Predicting Glycemia in Type 1 Diabetes Patients: Experiments with XGBoost

Cooper Midroni, Peter J. Leimbigler, Gaurav Baruah, Maheedhar Kolla, Alfred J. Whitehead, Yan Fossat
Klick Inc., 175 Bloor Street East, Toronto, Ontario, Canada
{cmidroni,pleimbigler,gbaruah,mkolla,awhitehead,yfossat}@klick.com

Abstract

Type 1 diabetes patients must self-administer insulin through injections or insulin-pump therapy, requiring careful lifestyle management around meals and physical activity. Accurate blood glucose prediction could increase patient quality of life, and foreknowledge of hypoglycemia or hyperglycemia could mitigate risks and save lives. For the 2018 BGLP Challenge, we experiment primarily with XGBoost to predict blood glucose levels at a 30-minute horizon in the OhioT1DM dataset. Our experiments show that XGBoost can be a competitive predictor of blood glucose levels, as compared to prior research, and that feature signals from different sources contribute in varying capacity for improved predictive ability of XGBoost.

1 Introduction

Diabetes affects over 400 million people worldwide [World Health Organization, 2016], with near 5% of diabetics suffering from type 1 diabetes (T1D) [American Diabetes Association, 2018]. Patients with T1D are incapable of producing insulin, a hormone generated by the pancreas, which acts as the primary regulator of blood glucose metabolism. This dysfunction can lead to both hypoglycemia (low blood sugar) and hyperglycemia (high blood sugar), resulting in a significant patient burden to regulate carbohydrate consumption and supplemental insulin delivery. Hyperglycemia can lead to medical complications such as vision loss and kidney failure, and increases risk of heart disease and stroke. Hypoglycemia can lead to loss of consciousness and even death.

An increasing number of T1D patients are adopting insulin pump therapy, wherein a wearable device releases insulin subcutaneously to mimic pancreatic response. Current insulin pumps require patient input on carbohydrate intake and approval of each recommended insulin dose. Driven by the outstanding need for closed-loop insulin therapy, the notion of an artificial pancreas has gained traction in diabetes-related research [Graf *et al.*, 2017; Juvenile Diabetes Research Foundation, 2018].

The dysregulation of blood glucose in T1D patients is further complicated by daily variations in the magnitude and timing of meals, physical activity, and insulin self-administration. This, along with the altered pharmacokinetics of subcutaneous insulin, add further layers of complexity to the task of predicting blood glucose. As such, the Blood Glucose Level Prediction Challenge represents an important step toward the realization of an artificial pancreas. Herein lies the objective of restoring homeostasis through accurately dosed insulin, and the creation of a model which captures the complexities of the disease. This challenge was particularly motivating for us, not simply from the perspective of predictive modeling, but also for the potential applications in providing tangible benefits to T1D patients.

To predict glucose at a 30-minute time horizon, we processed the raw features of the OhioT1DM dataset [Marling and Bunescu, 2018] to create 3 different feature sets, and experimented with gradient-boosted trees [Chen and Guestrin, 2016a] (XGBoost), random forests [Breiman, 2001], and recurrent neural network variants. We find that:

- XGBoost performs on par with prior models [Bunescu *et al.*, 2013; Mirshekarian *et al.*, 2017] for this task.
- Many of the provided features do not contribute to improved predictive performance. In essence, when using XGBoost, past glucose is the most important predictor of future glucose.
- Using ϵ-insensitive loss for training LSTMs improves predictive performance compared to mean squared error.

2 Data

The OhioT1DM dataset comprises 19 features collected from 6 patients with T1D [Marling and Bunescu, 2018]. Patients wore Medtronic 530G insulin pumps, Medtronic Enlite continuous glucose monitors (CGM), and Basis Peak fitness wristbands. 8 weeks of data were provided per patient, of which the final 10 days were provided separately as a test set.

