Machine learning prediction of severity and duration of hypoglycemic events in type 1 diabetes patients

Annie S. Wu, Eashan Singh, Ivy Zhang
Department of Computer Science
University of Central Florida
Orlando, FL 32816
aswu@cs.ucf.edu, {ea363606, Ivy.Zhang}@ucf.edu

Anika Bilal, Anna Casu, Richard Pratley
AdventHealth Translational Research Institute
Orlando, FL 32789
{Anika.Bilal, Anna.Casu}@AdventHealth.com
Richard.Pratley.MD@AdventHealth.com

Abstract

We compare the performance of machine learning methods for building predictive models to estimate the expected characteristics of hypoglycemic or low blood glucose events in type 1 diabetes patients. We hypothesize that the rate of change of blood glucose ahead of a hypoglycemic event may affect the severity and duration of the event and investigate the utility of machine learning methods on using blood glucose rate of change, in combination with other physiological and demographic factors, to predict the minimum glucose value and the duration of a hypoglycemic event. This work compares the performance of six state-of-the-art methods on prediction accuracy and feature selection. Results find that XGBoost delivers the best performance in all cases. Examination of the XGBoost feature importance scores show that glucose rate of change is the most used feature in the models generated by XGBoost.

Introduction

In this work, we examine the effectiveness of six machine learning (ML) methods on building a model to predict the expected characteristics of hypoglycemic or low blood glucose events in patients with type 1 diabetes (T1D). This model will use patient medical history, demographic data, and physiological measures to predict the severity and duration of hypoglycemic events. In addition commonly collected patient data, we hypothesize that the rate of change of blood glucose ahead of a hypoglycemic event may affect the severity and duration of the event and include this novel measure as an additional input to our models. Our primary goal compares the performance of state-of-the-art ML methods on this problem to identify the most effective methods. A secondary goal examines whether the blood glucose rate of change is a relevant input to the learned models.

T1D is the inability of the body to make insulin which is needed to process blood glucose into energy. This condition results in a build up of blood glucose in the bloodstream that must be treated via external injections of insulin. Patients with T1D must manually manage their blood glucose levels within target levels to avoid complications from both hypoglycemia (low blood glucose) and hyperglycemia (high blood glucose). T1D affects over 1.7 million Americans (CDC 2024) and impaired awareness of hypoglycemia (IAH) occurs in 20-30% of T1D patients (McNeilly and McCrimmon 2018). IAH is a diminished ability to recognize the symptoms of hypoglycemia which, if untreated, can put the patient at significant immediate risk. Previous work examines the use of continuous glucose monitoring (CGM) data and other measures to predict the occurence of hypoglycemic events (Cichosz et al. 2014; Dave et al. 2020; Fleischer, Hansen, and Cichosz 2022; Woldaregay et al. 2019). Understanding the factors that affect the severity and duration of hypoglycemic events may help us to develop predictive tools that can aid an IAH patient in recognizing the onset of a hypoglycemic event or possibly avoiding an event altogether.

The Wireless Innovation for Seniors With Diabetes Mellitus (WISDM) study (Pratley et al. 2020) is a randomized clinical trial (RCT) investigating the potential benefits of CGM over standard blood glucose monitoring (BGM) on reducing hypoglycemia in adults aged 60 or higher. This study followed 203 participants over a period of one year. Participants were randomly assigned to one of two groups: the BGM group used BGM to regulate blood glucose for first 26 weeks of the study, the CGM group used CGM. All participants used CGM to regulate blood glucose for the second 26 weeks of the study. CGM data was collected for all participants[1] in 10-14 day intervals at weeks 0 (baseline), 4, 8, 16, 26, 39, and 52 and analyzed with respect to the percent time that measured glucose values were less than 70 mg/dL (hypoglycemia) and the percent time within 70-180 mg/dL (in target range). Within each collection interval, CGM blood glucose measurements are recorded every five minutes.

Analysis of the WISDM participant baseline data indicates that participants with IAH spend more than twice as much time in hypoglycemia as those who do not experience IAH (Carlson et al. 2021). RCT results find that the use of CGM results in notable and sustainable drops in participants' percent time spent below 70 mg/dL as well as moderate improvement in percent time spent in range (Pratley et al. 2020; Miller et al. 2022). These findings suggest that cou-

[1]The CGM device collecting data from BGM participants was masked and did not reveal measured values to participants.

