}^{j,k} = \sum_{i \in P^{j,k}} (x_{i,k}^{(j)} - \mu(\mathbf{x}_{:,k}^{(j)}))$
6. Peak Coefficient of Variation: $CV_{peak}^{j,k} = \frac{\sigma(\{x_{i,k}^{(j)} : i \in P^{j,k}\})}{\mu(\{x_{i,k}^{(j)} : i \in P^{j,k}\})}$
7. Peak Maximum Prominence: $P_{prom}^{j,k} = \max_{i \in P^{j,k}}(x_{i,k}^{(j)} - \max(\text{neighboring troughs}))$

Category 4: Rate Analysis (21 metrics)

Discrete Rate Calculations

- Point-to-Point Rate: $r_i^{j,k} = \frac{x_{i+1,k}^{(j)} - x_{i,k}^{(j)}}{\Delta t}$
- Smoothed Rate (3-point derivative): $r_{smooth,i}^{j,k} = \frac{x_{i+1,k}^{(j)} - x_{i-1,k}^{(j)}}{2\Delta t}$
- Acceleration (second derivative): $a_i^{j,k} = \frac{r_{i+1}^{j,k} - r_i^{j,k}}{\Delta t}$

Rate Metrics (7 per biomarker)

1. Maximum Increase Rate: $R_{max,inc}^{j,k} = \max_i(r_i^{j,k})$
2. Maximum Decrease Rate: $R_{max,dec}^{j,k} = \max_i(|r_i^{j,k}|)$ where $r_i^{j,k} < 0$
3. Mean Absolute Rate: $R_{mean,abs}^{j,k} = \frac{1}{5} \sum_{i=1}^{5} |r_i^{j,k}|$
4. Rate Variability: $\sigma_{rate}^{j,k} = \sqrt{\frac{1}{4} \sum_{i=1}^{5} (r_i^{j,k} - \bar{r}^{j,k})^2}$
5. Maximum Acceleration: $A_{max}^{j,k} = \max_i(a_i^{j,k})$
6. Trend Slope (linear regression): $\beta_{trend}^{j,k} = \frac{\sum_{i=1}^{6}(t_i - \bar{t})(x_{i,k}^{(j)} - \bar{x}^{j,k})}{\sum_{i=1}^{6}(t_i - \bar{t})^2}$
7. Trend R-squared: $R_{trend}^{2,j,k} = \frac{[\sum_{i=1}^{6}(t_i - \bar{t})(x_{i,k}^{(j)} - \bar{x}^{j,k})]^2}{\sum_{i=1}^{6}(t_i - \bar{t})^2 \sum_{i=1}^{6}(x_{i,k}^{(j)} - \bar{x}^{j,k})^2}$

Category 5: Area Under Curve (AUC) Metrics (15 metrics)

AUC Calculations

- Total AUC (Trapezoidal Integration): $\text{AUC}_{total}^{j,k} = \frac{\Delta t}{2} \sum_{i=1}^{5} (x_{i,k}^{(j)} + x_{i+1,k}^{(j)})$
- AUC Above Baseline (positive excursions):
  $\text{AUC}_{above}^{j,k} = \frac{\Delta t}{2} \sum_{i=1}^{5} [\max(x_{i,k}^{(j)} - 1.0, 0) + \max(x_{i+1,k}^{(j)} - 1.0, 0)]$
- AUC Below Baseline (negative excursions):
  $\text{AUC}_{below}^{j,k} = \frac{\Delta t}{2} \sum_{i=1}^{5} [\max(1.0 - x_{i,k}^{(j)}, 0) + \max(1.0 - x_{i+1,k}^{(j)}, 0)]$

AUC Metrics (5 per biomarker)

1. Glucose-Corrected AUC: $\text{AUC}_{glucose}^{j,k} = \text{AUC}_{total}^{j,k} \times \frac{5.6}{\bar{G}_k}$
2. AUC Glucose Efficiency: $\eta_{AUC}^{j,k} = \frac{\text{AUC}_{total}^{j,k}}{\bar{G}_k}$
3. Time-Weighted Average: $\text{TWA}^{j,k} = \frac{\text{AUC}_{total}^{j,k}}{T_{exp}}$
4. AUC Balance Ratio: $B_{AUC}^{j,k} = \frac{\text{AUC}_{above}^{j,k}}{\text{AUC}_{below}^{j,k} + \epsilon}$
   where $\epsilon = 10^{-6}$ prevents division by zero
5. Net AUC: $\text{AUC}_{net}^{j,k} = \text{AUC}_{above}^{j,k} - \text{AUC}_{below}^{j,k}$

Category 6: Oscillation Pattern Analysis (18 metrics)

Oscillatory Characteristics (2 per biomarker)

