Heart Rate and CGM Feature Representation Diabetes Detection From Heart Rate: Learning Joint Features of Heart Rate and Continuous Glucose Monitors Yields Better Representations

HOOTAN RASHTIAN[ID], **SOLMAZ SHARIAT TORBAGHAN**, **SALAR RAHILI**[ID], **(Member, IEEE)**, **MICHAEL SNYDER, AND NIMA AGHAEEPOUR**
January AI, Menlo Park, CA 94025, USA

Corresponding author: Hootan Rashtian (hootan@january.ai)

ABSTRACT Continuous Glucose Monitors (CGMs) provide tremendous value for diabetes detection and management. However, their high cost and regulatory complications have prevented the widespread use of CGMs. On the other hand, Heart Rate (HR) monitors are in wide use and growing in popularity. In this work, we investigate the connection between HR monitor and CGM devices to find a cheaper alternative for measuring glucose dysregulation. We recruited 550 volunteers that included healthy, type 2 diabetic, pre-diabetic, and gestational diabetic cohorts to wear CGM and HR monitors for 10 days. Although the physiological mechanisms underlying glucose regulation and heart rate share many commonalities, we find that commonly used features in time series analysis yield poor correlations between CGM and HR signals. However, by learning a joint representation between CGM and HR using Canonical Correlation Analysis (CCA), we can learn CGM and HR features in CCA space respectively that have a statistically significant correlation. Finding HR features that maximize the CCA objective with CGM enables us to learn about a subject's glucose regulatory system using an HR monitor alone and not require cumbersome CGMs. Here we consider the detection of diabetes with heart rate monitors as a particular application. We find that CCA representations of heart rate can serve as a proxy for using CGMs in diabetes classification.

INDEX TERMS CCA (canonical correlation analysis), diabetes, continuous glucose monitors, heart rate monitors.

I. INTRODUCTION

Diabetes is an increasing problem worldwide globally affecting 422 million adults and costing over $82b billion [8]. In the United States alone nearly 10% of the population are diagnosed with diabetes and another 84 million are pre-diabetic. Without intervention up to 70% of pre-diabetics may progress to type 2 diabetes [15]. The underlying cause of prediabetes and type 2 diabetes is the dysregulation of glucose levels in the blood. Continuous Glucose Monitors (CGMs) provide a continuous window into the

The associate editor coordinating the review of this manuscript and approving it for publication was Shen Yin.

physiological mechanism that regulates glucose in the blood allowing patients and clinicians to make informed health decisions.

While CGMs have shown great promise for early detection of glucose dysregulation and cardiovascular diseases [4], their use remains limited due to cost and limited healthcare access. Patients with limited or no insurance will have to pay out of pocket and CGMs have to be replaced roughly every 10 days which adds to the cost. Unlike CGMs, Heart Rate (HR) monitors do not require prescriptions, do not require regular replacement, and are significantly cheaper in cost. The underlying physiological mechanisms driving heart rate and glucose in blood suggest that both glucose in the

blood and heart rate values can be correlated under certain circumstances [7], [18].

To understand the relationship between heart rate and glucose regulation, we recruited 550 volunteers to wear CGM and HR devices for 10 days. The volunteers included healthy, pre-diabetic, type 2 diabetic, and gestational diabetic subjects. Previous studies have studied CGM and HR for type 1 diabetics [3], [13]. The large dataset that we have collected provides a unique opportunity to understand how wearable devices capture underlying physiological mechanisms to a broad set of cohorts. To understand the relationship between HR and CGM, in this paper we propose a machine learning framework to *learn features* from HR that correlate with *learned* CGM feature representations. Learning joint features between HR and CGM can be used in many downstream applications of measuring glucose regulation. As an example downstream task, we demonstrate here using the learned features to train a diabetes classifier from HR alone.