Label	Data type	Description
Gender	Binary	0 = Male, 1 = Female
AgeAsOfEnrollDt	Integer	Age as of enrollment in study
DiagAge	Integer	Age at diagnosis for type 1 diabetes
DiabDuration	Integer	Diabetes duration. (current age - DiagAge)
BMI	Float	Body mass index
PumpUse	Binary	Does participant use an insulin pump: 0 = No, 1 = Yes
TrtGroup	Binary	WISDM study treatment group in weeks 0-26: 0 = BGM, 1 = CGM
Increase/Decrease	Binary	Did glucose management improve from week 0-26: 0 = no, 1 = yes
InsulinDosesKg	Float	Total daily insulin dose in units/kg at baseline
Cpep_detected	Binary	Detectable C-peptide levels at baseline: 0 = No, 1 = Yes
HbA1cTestRes	Float	Baseline HbA1c test result
gluBelow70	Float	Percent glucose measurements < 70 mg/dL at baseline
gluInRange	Integer	Percent glucose measurements within the range of 70-180 mg/dL at baseline
gluCV_num	Integer	Glucose coefficient of variation (glycemic variability) at baseline
SHNumLast12Months	Binary	Experienced severe hypoglycemic events in the last 12 months: 0 = No, 1 = Yes
DKANumLast12Months	Binary	Experienced diabetic ketoacidosis events in the last 12 months: 0 = No, 1 = Yes
week-$n^{\ddagger}$	Integer	n specifies week in which hypoglycemic event occurred, $n \in$ 0, 4, 8, 16, 26, 39, 52
LessThanBachelors†	Binary	Maximum education level less than Bachelor's degree: 0 = False, 1 = True
Bachelors†	Binary	Maximum education level at Bachelor's degree: 0 = False, 1 = True
Private_Ins	Binary	Private insurance: 0 = False, 1 = True
Medicare_Ins	Binary	Medicare insurance: 0 = False, 1 = True

Table 1: Independent variables from (Kahkoska et al. 2023) study. ‡Seven different week-n variables are included. †If LessThanBachelors = 0 and Bachelors = 0, then the maximum education level is greater than a Bachelor's degree.

pling CGM with predictive models for estimating the onset or severity of hypoglycemic events could be a useful tool for helping patients reduce time spent in hypoglycemia.

We investigate the ability of six ML methods to build a predictive model for estimating the minimum glucose value and the duration of hypoglycemic events. We train and test these algorithms on a dataset of individual hypoglycemic events extracted from the data collected in the WISDM study and evaluate the root mean square error (RMSE) of the predicted values as compared to the actual values. Each data point in our dataset represents the data for a single hypoglycemic event and our dataset consists of hypoglycemic event data collected from multiple patients over a period of one year. As a result, the models that we build are meant to be representative across patients, and are not patient-specific models.

Description of dataset

The dataset that we use in this study derives from the data from the WISDM study. We define a hypoglycemic event to be at least 15 consecutive minutes with a CGM glucose value lower than 60 mg/dL. The end of a hypoglycemic event is defined as a minimum of 15 consecutive minutes with a CGM glucose concentration greater than 70 mg/dL. Based on this definition, we identify 7973 instances of hypoglycemic events in the data from the WISDM study. These instances form our experimental dataset.

Each data point in our dataset represents a single hypoglycemic event and consists of 27 independent variables and two dependent variables. The independent variables are selected based on a previous ML study on the WISMD data (Kahkoska et al. 2023). Table 1 lists the independent variables. The two dependent variables that our models will predict are the event severity (or minimum event glucose value) and the event duration (or time until return to safe glucose levels).

In order to examine our hypothesis that the rate of change of blood glucose may affect the severity and duration of the event, we include additional independent variables representing two methods for measuring the rate of change in blood glucose in the 30 minutes immediately preceeding a hypoglycemic event, R: (1) the slope of the blood glucose calculated from measures taken 30 minutes before and at the start of an event and (2) individual glucose measures taken in five minute intervals during the 30 minutes before an event. These additional independent variables are given in Table 2.

Experimental details

We compare the performance of six ML methods on the following two predictive modeling problems:

- **Problem 1**: Predict the minimum event glucose value
- **Problem 2**: Predict the time to safe glucose levels

For each of these problems we test two different datasets:

- **Dataset A**: Each data point has 28 independent variables including all variables from Table 1 and the slope variable from Table 2.
- **Dataset B**: Each data point has 33 independent variables including all variables from Table 1 and the 5-minute interval glucose measurement variables from Table 2.