2.1 Data Analysis

Due to the heterogeneity of the data, we grouped features according to their frequency:

- *one-off data*: intermittent measurements with no fixed sampling frequency or duration. (e.g., finger stick glucose, insulin bolus time and dose, sleep times and quality, work intensity, exercise intensity and duration, meal

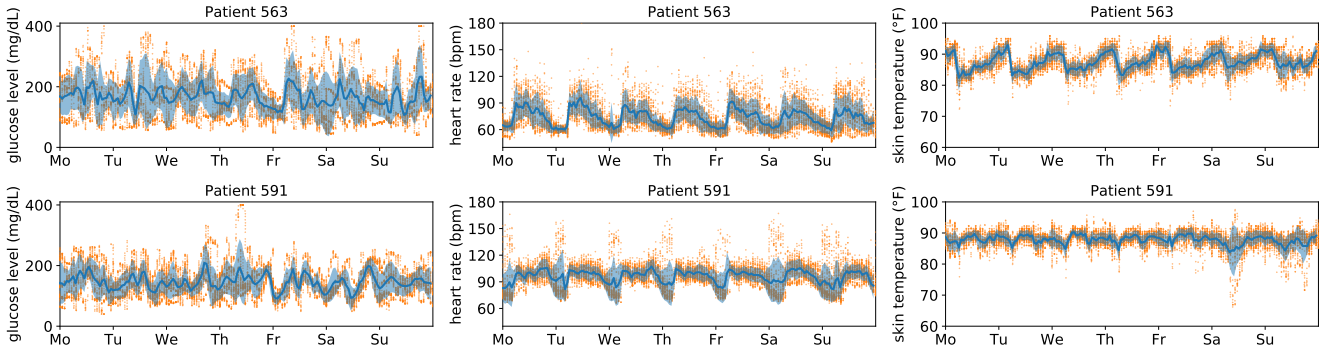

Figure 1: Selected features for two patients, illustrating differences in mean values and waveforms. Lines and shaded regions indicate mean ±1 SD. Orange dots show individual data samples.

type and carbohydrate content, hypoglycemic events, illnesses, stressors)

- *quasi-continuous data*: signals in continuous effect, and signals aggregated at 5-minute intervals. (e.g., CGM glucose level, basal and temporary basal rates of insulin infusion, heart rate, steps taken, galvanic skin response (GSR), skin and air temperature)

Analysis of quasi-continuous data showed unique temporal patterns in each patient (Figure 1). These bio-signals displayed properties characteristic to each individual, such as repeated waveforms and unique signal means. These qualities reflect the hour-of-day periodicities and homeostatic norms that vary across patients. Accordingly, we were motivated to build a distinct model for each patient, as opposed to a general model trained on the data of several patients.

2.2 Feature Engineering

Expanded Feature Set

We found that due to the variable sampling frequency of one-off features and missing values in the data, the feature vector at any given timestamp was not guaranteed to contain values for all fields. Converting the data into a feature matrix resulted in rows with missing values, hindering analysis. We thus chose to resample our data to 5-minute intervals, reflecting the 5-minute aggregation frequency of quasi-continuous variables.

Within each 5-minute resampling window, we aggregated each feature with either its maximum, mean, or last valid value, depending on its nature. For instance, we took the sum of steps, whereas we computed the mean of heart rate. Some features are only in effect for specified durations. For instance, basal insulin infusion rates were overridden with respective temporary basal overrides, if any, and each square-wave insulin bolus dose was spread evenly across its specified time interval. Missing values for glucose were imputed via linear interpolation.

Based on our observations (Sec. 2.1) of time-dependent patterns in the data, e.g., the dawn phenomenon, we included one-hot encoded features for hour-of-day and day-of-week. For each one-off feature, a binary indicator feature [Che *et al.*, 2018] was used to denote missing values.

Condensed Feature Set

We developed a *condensed* feature set based on pairwise correlations between features. This feature set included Basis Peak band and Medtronic CGM sensor data, based on the strength of their correlation with glucose. Several derivations of the glucose signal were added to the feature set, including five- and ten-minute time lags of glucose to provide the model with information of glucose history. A binary indicator feature marked when the present glucose level was in the upper or lower 20% of the patient's glucose distribution. Other features present in this set include the last bolus dose, mean basal rate over the past 5 minutes, and an indicator of whether the patient was asleep.