To find joint HR and CGM feature representations, we use Canonical Correlation Analysis (CCA) based on a dataset collected by volunteers wearing CGM and HR devices. With this dataset, CCA learns joint CGM and HR latent representations that are maximally correlated. We find that these latent feature representations can be used to identify different cohorts with glucose dysregulation. As a result of these groupings, we further show that using CCA-based feature representations can be applied in the downstream task of diabetes classification. We train a linear classifier on the HR CCA latent features and show accuracy degrades slightly compared to CGM representations that use Principal Component Analysis (PCA). Our primary contributions are:

- Collected CGM and HR data from 550 healthy, pre-diabetic, type 2 diabetic, and gestational diabetic volunteers
- A CCA framework for learning a joint latent representation of HR and CGM data
- Comparing how learned features are distributed amongst different cohorts
- Showing that HR features learned with our framework can be used for diabetes classification

In Section II we describe the CGM and HR dataset we collected from 550 volunteers to validate our methodology. In Section III we describe our CCA framework for extracting heart rate features and training a diabetes classifier. In Section IV we show the quantitative results of our methodology on the collected dataset. We also compare the diabetes classifier we trained using our proposed latent features from CCA against an HR alone baseline. In Section V we conclude that our framework of collecting CGM and HR data can be used to train ML algorithms that can learn to predict glucose dysregulation from HR alone.

II. COLLECTING CGM AND HR DATASET

We recruited 550 volunteers to participate in a "sugar challenge": an educational opportunity for volunteers to learn about how their food consumption and activities affect their

TABLE 1. Number of subjects and disease classifications. We later group the data of users in "GEST", "PRE", and "TYPE2" into "unhealthy" group for the analysis of the downstream task explained in Section III-C.

	HEALTHY	TYPE 2	PRE	GEST.
N	312	217	14	7
AGE	34 ± 8	43 ± 9	35 ± 8	39 ± 8
BMI	29 ± 7	39 ± 8	37 ± 6	32 ± 7
GENDER (F/M/O)	189/122/1	154/63/0	8/6/0	7/0/0

blood glucose levels. The participants used the health tracking mobile application developed by January AI for this specific study to monitor and store their heart rate, glucose, and food logs to gain summaries and insights. The participants in the study were shipped with a starter package that included: a Continuous Glucose Monitor (CGM), a Heart Rate Monitor, a glucose shot, and two bars as control meals. Once the package arrived, participants scheduled a remote meeting with a member of the clinical team to help them wear the CGM, heart rate device, and activate the app to integrate the two devices. After onboarding, over the course of 10 days participants logged their meals and activities in the app. If and when participants successfully completed the challenge, they were rewarded with a $50.00 Amazon.com gift card.

Table 1 shows the number of cohorts in each cohort disease classification with their respective age (mean and standard deviation), Body Mass Index (BMI in kg/m^2 and the mean and standard deviation), and gender distributions. All demographics information was self-reported by participants. The disease classifications are: healthy (no form of diabetes), type 2 diabetes, pre-diabetes, and gestational diabetes.

III. METHODS

In this section, we describe our methodology for extracting features from heart rate (HR) and continuous glucose monitors (CGM). We then describe how we learn joint HR and CGM latent representations using Canonical Correlation Analysis (CCA) and apply the aggregated features for the downstream task of diabetes classification.

A. FEATURE EXTRACTION

For each user u we represent their CGM values as the time series $x_u^{cgm}(t)$. Similarly, we represent the user's HR with the time series $x_u^{hr}(t)$. Over the course of a time window of T minutes, we extract summary statistics commonly used in signal processing and time series analysis to represent the windowed time series as a feature vector. That is, for CGM values from start time t_0 to end time $t_0 + T$, we extract D features to represent the windowed time series as the feature vector $v_{u,t_0}^{cgm} \in \mathbf{R}^D$. Similarly, for heart rate during the same window we extract features $v_{u,t_0}^{hr} \in \mathbf{R}^D$. We extract features using the `tsfresh` Python package [2]. The types of features that are extracted include summary statistics (such as mean, minimum, maximum), Fourier and Wavelet coefficients, amongst many others. For the full list of features see Appendix A in [2].