Each dataset is randomly divided into a training set (80%) and testing set (20%).

Label	Data type	Description
Slope	Float	Slope of glucose measurements in the 30 minutes immediately preceding the hypoglycemic event
glu_value5before	Float	Measured blood glucose 5 minutes before start of hypoglycemic event.
glu_value10before	Float	Measured blood glucose 10 minutes before start of hypoglycemic event.
glu_value15before	Float	Measured blood glucose 15 minutes before start of hypoglycemic event.
glu_value20before	Float	Measured blood glucose 20 minutes before start of hypoglycemic event.
glu_value25before	Float	Measured blood glucose 25 minutes before start of hypoglycemic event.
glu_value30before	Float	Measured blood glucose 30 minutes before start of hypoglycemic event.

Table 2: Independent variables measuring the rate of change of glucose in the 30 minutes before a hypoglycemic event, R.

XGBoost				
Problem	**1**	**1**	**2**	**2**
Dataset	**A**	**B**	**A**	**B**
colsample_bytree	0.8	0.7	0.9	0.9
gamma	0.0	0.0	0.0	0.0
learning_rate	0.1	0.1	0.01	0.01
max_depth	3	3	9	7
min_child_weight	1	1	3	5
subsample	0.7	0.9	0.6	0.7

Deep Learning		
DNN Architecture	**Small**	**Medium**
Optimizer	AdamW	AdamW
Learning Rate	0.001	0.005
Loss Function	MSEL	MSEL
Epochs	100	100
Batch size	32	32
Dropout rate (for regularization)	0.5	0.5
Activation function	ReLU	ReLU

Table 3: Independently optimized parameter settings for each tested method. MSEL = Mean Squared Error Loss. For decision tree and random forest, the random_state variable is set to 42. All linear regression parameters use default library values.

The ML methods that we study are decision trees (Quinlan 1986), random forest (Ho 1995), extreme gradient boosting (XGBoost) (Chen and Guestrin 2016)), linear regression (Weisberg 2005), and two deep neural network (DNN) architectures (LeCun, Bengio, and Hinton 2015). We implement the first four methods using the scikit-learn library and the deep learning methods using the PyTorch library. Each method is optimized empirically. Table 3 gives the optimized parameter settings for each method. Parameters that are not listed are set to the library default values. Each ML method is trained on the training set of each dataset, then tested on the corresponding test set. Performance is given as the RMSE of the test set. RMSE measures the average difference between the predicted and actual values. Lower RMSE values indicate better performance.

Results

Comparison of prediction error

Table 4 shows the performance of the tested models on predicting the minimum glucose value or severity of a hypoglycemic event. Column two gives the RMSE in mg/dL for the models created using Dataset A; column three, for Dataset B. The best performance is indicated in bold type. The models are ordered from best to worst performance and, interestingly, the relative performance of the six models is the same regardless of whether R is specified as a single slope value or as multiple measurements taken in five minute intervals. XGBoost delivers the best performance on both datasets with an RMSE of 5.982 mg/dL on Dataset A and 5.879 mg/dL on Dataset B. Among the hypoglycemic events identified in the WISDM data, the range of minimum glucose values runs from 39 to 60 mg/dL. As a result, the best RMSE of 5.879 mg/dL represents a prediction error rate that is 5.87/21 = 27.95% of the data range.

Model	Dataset A	Dataset B
XGBoost	**5.982**	**5.879**
Random forest	6.006	5.964
Linear regression	6.041	6.156
Medium DNN architecture	6.317	6.409
Small DNN architecture	6.628	6.638
Decision tree	8.374	8.516

Table 4: Model performance on predicting severity of a hypoglycemic event. Values indicate RMSE in mg/dL. Best performance shown in bold.

Table 5 shows the performance of the tested models on predicting the time to safe glucose levels or duration of a hypoglycemic event. Column two gives the RMSE in minutes for the models created using Dataset A; column three, for Dataset B. The best performance is indicated in bold type. Once again, the relative performance of the six models is the same regardless for both R. XGBoost again delivers the best performance on both datasets with an RMSE of 46.058 minutes on Dataset A and 44.895 minutes on Dataset B. The duration of the hypoglycemic events identified in the WISDM data ranges from 25 to 709 minutes. Thus, the best RMSE of 44.895 minutes represents a prediction error rate that is 44.895/684 = 6.56% of the data range.