1. Zero Crossings (baseline crossings): $Z^{j,k} = |\{i : \text{sign}(x_{i,k}^{(j)} - 1.0) \neq \text{sign}(x_{i-1,k}^{(j)} - 1.0)\}|$
2. Oscillation Frequency: $f_{osc}^{j,k} = \frac{Z^{j,k}}{T_{exp}/60}$

Pattern Regularity Metrics (2 per biomarker)

1. Autocorrelation-Based Regularity: $\rho_{auto}^{j,k}(\tau) = \frac{\sum_{i=1}^{6-\tau} (x_{i,k}^{(j)} - \bar{x}^{j,k})(x_{i+\tau,k}^{(j)} - \bar{x}^{j,k})}{\sum_{i=1}^{6} (x_{i,k}^{(j)} - \bar{x}^{j,k})^2}$
2. Pattern Regularity Index: $I_{reg}^{j,k} = \max_{\tau \in \{1,2,3\}} |\rho_{auto}^{j,k}(\tau)|$

Oscillation Metrics (3 per biomarker)

1. Rhythmicity Index (Fast Fourier Transform-based): $I_{rhythm}^{j,k} = \frac{\max(|\text{FFT}(\mathbf{x}_{:,k}^{(j)})|)}{\text{mean}(|\text{FFT}(\mathbf{x}_{:,k}^{(j)})|)}$
2. Signal Entropy: $H^{j,k} = -\sum_{i=1}^{6} p_i \log_2(p_i)$
   where $p_i = \frac{|x_{i,k}^{(j)}|}{\sum_{j=1}^{6} |x_{j,k}^{(j)}|}$
3. Temporal Stability: $S_{temp}^{j,k} = 1 - \frac{\sigma(\mathbf{x}_{:,k}^{(j)})}{\max(\mathbf{x}_{:,k}^{(j)}) - \min(\mathbf{x}_{:,k}^{(j)}) + \epsilon}$

Category 7: Inter-Biomarker Correlations

For each pair of biomarkers $(j_1, j_2) \in \{(I, Glu), (I, CP), (Glu, CP)\}$ and animal k we have:

Basic Correlations:

1. Pearson Correlation: $\rho_{j_1,j_2}^{k} = \frac{\text{cov}(\mathbf{x}_{:,k}^{(j_1)}, \mathbf{x}_{:,k}^{(j_2)})}{\sigma(\mathbf{x}_{:,k}^{(j_1)}) \sigma(\mathbf{x}_{:,k}^{(j_2)})}$
2. Cross-Correlation Maximum: $\rho_{max}^{j_1,j_2,k} = \max_{\tau} |\rho_{j_1,j_2}^{k}(\tau)|$
   where $\rho_{j_1,j_2}^{k}(\tau)$ is the cross-correlation at lag τ

Category 8: ELISA Quantitative Metrics

ELISA Calibration Parameters:

- Calibration Slope: $\beta_{cal}^{j,k} = \frac{\text{known concentration}}{\text{measured signal}}$
- Calibration Error: $\epsilon_{cal}^{j,k} = \frac{|\text{predicted} - \text{actual}|}{|\text{actual}|} \times 100\%$

Concentration Metrics (7 per biomarker)

1. Mean Concentration (pM): $C_{mean}^{j,k} = \frac{C_{start}^{j,k} + C_{end}^{j,k}}{2}$
2. Concentration Range: $C_{range}^{j,k} = C_{end}^{j,k} - C_{start}^{j,k}$
3. CV Concentration: $CV_{conc}^{j,k} = \frac{|C_{range}^{j,k}|}{C_{mean}^{j,k}} \times 100\%$
4. Maximum Concentration: $C_{max}^{j,k} = \max(C_{start}^{j,k}, C_{end}^{j,k})$
5. Maximum Secretion Rate: $SR_{max}^{j,k} = \frac{|C_{range}^{j,k}|}{T_{exp}}$
6. Concentration-Device Correlation: $\rho_{conc-dev}^{j,k} = \text{corr}([C_{start}^{j,k}, C_{end}^{j,k}], [x_{1,k}^{(j)}, x_{6,k}^{(j)}])$
7. Calibration Quality Score: $Q_{cal}^{j,k} = \frac{1}{1+\epsilon_{cal}^{j,k}/100}$

Category 9: Peak Kinetics Analysis (21 metrics)

Kinetic Parameter Definitions

- Rise Time (to peak): $T_{rise,i}^{j,k} = t_i - t_{start}$
  where t_{start} is the timepoint before peak i begins to rise.
- Decline Time (from peak): $T_{decline,i}^{j,k} = t_{end} - t_i$
  where t_{end} is the timepoint where peak i returns to baseline.