There are occasions where we have missing values in the time window (for example, when the user has taken off the HR device for recharging). For the cases with partially missing values in the window, we impute the missing values using mean imputation. However, if the entire window is missing (for cases where the window length is small or the device was not worn for long duration), then we discard the entire window.

The windowed features in our analyses have no overlap and we use the index t_0 as a reference to the window. For example, if the length of the window is 60 minutes, and the volunteer wore the devices for 10 days, we have a tuple of 240 CGM and HR feature vectors. However, some users did not wear the devices for 10 full days and some wore the devices for even longer than 10 days. Regardless of the number of days a user wore the CGM and HR devices, we use data from all users that have at least one full window of CGM and HR values.

B. JOINT FEATURE REPRESENTATION

For each feature vector tuple $(v_{u,t_0}^{cgm}, v_{u,t_0}^{hr})$, we train embedding functions $f_{hr}(\cdot)$ and $f_{cgm}(\cdot)$ that learn a joint low dimensional representation $(z_{u,t_0}^{cgm}, z_{u,t_0}^{hr})$. Our objective is to find low dimensional CGM and HR representations that are maximally correlated in the latent space. A well-known method for finding maximally correlated latent representations is Canonical Correlation Analysis (CCA) [5], [10].

CCA is a statistical method employed to investigate relationships among two or more variable sets, each consisting of at least two variables. CCA is the multivariate form of the general linear model, which presumes that all analyses are correlational, derive estimates by applying weights to measured variables, and yield variance-accounted-for effect sizes [16].

CCA is similar to Principal Component Analysis (PCA) in that a high dimensional vector is linearly projected to a low dimensional vector space. Whereas in PCA the low dimensional mapping aims to maximize the variance of the high dimensional dataset, in CCA the low dimensional mapping maximizes the correlation between the CGM and HR latent representations.

C. FEATURE AGGREGATION AND DIABETES CLASSIFICATION

Over the duration of the study, each user will have a set of latent feature vectors $z_{u,i}^{cgm}$ and $z_{u,i}^{hr}$ for $i \in 1, 2, \ldots, N_u$ where N_u is the total number of windows available for user u. To have a single latent representation for a user, we take the average of all available latent vectors for the CGM and HR sets respectively:

$$\bar{z}_u^{cgm} = \frac{1}{N_u} \sum_{i=1}^{N_u} z_{u,i}^{cgm} \qquad (1a)$$

$$\bar{z}_u^{hr} = \frac{1}{N_u} \sum_{i=1}^{N_u} z_{u,i}^{hr} \qquad (1b)$$

This aggregation operation on latent features is typically referred to as "mean pooling" in representation learning and other aggregation operations can be considered in future work such as in [6], [17].

The tuple $(\bar{z}_u^{cgm}, \bar{z}_u^{hr})$ can be thought of as a joint vector representation capturing the physiological state of a user u. As is often done in representation learning, we can evaluate the utility of the aggregated latent representations in a downstream supervised task [1]. In this paper, the downstream task we consider is diabetes classification (which has been receiving attention in research such as [11], [12], [14]) from heart rate data alone.

We train a binary linear logistic classifier on the aggregated latent heart rate features $\bar{z}_u^{hr}$ to classify if a user self-reports as "healthy" or self-reports as having glucose dysregulation (that is, self-report having either type-2 diabetes, gestational diabetes, or prediabetes). Since we did not require volunteers to provide us with their medical records or get a Hemoglobin A1C test, we can only rely on self-reporting. To avoid any confusion, we emphasize that only the demographic information and disease class were self-reported. The CGM and HR data were collected objectively through CGM and HR wearable devices.