Feature importance in XGBoost

Because XGBoost consistently exhibits the best performance of all methods tested, we further examine the models created by this method. For each feature in each model

Model	Dataset A	Dataset B
XGBoost	**46.058**	**44.895**
Linear regression	47.644	46.621
Medium DNN architecture	50.978	47.671
Random forest	51.119	50.056
Small DNN architecture	51.259	50.272
Decision tree	76.075	74.842

Table 5: Model performance on predicting duration of a hypoglycemic event. Values indicate RMSE in minutes. Best performance shown in bold.

learned, XGBoost provides an F-score that indicates the importance of that feature in that model. The F-score specifies the frequency with which a given feature is used as a node in an XGBoost classification tree. Higher F-scores indicate greater importance.

Figure 1 shows the number of times features are used as decision nodes to predict the minimum glucose value for a hypoglycemic event. The x-axis indicates F-score and the y-axis specifies the features ordered from lowest to highest F-score. Only features used one or more times are included in the plots. For Dataset A, 27 of the 28 features are used in the model. For Dataset B, 32 of the 33 features are used in the model. For both Datasets A and B, the features that specify R are the most used features in the XGBoost classification trees. The top feature for Dataset A is the slope of the blood glucose measurements in the 30 minutes immediately preceeding the hypoglycemic event. Interestingly, for Dataset B, glu-value30before and glu-value5before are the top features; these features are the values used to calculate the slope values in Dataset A. InsulinDosesKg, gluBelow70, gluInRange, and HbA1cTestRes also rank highly for both datasets. The relative positioning of the features is similar for both datasets which indicates that XGBoost's results are consistent and reproducible. All of the week-n features are clustered, suggesting that all weeks have similar importance.

Figure 2 shows the number of times features are used as decision nodes to predict the expected duration of a hypoglycemic event. The x-axis indicates F-score and the y-axis specifies the features ordered from lowest to highest F-score. Only features used one or more times are included in the plots. For Dataset A, 27 of the 28 features are used in the model. For Dataset B, 31 of the 33 features are used in the model. Once again, the top feature for Dataset A is the slope and the top features for Dataset B are glu-value30before and glu-value5before. The relative ordering of the features are again similar for both datasets. The selected features in these two plots differ slightly from the selected features in Figure 1; however, the relative position of the features that appear in both plots are similar.

Conclusions

In this paper, we compare the performance of machine learning methods for building predictive models to estimate the severity and duration of hypoglycemic events in type 1 diabetes patients. We examine the performance of XGBoost, random forest, decision trees, linear regression, and two DNN architectures, and conduct experimental studies on two data sets consisting of hypoglycemic event data collected from multiple patients over a duration of one year. The two datasets differ in the independent variables of the data points. Although most independent variables are identical, the variables that represent the rate of change in blood glucose preceeding a hypoglycemic event are recorded as a single slope value in one dataset and as multiple data points in the second.

Experimental results find that XGBoost consistently performs better than all other methods tested. Random forest, linear regression and the medium DNN architecture perform slightly worse than XGBoost. The small DNN architecture and decision trees consistently perform the worst. All algorithms perform better on the problem of predicting the duration of a hypoglycemic event and perform worse on predicting the lowest glucose value of an event.

Examination of the F-scores of the XGBoost models find that almost all available features (independent variables) are used in the XGBoost classification trees. In all experiments, the features related to the rate of change of blood glucose levels emerge as the most important features. After that, a set of five features from participant baseline measures make up the next most important features. The relative ranking of feature importance is similar across all experiments, but not identical. Participant demographic features appear to be the least important features according to XGBoost.

That the features related to the rate of change of blood glucose are the most important features in all of our experiments may be due in part to the fact that those features are event-specific features, which is likely to be more useful in distinguishing events. Most of the other features are patient-specific features. The only other event-specific features are the week-n features which appear to have similar relevance as they are typically clustered together in the middle of the importance ordering. If we consider all of those features as a single *timing* feature and sum all of their F-scores into a single score, the combined F-score would make the timing feature the second most important feature in three of our four experiments and third in the fourth.

The importance that XGBoost places on the slope, the 5-minute interval measurements, and the combined timing features suggests that including additional event-specific measures such as synchronized physiological or activity data could improve the predictive power of the learned models. This work is part of a larger study that will be collecting and analyzing data on a new cohort of participants. In future work, we hope to collect additional synchronized streams of time-series data to be included as additional inputs to building these models.