Dimensionality-Reduced Features (PCA-reduced)

We used Principal Component Analysis (PCA) to transform our expanded feature set to remove features with minimal variance. Such features are less likely to be predictive in a model. On average, the first 55 principal components accounted for 99% of the variance in each patient's dataset.

3 Machine Learning Models

For all models, we used the final week of each patient's training data for validation. In training personalized models, hyperparameters and model structure (e.g. learning rates, number of LSTM nodes) were kept consistent across patients. All models in this study were implemented on all three of the engineered feature sets.

We investigated the following models:

- Tree ensembles using the Random Forest Regressor implementation of Scikit-Learn [Pedregosa *et al.*, 2011].
- Regression-based gradient-boosted decision trees using XGBoost [Chen and Guestrin, 2016b].
- Recurrent neural network variants using Keras [Chollet, 2015].

Keras was used to create several RNN models: multi-layer LSTMs, GRUs, a Bidirectional GRU, and LSTMs with dropout. These were implemented to evaluate their performance on the data, and were tested on varying durations of look-back windows (5 minutes–1 day).

Table 1: Submitted system performances.

Model: Featureset:	RF condensed	XGBoost expanded	XGBoost PCA-reduced	LSTM expanded ϵ-insens.	LSTM expanded MAE loss
559	37.236	19.810	21.816	23.949	23.76
563	28.095	18.415	19.375	25.121	24.901
570	24.625	18.140	19.614	19.562	20.497
575	28.688	24.172	24.242	25.923	27.796
588	23.894	19.240	22.141	19.012	20.303
591	28.976	22.487	23.391	27.333	30.256
AVG	**28.586**	**20.377**	**21.763**	**23.483**	**24.586**

3.1 System Performance Results

We found that no one feature set (either expanded, condensed or PCA-reduced) produced consistently better glucose predictions, and different models performed better on different feature sets. Table 1 lists our submitted systems and feature sets.

XGBoost was the best-performing model on both the expanded and PCA-reduced feature sets, achieving a mean RMSE across all patients of 20.377. These results are at par with previously published models based on Support Vector Regression [Bunescu *et al.*, 2013].

Experiments with LSTM Loss Functions

Our LSTM models were simple and did not perform as well as recent LSTM models for blood-glucose prediction [Mirshekarian *et al.*, 2017]. We submitted results for an LSTM model that was composed of: layers `LSTM(64 nodes)`, `LSTM(64 nodes)`, `Fully-connected(32 nodes)`; a `dropout` rate of 0.2; an `Adadelta` optimizer; and a look-back of 5 minutes.

We observed that Mean Absolute Error (MAE) improved the performance of trained LSTMs over using Mean Squared Error (MSE) as the loss function. The models trained with MSE showed a degradation which was particularly severe for glucose values near hypoglycemic and hyperglycemic levels [Medtronic, 2010]. MAE, in contrast to MSE, does not penalize large errors as heavily as MSE, which likely helped improve performance over outlying cases.

The generally accepted error rate for finger-stick blood glucose measurements is 15 mg/dL [Food and Drug Administration, 2016]. Thus, for predictions within a 15 mg/dL window of the ground truth, the the loss for such values can be considered less impactful. We therefore investigated the use of an ϵ-insensitive loss function for training our LSTM, with ϵ set as 5 mg/dL, for a more stringent boundary than 15 mg/dL.

We compared the three loss functions: and found that training LSTMs with MAE loss improved results (RMSE 24.586) over MSE loss (RMSE 30.097) with ϵ-insensitive loss performing the best with an RMSE of 23.483.

4 Follow-up Experiments and Results

4.1 Post Challenge Submission

We tuned the hyperparameters of our best model, XGBoost, and added the following features to the expanded featureset: first difference of CGM glucose; time since last bolus, meal, hypo event, and hypo correction; size of last bolus; carbs in last meal; and lagged features, up to 2 hours. 8-fold cross-validation without shuffling was used on the full training set to optimize the number of boosting rounds.

Table 2: Results of XGBoost on ablated feature set combinations of Self Reported (S), Medtronic Insulin pump (P), Basis Peak band (B), and Continuous Glucose Monitor (G) features (Sec. 2.2).