To understand how our features derived from HR compare to CGMs, we train a classifier that uses latent features from PCA of CGM. We compute the PCA transform for each heart rate feature vector $v_{u,i}^{cgm}$ and aggregate for each user as done in Equation (1a). Recall that latent features from PCA result in a subspace that maximizes the variance of the CGM dataset. We consider the CGM-derived features to be "state-of-the-art" since CGM signals are directly measuring glucose and diabetes is a condition related to glucose dysregulation. Since CCA transforms heart rate to a latent representation that is maximally correlated with latent CGM features $z_{u,i}^{cgm}$, we hypothesize that the aggregated CCA features in Equation (1b) to be also predictive of glucose symptoms. We stress that during inference using the CCA HR features no CGM data is used.

1) LIMITATIONS

Since we rely on self-reporting of "disease class", many of the users that claim to be healthy may have glucose dysregulation but have yet to be diagnosed (either because symptoms are not showing or do not have access to adequate healthcare). Like other studies that rely on volunteers, we are limited to the biased sample of users that have agreed to participate and may not be representative of the general population. Finally, many factors contribute to diabetes that we do not have sufficient samples to control for. Future studies and data collection efforts should consider these limitations.

IV. RESULTS

As discussed in Section III-B, for a given window of length T, we compute the CCA latent representations of HR and CGM ($z_{u,i}^{hr}$ and $z_{u,i}^{cgm}$ respectively) for each user u and each window i.

FIGURE 1. This figure shows how the first CCA component of Heart Rate and CGM for a user varies compared to the rest of the population. Each dot corresponds to features extracted over the course of a 3-hour window. Light blue dots are samples from all users in our dataset and represent the distribution of the CCA components (and do not change between each subplot). Each subplot shows the overlay of the distribution of CCA components of a randomly selected user compared to the rest of the population. Since each user occupies a specific region of the latent CCA space, the CCA representation of CGM and HR can be used to characterize and represent the user. Horizontal axis: HR first CCA component; Vertical axis: CGM first CCA component.

Table 2 shows how the Pearson correlation between the first component of $z_{u,i}^{hr}$ and $z_{u,i}^{cgm}$ varies as we change the window length T. The Pearson correlation is averaged over 10 fold cross-validation across users. A window length of 3 hours yields the best subspace that has the most correlation.

Figure 1 shows the 3-hour time window first CCA components for $z_{u,i}^{hr}$ and $z_{u,i}^{cgm}$ for all users in light blue and for a sample user with the orange marker. We observe that each user occupies a specific sub-region of the latent CGM and HR space. In Figure 2 we show the aggregation of the two CCA components $\bar{z}_u^{hr}$ with a time window of 3 hours as done in

Equations (1a) and (1b). Since glucose dysregulation such as type 2 diabetes and prediabetes are defined based on glucose in the blood, it should not be surprising that features derived from CGM are predictive of glucose dysregulation. For this reason, we investigate how predictive are HR features $\bar{z}_u^{hr}$ alone without any CGM data for glucose dysregulation.

A. DIABETES CLASSIFICATION
Section III-C describes our procedure for training a linear logistic classifier using the heart rate representation in Equation (1b) to classify users as healthy ($N = 312$) or
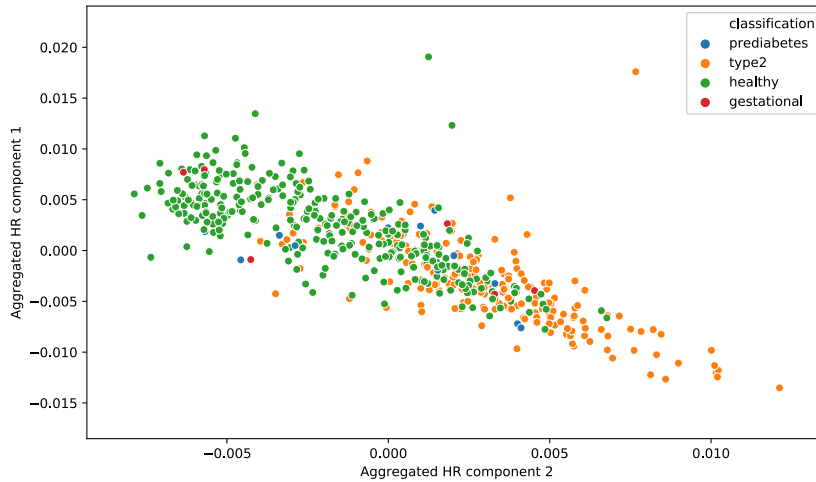
FIGURE 2. Aggregation of HR features as done in Equation (1b) groups users based on disease cohort: each dot is a representation of the users using the HR CCA representation. We can see that the HR CCA representation of users can discriminate the cohort of users between healthy and not healthy groups.