Acknowledgments

This work is supported by the National Institutes of Health.

References

Carlson, A. L.; Kanapka, L. G.; Miller, K. M.; Ahmann, A. J.; Chaytor, N. S.; Fox, S.; Kiblinger, L.; Kruger, D.; Levy, C. J.; Peters, A. L.; Rickels, M. R.; Salam, M.; Shah,

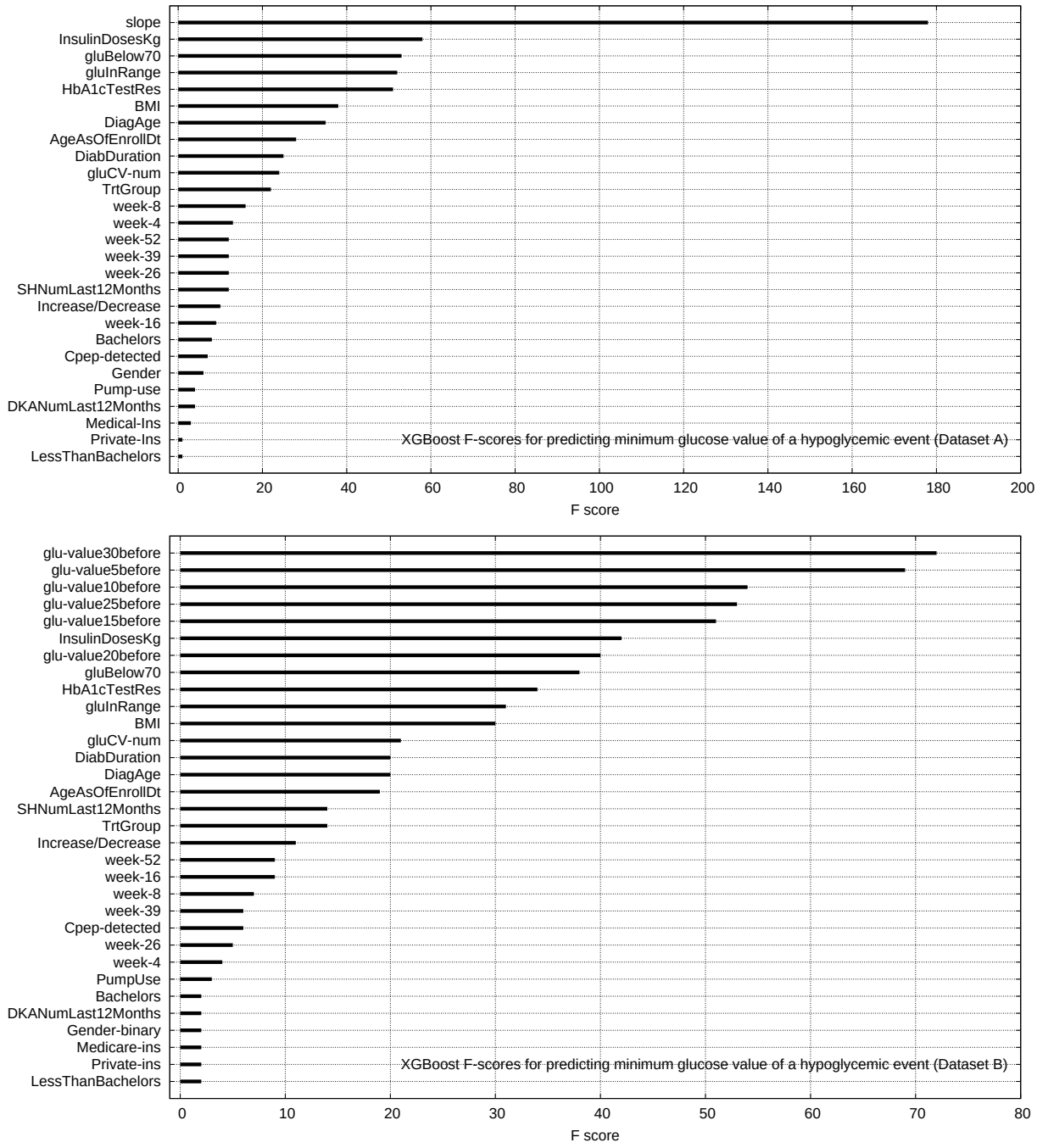

Figure 1: Frequency at which features are used as decision nodes in XGBoost classification trees in models trained to predict the severity of a hypoglycemic event.