Features	val RMSE	test RMSE
S	53.549	54.510
B	55.829	55.390
G	22.692	*19.597*
P	56.85	54.711
S+B	53.184	53.607
S+G	22.099	**19.322**
S+P	54.087	54.371
B+G	22.764	*19.842*
P+B	56.779	54.128
P+G	22.388	*19.470*
S+B+P	52.481	53.238
S+B+G	22.504	*19.484*
S+P+G	22.125	*19.418*
B+P+G	22.616	*19.577*
S+B+P+G	22.45	*19.573*

4.2 Feature Ablation Experiments

Given the diversity of features sourced from biological measurements, self-reported events, and non-invasive physiological signals, we conducted an ablation study to determine the relative performance of models on subsets of feature groups. This experiment was performed with our best XGBoost model. From our expanded feature set, we created the following feature subsets:

- Self-reported features (S): meals, finger-stick glucose, illness, stress, exercise, and work, together with missing-value indicator columns, and one-hot encodings for meal type (41 features)
- Basis Peak band features (B): heart rate, GSR, skin and air temperature, steps, and sleep (6 features)
- Pump features (P): basal and temporary basal infusion rates, bolus doses, together with missing-value indicator and one-hot encoding columns (10 features)
- CGM glucose feature (G): blood glucose level recorded via CGM sensor (1 feature)
- Time features: one-hot encodings for hour-of-day and day-of-week (31 features)

The XGBoost model was trained on a feature vector containing the current feature value as well as the previous 12 values, lagged at 5 minutes apart. In total, we investigated 15 combinations of the S, P, B, and G features (Table 2). Time features were included in all combinations.

Table 2 shows that:

- Prediction suffers greatly when glucose (G) is ablated.
- The best model *does not* include insulin (P) or band (B) features.

Table 3: Average feature ranking for XGBoost for top 25 features (with average rank of feature in brackets). lag$N$ indicates a feature value $(N \times 5)$ minutes in the past.

Feature importance *without* band features		Feature importance *with* band features	
(1.0)	glucose_level	(1.0)	glucose_level
(2.1)	glucose_level_lag12	(2.1)	glucose_level_lag12
(6.0)	glucose_level_lag6	(7.3)	glucose_level_lag5
(6.7)	glucose_level_lag5	(8.5)	glucose_level_lag6
(7.7)	glucose_level_lag4	(10.4)	glucose_level_lag11
(8.8)	glucose_level_lag3	(13.0)	glucose_level_lag10
(10.6)	glucose_level_lag1	(13.5)	glucose_level_lag4
(10.8)	glucose_level_lag11	(14.8)	glucose_level_lag7
(12.2)	glucose_level_lag7	(15.0)	glucose_level_lag8
(13.2)	glucose_level_lag9	(15.8)	glucose_level_lag1
(13.8)	glucose_level_lag10	(17.5)	glucose_level_lag2
(14.8)	glucose_level_lag2	(17.8)	glucose_level_lag9
(14.8)	glucose_level_lag8	(18.1)	glucose_level_lag3
(37.2)	meal_carbs_lag3	(27.1)	basis_air_temp
(40.2)	finger_stick_lag5	(27.4)	basis_skin_temp_lag12
(43.7)	finger_stick_lag7	(33.2)	basis_gsr_lag12
(44.3)	meal_carbs_lag2	(34.9)	basis_gsr
(49.1)	finger_stick_lag8	(37.1)	basis_heart_rate
(50.8)	meal_carbs	(38.9)	basis_heart_rate_lag12
(57.2)	basal	(48.2)	basis_heart_rate_lag4
(59.4)	meal_carbs_lag4	(51.3)	basis_heart_rate_lag2
(60.6)	finger_stick_lag6	(59.3)	basis_heart_rate_lag3
(64.2)	meal_carbs_lag1	(62.1)	basis_heart_rate_lag1
(66.4)	finger_stick_lag3	(62.8)	meal_carbs_lag2
(67.8)	finger_stick_lag4	(63.8)	meal_carbs_lag4

- The model trained with only band features (B) and glucose (G) performed the worst among the glucose cohort.
- In general, adding band features seems to reduce performance.