TABLE 2. Pearson correlation of CCA features across various window lengths.

Window length (hours)	Correlation
1.0	0.275586
2.0	0.289476
3.0	0.385907
5.0	0.213667
7.0	0.132562
9.0	0.028756

TABLE 3. ROC AUC scores across the three scenarios. The results indicate slight degradation of performance in the downstream task of "diabetes classification" if only using HR CCA features compared to the other two scenarios that have CGM-based features (CGM PCA and CGM+HR CCA scenarios). Figure 4 shows the ROC curves corresponding to the cross-validated folds presented in this table.

	ROC AUC Score		
Corresponding Fold #	CGM PCA	HR CCA	CGM+HR CCA
Fold #1	0.969	0.927	0.976
Fold #2	0.901	0.920	0.931
Fold #3	0.891	0.930	0.944
Fold #4	0.943	0.903	0.967
Fold #5	0.959	0.905	0.971
Max	0.969	0.930	**0.976**
Min	**0.891**	0.903	0.931
Mean	0.932	0.917	**0.958**

glucose dysregulated (users that report having type 2 diabetes, prediabetes, or gestational diabetes, $N = 238$). As discussed at the end of Section III-C, we also compare classifying with latent CGM features that are based on PCA. The key distinction between training a classifier with the PCA latent features and Equation (1a) is that latent features $z_{u,i}^{hr}$ in Equation (1b) are learned to be maximally correlated with the CGM latent features $z_{u,i}^{cgm}$ according to the CCA objective.

We also control for BMI and age as these are known to be strong indicators of diabetes [9]. We concatenate the latent feature vectors with the BMI and age of the user. Since logistic regression can be sensitive to scaling, we normalize each feature to be centred at zero and have unit variance. We use L2 regularization and select the regularization strength based on nested cross-validation. We search for the best L2 regularization penalty using grid search swept logarithmically between 10^{-10} and 10^{10} spaced at 50 points. We use the scikit-learn implementation of CCA, PCA, and logistic regression [10].

Figures 3 shows the Area Under the Receiver Operator Characteristic Curve (AUC) as the number of components for CCA and PCA are varied. AUC values are averaged over 5 fold cross-validation across users and the error bars indicate one standard deviation. The scenarios that have CGM features on average have higher AUC than using the CCA-derived HR features. This is to be expected and thought to perform better and in the case of CGM+HR features be as an upper-bound since it benefits from both HR features and CGM features that are directly measuring glucose and glucose dysregulation is directly related to diabetes. We see however when using less than 5 components, the accuracy in diabetes detection for the CGM-only scenario is worse than using HR. This suggests the top components in CGM while capturing maximal variance, are not adequate in the downstream task of diabetes classification. However, as we increase the number of components in the PCA CGM features, we see a significant increase as expected.