Kinetic Metrics (7 per biomarker)

1. Mean Rise Time: $\bar{T}_{rise}^{j,k} = \frac{1}{|P^{j,k}|} \sum_{i \in P^{j,k}} T_{rise,i}^{j,k}$
2. Mean Decline Time: $\bar{T}_{decline}^{j,k} = \frac{1}{|P^{j,k}|} \sum_{i \in P^{j,k}} T_{decline,i}^{j,k}$
3. Mean Rise Rate: $\bar{R}_{rise}^{j,k} = \frac{1}{|P^{j,k}|} \sum_{i \in P^{j,k}} \frac{x_{i,k}^{(j)} - x_{baseline,i}^{j,k}}{T_{rise,i}^{j,k}}$
4. Mean Decline Rate: $\bar{R}_{decline}^{j,k} = \frac{1}{|P^{j,k}|} \sum_{i \in P^{j,k}} \frac{x_{i,k}^{(j)} - x_{baseline,i}^{j,k}}{T_{decline,i}^{j,k}}$
5. Rise-Decline Ratio: $\gamma_{kinetic}^{j,k} = \frac{\bar{R}_{rise}^{j,k}}{|\bar{R}_{decline}^{j,k}| + \epsilon}$
6. Peak Count: $N_{peaks}^{j,k} = |P^{j,k}|$
7. Peak Frequency: $f_{peak}^{j,k} = \frac{N_{peaks}^{j,k}}{T_{exp}/60}$

A.2 Fed-State Data Measurements by QIRT-ELISA

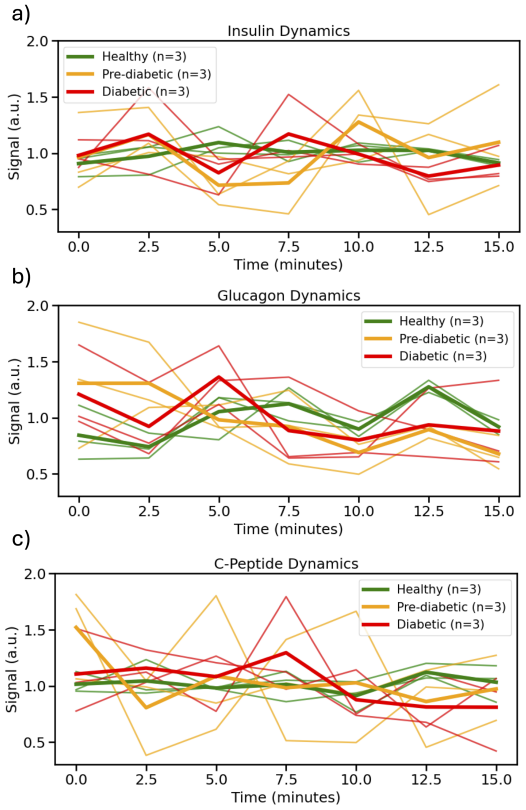

Figure 4: Monitored hormonal/peptide data. The measurements during $T_{exp} = 15\ min$ for a) insulin, b) glucagon, and c) C-peptide. Here, we assume that the data from the fed-state experiments represent a quasi-steady metabolic condition where biomarker fluctuations reflect intrinsic regulatory dynamics since we do not have external perturbations such as glucose gavage or injection.

A.3 Comparison of Classification Models

Model 1. Multinomial logistic regression model (with softmax parameterization), $P\left(y_k = c|\phi_k^{(final)}\right) = \frac{exp(w_c^T \phi_k^{(final)} + b_c)}{\sum_{j=0}^{2} exp(w_j^T \phi_k^{(final)} + b_j)}$ for $c \in \{0, 1, 2\}$, where $W = [w_0, w_1, w_2] \in \mathbb{R}^{15\times 3}$ and bias vector of $b = [b_0, b_1, b_2]^T \in \mathbb{R}^3$. Considering a log-likelihood for subject k with the true label y_k, and L2 regularization to avoid overfitting, the complete objective is $\mathcal{J}(W, b) = -\sum_{k=1}^{9}\sum_{c=0}^{2} \mathbb{I}[y_k = c]\, logP\left(y_k = c|\phi_k^{(final)}\right) + \lambda \sum_{c=1}^{2}\sum_{j=1}^{15} w_{cj}^2$.

Model 2. Support Vector Machine Here the model aims to solve the optimization problem of $\underset{w,b,\xi}{min} \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n} \xi_i$ subject to $y_i\left(w^T\phi(x_i) + b\right) \geq 1 - \xi_i, \xi_i \geq 0$, where for inputs features are mapped by $\phi(x_i)$: $K(x_i, x_j) = \phi(x_i)^T \phi(x_j)$ with a hyperparameter optimization through grid seach over $C \in \{0.1, 1.10\}$ and $\gamma \in \{scale, auto\}$.