We used XGBoost's feature importance rank to observe which features contributed to the most decision splits within the trees of the model. More splits within the trees infer a higher importance in decision making. Table 3 shows the mean rank of a feature as determined by XGBoost's importance score. We observe that, on average, XGBoost's decisions are most influenced by: (i) current glucose, (ii) glucose one hour ago, and (iii–xii) other glucose values within the past hour. This remains true irrespective of inclusion of Basis Peak band features.

4.3 Data Imputation Revisited

In each of the patient's data, missing values were observed in CGM measurements. Initially, these data gaps were filled via linear interpolation for both the training and test sets. However, such an imputation across gaps is only valid in a training and batch prediction setting. In online prediction, new feature vectors stream into the model in real time. Thus, to reflect realistic online prediction, missing values should only be imputed using past data.

To this point, any time intervals with missing glucose values should be excluded from the test-RMSE metric. As such, we corrected our implementation of test-RMSE to include only data periods with available glucose. In Figure 2 (left), unfiltered RMSE is computed with the interpolated points (black over gray) included in the test set, whereas *corrected RMSE* is computed with the interpolated points excluded from the test set (no predictions when glucose values are missing). Under the linear interpolation scheme, mean RMSE improves from an unfiltered value of 18.540 to 16.214 mg/dL (Table 4). Note that other features may have values even when glucose is absent (Sec. 2.2).

We then ran our best model on three test-set imputation schemes, the latter two of which are compatible with *online* prediction: linear interpolation (*linear*), persisting the last valid value (*ffill*), and leaving gaps unchanged (*none*).

Figure 2 and Table 4 compare the RMSEs of these imputation schemes. The unfiltered RMSE column in Table 4 gives model performance on interpolated test glucose values, without using the corrected RMSE function. The corrected RMSE columns list performance for three imputation schemes, using the corrected RMSE function. Interpolation with corrected RMSE performs best, but only forward-filled and non-imputed schemes can be implemented in an online context. Figure 2 (left) illustrates incorrect interpolation of glucose values. Figure 2 (center) and (right) show the *ffill* and *none* schemes, both of which are compatible with online prediction.

As an interesting exercise, missing test-set glucose values were left unchanged, and XGBoost was allowed to make predictions on the remaining features in the absence of glucose (Figure 2, right panel, orange over gray). Predictions were more variable in the absence of glucose signal, but seem to recover within two hours of the end of the data gap.

5 Context, Related Work, and Discussion

In this work, our aim was to deepen our understanding of the predictive modeling of bio-signals, for which the Blood Glucose Prediction Challenge was ideally suited. We acknowledge that a rich body of research exists, which explores the prediction of glycemia in depth.

Researchers have previously implemented Support Vector Regression [Plis *et al.*, 2014; Bunescu *et al.*, 2013], neural networks [Pappada *et al.*, 2008], recurrent neural networks [Allam *et al.*, 2011; Mirshekarian *et al.*, 2017], as well as genetic algorithms [Hidalgo *et al.*, 2017]. Feature engineering approaches include using expectation maximization for missing data imputation [Tresp and Briegel, 1998], as well as physiologically modeling glucose response signals as features [Bunescu *et al.*, 2013; Zecchin *et al.*, 2012; Contreras *et al.*, 2017].

For this work, we applied conventional feature-engineering methods. In the future, we would like to explore the inclusion of features based on physiological models of bio-signals for prediction. As an extension to our ablation and feature importance study, we would also like to explore non-glucose signals in-depth for glucose level prediction.

6 Conclusion

Our main finding is that XGBoost remains competitive with previously reported ARIMA models [Bunescu *et al.*, 2013], which supports glucose-derived features as the strongest predictors of future glucose levels.

Table 4: Post-submission XGBoost scores on test-set glucose with unfiltered and NaN-masked (corrected) RMSE, for three interpolation schemes.