In Figure 4 we show the Receiver Operator Characteristic Curves for each fold for three scenarios: HR CCA features (using 2 components), HR+CGM CCA features (using 2 components), and the CGM PCA features (using 10 components). As can be seen from the figure, for 3 out of 5 folds the ROCs for CGM PCA and HR CCA are similar. Also, for 4 out
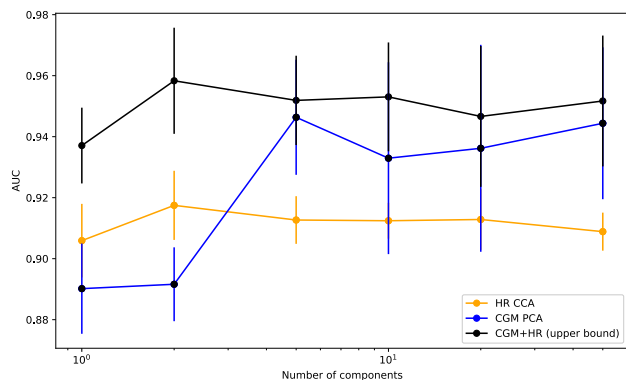
FIGURE 3. Comparing the AUC of three scenarios where there are: 1) HR features with a varying number of CCA components, 2) HR+CGM features with a varying number of CCA components, and 3) CGM features with varying number of PCA components. The orange curve does not use any CGM data for diabetes classification; it instead relies on the heart rate features that we have learned according to Equation (1b). For diabetes classification, we expect scenarios with CGM-derived features to perform better since the features are measuring glucose directly. Although CGM-derived features have much higher AUC for diabetes classification as expected, we see that features derived from HR can also become close in terms of accuracy.
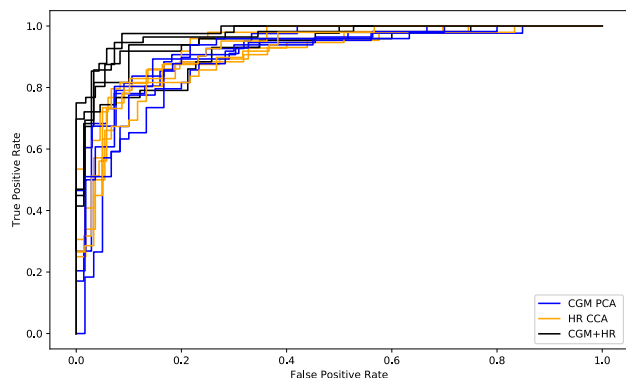

FIGURE 4. Comparing the ROC curves of three scenarios where there are: 1) HR features with a varying number of CCA components, 2) HR+CGM features with a varying number of CCA components, and 3) CGM features with varying number of PCA components. The orange curves are based on features derived from HR according to Equation (1b) whereas the blue curves are using CGM features, and the black curves are using both CGM and HR (CCA) features. Since CGM features are directly measuring glucose dysregulation, we expect better performance. However, the HR-derived features show there is a slight degradation in performance. Table 3 provides a numerical comparison of ROC AUC scores.

of 5 folds, the ROCs for the CGM+HR scenario outperform the other two scenarios.

Also, Figure 5 indicates the AUC Precision-Recall Curves for each fold for three scenarios: HR CCA features (using 2 components), HR+CGM CCA features (using 2 components), and the CGM PCA features (using 10 components). As can be seen from the figure, the AUCs for CGM PCA and HR CCA are similar for three of the cases. Also, for 4 out of 5 folds, the ROCs for the CGM+HR scenario outperform the HR CCA scenario and for 3 out of 5 folds outperform the CGM PCA scenario.

We stress that the CGM features - as our observations shown in Figures 4 and 5 confirm - for diabetes classification
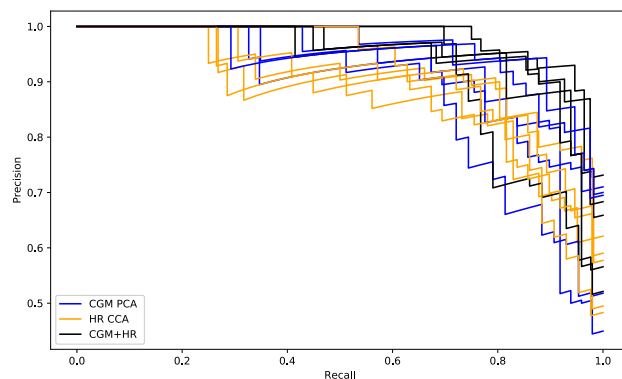

FIGURE 5. Comparing the AUC precision-recall curves of three scenarios where there are: 1) HR features with a varying number of CCA components, 2) HR+CGM features with a varying number of CCA components, and 3) CGM features with varying number of PCA components. The orange curves are based on features derived from HR according to Equation (1b) whereas the blue curves are using CGM features, and the black curves are using both CGM and HR (CCA) features. Since CGM features are directly measuring glucose dysregulation, we expect better performance. However, the HR-derived features show there is a slight degradation in performance.

is an upper-bound for diabetes classification using wearable devices since CGMs are measuring glucose directly. We see, however, the HR features that we have learned have a slight degradation in terms of accuracy.