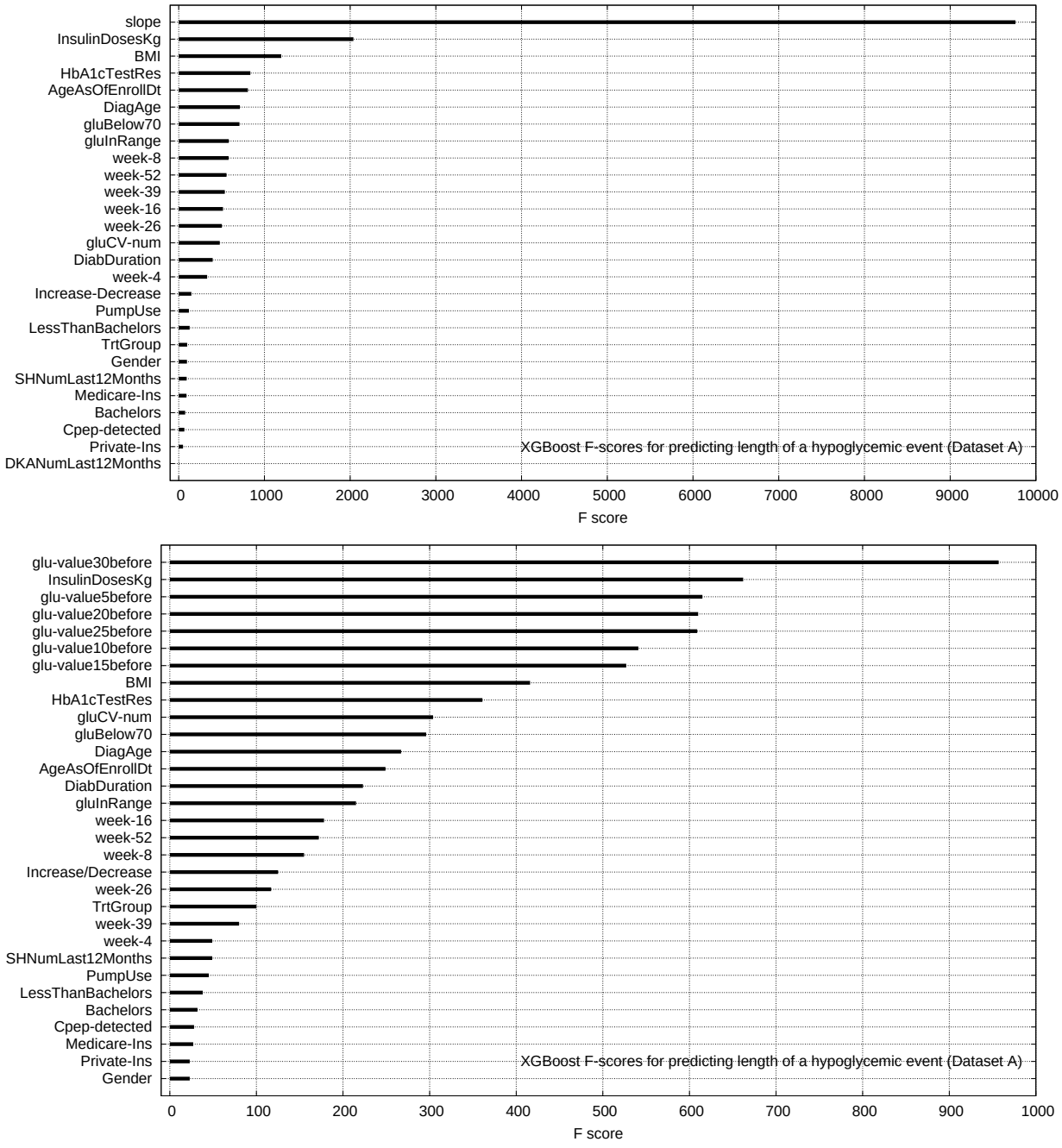

Figure 2: Frequency at which features are used as decision nodes in XGBoost classification trees in models trained to predict the duration of a hypoglycemic event.

V. N.; Young, L. A.; Kudva, Y. C.; and Pratley, R. 2021. Hypoglycemia and glycemic control in older adults with type 1 diabetes: Baseline results from the WISDM study. *Journal of Diabetes Science and Technology* 15(3):582–592.