Interpolation: Patient-id	number of test points	*linear* unfiltered RMSE	number of test points	*linear* corrected RMSE	*ffill* corrected RMSE	*no interpolation* corrected RMSE
559	2985	16.598	2514	17.107	17.391	17.230
563	2884	18.509	2570	16.018	16.033	16.026
570	2972	15.401	2745	14.315	14.709	14.493
575	2758	21.217	2590	17.556	17.611	17.749
588	2875	18.964	2791	16.500	16.500	16.519
591	2949	20.552	2760	15.162	15.140	15.269
AVG		**18.540**		**16.110**	**16.231**	**16.214**

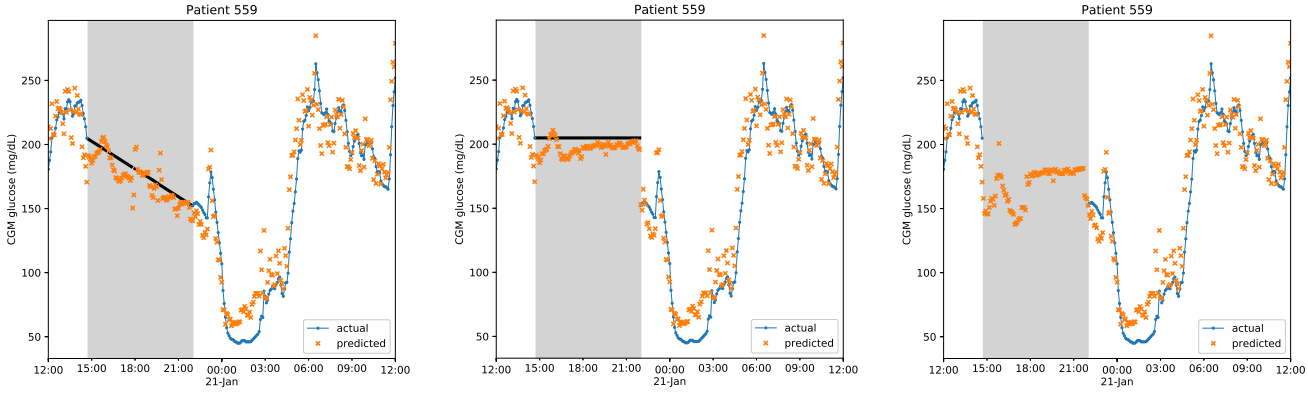

Figure 2: A representative data period in the test set of Patient 559, showing CGM glucose (blue), predicted glucose (orange), interpolated glucose (black), and spans of missing glucose (grey shaded regions). Three imputation schemes are compared: linear interpolation (left), forward-filling (center), and no imputation (right).

We observed that in LSTMs, ϵ-insensitive loss proved a more effective loss function than MAE, as inspired by the notion of incorporating an error tolerance corresponding to finger-stick measurement error. Interestingly, XGBoost models outperformed LSTMs in our study.

The collaboration of life sciences with the practice of data science offers the possibility of developing truly individualized proactive medicine. By personalizing such predictive models, we endeavour to further explore key signals—digital biomarkers, digital surrogate measurements—which reflect the strength of this interdisciplinary collaboration, and our ability to transform the future of healthcare.

Acknowledgments

We would like to thank Michael Li and the rest of the Klick-Labs team at Klick Inc. for their feedback, support, and encouragements.

References

[Allam *et al.*, 2011] Fayrouz Allam, Zaki Nossai, Hesham Gomma, Ibrahim Ibrahim, and Mona Abdelsalam. A recurrent neural network approach for predicting glucose concentration in type-1 diabetic patients. In *Engineering Applications of Neural Networks*, pages 254–259. Springer, 2011.

[American Diabetes Association, 2018] American Diabetes Association. Type 1 diabetes. `http://www.diabetes.org/diabetes-basics/type-1/`, 2018.

[Breiman, 2001] Leo Breiman. Random forests. *Machine Learning*, 45(1):5–32, Oct 2001.

[Bunescu *et al.*, 2013] Razvan Bunescu, Nigel Struble, Cindy Marling, Jay Shubrook, and Frank Schwartz. Blood glucose level prediction using physiological models and support vector regression. In *ICMLA, 2013*, volume 1, pages 135–140. IEEE, 2013.

[Che *et al.*, 2018] Zhengping Che, Sanjay Purushotham, Kyunghyun Cho, David Sontag, and Yan Liu. Recurrent neural networks for multivariate time series with missing values. *Scientific reports*, 8(1):6085, 2018.