V. CONCLUSION
The growing ubiquity of heart rate monitors provides unique opportunities for consumers to manage exercise and activity habits. We provide a methodology for expanding the utility of heart rate data to diabetes detection for patients and health-care providers. While CGMs are the ideal wearable device for glucose monitoring, our study suggests that machine learning can be used to adapt HR monitors to be an alternative in cases where CGM availability is limited.

COMPETING INTERESTS
All authors are affiliated with January, Inc.

ETHICAL STANDARDS
This research was conducted under IRB approval (120190429).

REFERENCES
[1] Y. Bengio, A. Courville, and P. Vincent, "Representation learning: A review and new perspectives," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 8, pp. 1798–1828, Aug. 2013.
[2] M. Christ, N. Braun, J. Neuffer, and A. W. Kempa-Liehr, "Time series FeatuRe extraction on basis of scalable hypothesis tests (tsfresh—A Python package)," *Neurocomputing*, vol. 307, pp. 72–77, Sep. 2018.

[3] S. L. Cichosz, J. Frystyk, O. K. Hejlesen, L. Tarnow, and J. Fleischer, "A novel algorithm for prediction and detection of hypoglycemia based on continuous glucose monitoring and heart rate variability in patients with type 1 diabetes," *J. Diabetes Sci. Technol.*, vol. 8, no. 4, pp. 731–737, Jul. 2014.

[4] H. Hall, D. Perelman, A. Breschi, P. Limcaoco, R. Kellogg, T. McLaughlin, and M. Snyder, "Glucotypes reveal new patterns of glucose dysregulation," *PLOS Biol.*, vol. 16, no. 7, Jul. 2018, Art. no. e2005143.

[5] H. Hotelling, "Relations between two sets of variates," in *Breakthroughs in Statistics*. New York, NY, USA: Springer, 1992, pp. 162–190.

[6] J. Lee, Y. Lee, J. Kim, A. Kosiorek, S. Choi, and Y. W. Teh, "Set transformer: A framework for attention-based permutation-invariant neural networks," in *Proc. Int. Conf. Mach. Learn. (PMLR)*, 2019, pp. 3744–3753.

[7] Y.-Q. Li, C.-Q. Sun, L.-L. Li, L. Wang, Y.-R. Guo, A.-G. You, Y.-L. Xi, and C.-J. Wang, "Resting heart rate as a marker for identifying the risk of undiagnosed type 2 diabetes mellitus: A cross-sectional survey," *BMC Public Health*, vol. 14, no. 1, pp. 1–7, Dec. 2014.

[8] D. K. McCulloch, D. Nathan, and J. Mulder, "Overview of medical care in adults with diabetes mellitus," UpToDate Online, Boston, MA, USA, 2009.

[9] K. M. V. Narayan, J. P. Boyle, T. J. Thompson, E. W. Gregg, and D. F. Williamson, "Effect of BMI on lifetime risk for diabetes in the U.S." *Diabetes Care*, vol. 30, no. 6, pp. 1562–1566, Jun. 2007.

[10] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine learning in Python," *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, 2011.

[11] A. Seyd, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Automated diagnosis of diabetes using heart rate variability signals," *J. Med. Syst.*, vol. 36, no. 3, pp. 1935–1941, 2012.

[12] D. Sisodia and D. S. Sisodia, "Prediction of diabetes using classification algorithms," *Procedia Comput. Sci.*, vol. 132, pp. 1578–1585, 2018.