Model 3. Random Forest We used this model to construct an ensemble of B decision trees $\hat{f}_{RF}(x) = \frac{1}{B}\sum_{b=1}^{B} T_b(x)$, where each T_b is trained on a bootstrap sample with randomly subsampling features at splits. Since our sample number was limited ($n = 9$), we set the following parameters for shallow depth and minimum leaf samples to prevent overfitting: estimator numbers of $B \in \{20, 50\}$, maximum depth of $d_{max} \in \{3, 5\}$, minimum sample per leaf of $n_{leaf} \in \{2, 3\}$, and with feature subsampling of $m = \sqrt{15}$.

Model 4. K-Nearest Neighbors This classifier was utilized to classify the subjects based on the majority vote: $\hat{y} = arg\ \underset{c}{max} \sum_{x_i \in N_k(x)} I(y_i = c)$ where $N_k(x)$ denotes the k nearest neighbors to our query point x. Given our subject size ($n = 9$), chose a neighborhood size of $k \in \{3, 5\}$, distance weighting of $w_i = \frac{1}{d(x,x_i)^2}$ where $d(x_i, x_j) = \sqrt{\sum_{l=1}^{p}(x_{il} - x_{jl})^2}$.

We employed different ML classification models, initially for the experimental data, including logistic regression, random forest (RF), k-nearest neighbors (KNN), support vector machines (SVM). We followed a similar feature selection from the experimental data, and had priority-matched features for all classifiers. In the RF model, although we have a successful classification for training, but a low level of cross-validation accuracy suggests a possible overfitting to training data (Table 1). Moreover, when analyzing the decision boundaries in this model, the jagged and rectangular boundaries represent artifacts of the tree-based splitting process, that undermines a realistic precision in physiological interpretation and classification (Fig. 2 e). In KNN, we have moderate cross-validation accuracy, inferior to both LR and SVM results (Fig. 2 f). Moreover, in the KNN classification, the complex curved boundaries with abrupt transitions, create physiologically implausible "islands" of classification. This will imply that a metabolic classification of a subject depends more on which particular subject (animals) it resembles rather than underlying the actual biological mechanisms, and nearest neighbor logic will not capture causal physiological relationships among features (Fig. 2 f).

Comparing LR and SVM, that outperformed previous models, both show good accuracies in training and cross-validation (Table 1), and the similar linear boundaries in both suggest a continuously progression in α and β-cells efficiency and dysfunction, and they follow along a biological continua (Fig. 3 c-d).

With similar performance from LR and SVM, we investigated the most reliable model. For this aim, we tested the prediction confidence for a trained classifier h, input x_i, and class $j \in \{healthy, prediabetic, diabetic\}$: $C_i = \underset{j\in 0,1,2}{max} P(y_i = j|x_i, h)$, and also classification uncertainty: $U_i = 1 - C_i$. From the uncertainty heatmap, we can see the higher uncertainty for SVM than LR for most animals (Fig. 4 a-b).

Table 1: Model Performance Comparison

Classifier	Cross-Validation Accuracy
Logistic Regression	0.89
Random Forest	0.56
SVM	0.89
KNN	0.78

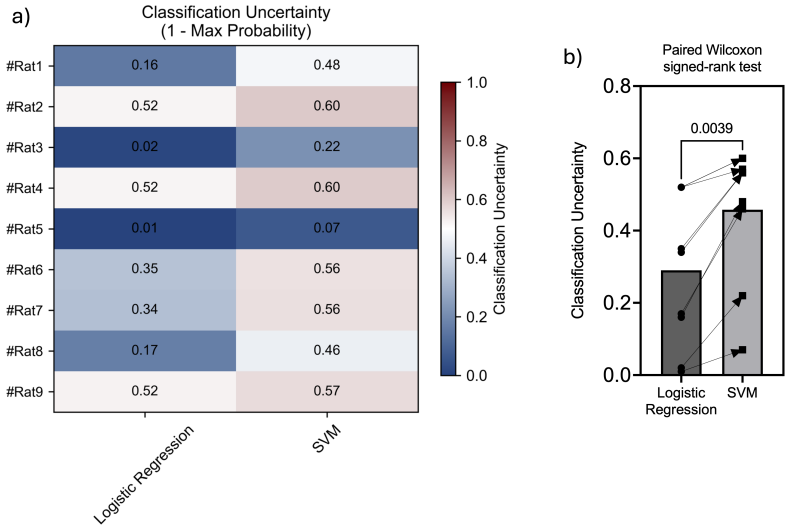

Figure 5: Robustness comparison. The classification uncertainty of LR and SVM are compared, and the significance level for these values are shown.