[Chen and Guestrin, 2016a] Tianqi Chen and Carlos Guestrin. Xgboost: A scalable tree boosting system. *CoRR*, abs/1603.02754, 2016.

[Chen and Guestrin, 2016b] Tianqi Chen and Carlos Guestrin. Xgboost: A scalable tree boosting system. In *KDD*, KDD '16, pages 785–794, New York, NY, USA, 2016. ACM.

[Chollet, 2015] François et al Chollet. Keras. `https://keras.io`, 2015.

[Contreras *et al.*, 2017] Iván Contreras, Silvia Oviedo, Martina Vettoretti, Roberto Visentin, and Josep Vehí. Personalized blood glucose prediction: A hybrid approach using grammatical evolution and physiological models. *PloS one*, 12(11):e0187754, 2017.

[Food and Drug Administration, 2016] U.S. Food and Drug Administration. Self-monitoring blood glucose test systems for over-the-counter use. https://www.fda.gov/downloads/ucm380327.pdf, 2016.

[Graf *et al.*, 2017] Anneke Graf, Sybil A. McAuley, Catriona Sims, Johanna Ulloa, Alicia J. Jenkins, Gayane Voskanyan, and David N. O'Neal. Moving toward a unified platform for insulin delivery and sensing of inputs relevant to an artificial pancreas. *Journal of Diabetes Science and Technology*, 11(2):308–314, 2017. PMID: 28264192.

[Hidalgo *et al.*, 2017] J Ignacio Hidalgo, J Manuel Colmenar, Gabriel Kronberger, Stephan M Winkler, Oscar Garnica, and Juan Lanchares. Data based prediction of blood glucose concentrations using evolutionary methods. *Journal of medical systems*, 41(9):142, 2017.

[Juvenile Diabetes Research Foundation, 2018] Juvenile Diabetes Research Foundation. http://www.jdrf.org/research/artificial-pancreas/. http://www.jdrf.org/research/artificial-pancreas/, 2018.

[Marling and Bunescu, 2018] C. Marling and R. Bunescu. The OhioT1DM dataset for blood glucose level prediction. In *The 3rd International Workshop on Knowledge Discovery in Healthcare Data*, Stockholm, Sweden, July 2018. CEUR proceedings in press, available at http://smarthealth.cs.ohio.edu/bglp/OhioT1DM-dataset-paper.pdf.

[Medtronic, 2010] Medtronic. The basics of insulin pump therapy - medtronic diabetes. https://www.medtronicdiabetes.com/sites/default/files/library/download-library/workbooks/BasicsofInsulinPumpTherapy.pdf, 2010.

[Mirshekarian *et al.*, 2017] Sadegh Mirshekarian, Razvan Bunescu, Cindy Marling, and Frank Schwartz. Using lstms to learn physiological models of blood glucose behavior. In *EMBC, 2017*, pages 2887–2891. IEEE, 2017.

[Pappada *et al.*, 2008] Scott M Pappada, Brent D Cameron, and Paul M Rosman. Development of a neural network for prediction of glucose concentration in type 1 diabetes patients. *Journal of diabetes science and technology*, 2(5):792–801, 2008.

[Pedregosa *et al.*, 2011] 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. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

[Plis *et al.*, 2014] Kevin Plis, Razvan C Bunescu, Cindy Marling, Jay Shubrook, and Frank Schwartz. A machine learning approach to predicting blood glucose levels for diabetes management. In *AAAI Workshop: Modern Artificial Intelligence for Health Analytics*, number 31, pages 35–39, 2014.

[Tresp and Briegel, 1998] Volker Tresp and Thomas Briegel. A solution for missing data in recurrent neural networks with an application to blood glucose prediction. In *NIPS*, pages 971–977, 1998.

[World Health Organization, 2016] World Health Organization. *Global report on Diabetes*. World Health Organization, 2016.

[Zecchin *et al.*, 2012] C. Zecchin, A. Facchinetti, G. Sparacino, G. De Nicolao, and C. Cobelli. Neural network incorporating meal information improves accuracy of short-time prediction of glucose concentration. *IEEE Transactions on Biomedical Engineering*, 59(6):1550–1560, June 2012.