[13] M. Stenerson, F. Cameron, D. M. Wilson, B. Harris, S. Payne, B. W. Bequette, and B. A. Buckingham, "The impact of accelerometer and heart rate data on hypoglycemia mitigation in type 1 diabetes," *J. Diabetes Sci. Technol.*, vol. 8, no. 1, pp. 64–69, Jan. 2014.

[14] G. Swapna, R. Vinayakumar, and K. P. Soman, "Diabetes detection using deep learning algorithms," *ICT Exp.*, vol. 4, no. 4, pp. 243–246, Dec. 2018.

[15] A. G. Tabák, C. Herder, W. Rathmann, E. J. Brunner, and M. Kivimäki, "Prediabetes: A high-risk state for developing diabetes," *Lancet*, vol. 379, no. 9833, p. 2279, 2012.

[16] B. Thompson, "Canonical correlation analysis," *Encyclopedia Statist. Behav. Sci.*, 2005.

[17] M. Zaheer, S. Kottur, S. Ravanbhakhsh, B. Póczos, R. Salakhutdinov, and A. J. Smola, "Deep sets," in *Proc. 31st Int. Conf. Neural Inf. Process. Syst.*, 2017, pp. 3394–3404.

[18] M. Zhang, B. Wang, Y. Liu, X. Sun, X. Luo, C. Wang, L. Li, L. Zhang, Y. Ren, Y. Zhao, J. Zhou, C. Han, J. Zhao, and D. Hu, "Cumulative increased risk of incident type 2 diabetes mellitus with increasing triglyceride glucose index in normal-weight people: The rural Chinese cohort study," *Cardiovascular Diabetology*, vol. 16, no. 1, pp. 1–11, Dec. 2017.

SOLMAZ SHARIAT TORBAGHAN was born in Tehran, Iran. She received the B.Sc. degree in computer science from the University of Tehran, in 2008, the M.Sc. degree in computational and systems biology from the KTH Royal Institute of Technology, Stockholm, Sweden, in 2010, and the M.A. degree in psychology/cognitive neuroscience from New York University (NYU), in 2016. She was a Neuroscientist with Princeton University and NYU, where she worked on visual neuroscience and AI. Her research interests include AI, active learning, and deep learning in healthcare. She received the Fellowship from the University of California at Los Angeles (UCLA), in 2010, to conduct research in neuroscience of visual attention.

SALAR RAHILI (Member, IEEE) received the B.S. and M.Sc. degrees in electrical engineering from the Isfahan University of Technology, Isfahan, Iran, in 2009 and 2012, respectively, and the Ph.D. degree in electrical engineering from the University of California at Riverside, Riverside, CA, USA, in 2016. From 2016 to 2018, he was a Postdoctoral Researcher with the Department of Aerospace, California Institute of Technology, USA. His research interests include deep reinforcement learning, deep learning, and machine learning.

MICHAEL SNYDER is a Stanford Ascherman Professor and the Chair of genetics and the Director of the Center of Genomics and Personalized Medicine. He is also the Co-Founder of January AI. He is a leader in the field of functional genomics and multiomics, and one of the major participants of the ENCODE project. He received the Ph.D. training from the California Institute of Technology and carried out postdoctoral training at Stanford University.

HOOTAN RASHTIAN received the M.Sc. and Ph.D. degrees in computer engineering from The University of British Columbia (UBC), Vancouver, Canada. He is currently a Machine Learning Lead at January AI, to provide personalized health insights and recommendations. His academic and industry experiences are in the areas of deep reinforcement learning, deep learning, and machine learning. He received the NSERC and the MITACS scholarships during his graduate studies.

NIMA AGHAEEPOUR received the B.Sc. degree from the University of Tehran and the Ph.D. degree from The University of British Columbia. He is a member of the Scientific Advisory Board at January AI. His career is focused on the development of machine learning and artificial intelligence methods to study clinical and biological modalities in translational settings. He received the Postdoctoral Fellowship from Stanford University.

• • •