CDC. 2024. Prevalence of diagnosed diabetes. Accessed 1/14/2024, https://www.cdc.gov/diabetes/data/statistics-report/diagnosed-diabetes.html.

Chen, T., and Guestrin, C. 2016. XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM International Conference on Knowledge Discovery and Data Mining*, 785–794.

Cichosz, S. L.; Frystyk, J.; Hejlesen, O. K.; Tarnow, L.; and Fleischer, J. 2014. A novel algorithm for prediction and detection of hypoglycemia based on continuous glucose monitoring and heart rate variability in patients with type 1 diabetes. *Journal of Diabetes Science and Technology* 8(4):731–737.

Dave, D.; DeSalvo, D. J.; Haridas, B.; McKay, S.; Shenoy, A.; Koh, C. J.; Lawley, M.; and Erraguntla, M. 2020. Feature-based machine learning model for real-time hypoglycemia prediction. *Journal of Diabetes Science and Technology* 15(4):842–855.

Fleischer, J.; Hansen, T. K.; and Cichosz, S. L. 2022. Hypoglycemia event prediction from cgm using ensemble learning. *Frontiers in Clinical Diabetes and Healthcare* 3:1066744.

Ho, T. K. 1995. Random decision forests. In *Proceedings of the 3rd International Conference on Document Analysis and Recognition*, 278–282.

Kahkoska, A. R.; Kosorok, M. R.; Miller, K. M.; Rickels, M.; Weinstock, R. S.; Young, L. A.; and Pratley, R. E. 2023. Estimation of a machine learning-based decision rule to reduce hypoglycemia among older adults with type 1 diabetes: A post hoc analysis of continuous glucose monitoring in the WISDM study. *Journal of Diabetes Science and Technology* 19322968221149040.

LeCun, Y.; Bengio, Y.; and Hinton, G. 2015. Deep learning. *Nature* 521:436–444.

McNeilly, A. D., and McCrimmon, R. J. 2018. Impaired hypoglycaemia awareness in type 1 diabetes: Lessons from the lab. *Diabetologia* 61(4):743–750.

Miller, K. M.; Kanapka, L. G.; Rickels, M. R.; Ahmann, A. J.; Aleppo, G.; Ang, L.; Bhargava, A.; Bode, B. W.; Carlson, A.; Chaytor, N. S.; Gannon, G.; Goland, R.; Hirsch, I. B.; Kiblinger, L.; Kruger, D.; Kudva, Y. C.; Levy, C. J.; andi Grenye O'Malley, J. B. M.; Peters, A. L.; Philipson, L. H.; Philis-Tsimikas, A.; Pop-Busui, R.; Salam, M.; Shah, V. N.; Thompson, M. J.; Vendrame, F.; Verdejo, A.; Weinstock, R. S.; Young, L.; and Pratley, R. 2022. Benefit of continuous glucose monitoring in reducing hypoglycemia is sustained through 12 months of use among older adults with type 1 diabetes. *Diabetes Technology and Therapeutics* 24(6):424–434.

Pratley, R. E.; Kanapka, L. G.; Rickels, M. R.; Ahmann, A.; Aleppo, G.; Beck, R.; Bhargava, A.; Bode, B. W.; Carlson, A.; Chaytor, N. S.; Fox, D. S.; Goland, R.; Hirsch, I. B.; Kruger, D.; Kudva, Y. C.; Levy, C.; McGill, J. B.; Peters, A.; Philipson, L.; Philis-Tsimikas, A.; Pop-Busui, R.; Shah, V. N.; Thompson, M.; Vendrame, F.; Verdejo, A.; Weinstock, R. S.; Young, L.; and Miller, K. M. 2020. Effect of continuous glucose monitoring on hypoglycemia in older adults with type 1 diabetes: A randomized clinical trial. *Journal of the America Medical Association* 323(23):2397–2406.

Quinlan, J. R. 1986. Induction of decision trees. *Machine Learning* 1:81–106.

Weisberg, S. 2005. *Applied Linear Regression*. John Wiley & Sons,.

Woldaregay, A. Z.; Årsand, E.; Walderhaug, S.; Albers, D.; Mamykina, L.; Botsis, T.; and Hartvigsen, G. 2019. Data-driven modeling and prediction of blood glucose dynamics: Machine learning applications in type 1 diabetes. *Artificial Intelligence in Medicine* 98:109–134.