Predicting the Risk of Inpatient Hypoglycemia With Machine Learning Using Electronic Health Records

Yue Ruan,[1,2,3] Alexis Bellot,[4,5]
Zuzana Moysova,[6] Garry D. Tan,[1,2]
Alistair Lumb,[1,2] Jim Davies,[6]
Mihaela van der Schaar,[4,5] and
Rustam Rea[1,2]

OBJECTIVE

We analyzed data from inpatients with diabetes admitted to a large university hospital to predict the risk of hypoglycemia through the use of machine learning algorithms.

RESEARCH DESIGN AND METHODS

Four years of data were extracted from a hospital electronic health record system. This included laboratory and point-of-care blood glucose (BG) values to identify biochemical and clinically significant hypoglycemic episodes (BG ≤3.9 and ≤2.9 mmol/L, respectively). We used patient demographics, administered medications, vital signs, laboratory results, and procedures performed during the hospital stays to inform the model. Two iterations of the data set included the doses of insulin administered and the past history of inpatient hypoglycemia. Eighteen different prediction models were compared using the area under the receiver operating characteristic curve (AUROC) through a 10-fold cross validation.

RESULTS

We analyzed data obtained from 17,658 inpatients with diabetes who underwent 32,758 admissions between July 2014 and August 2018. The predictive factors from the logistic regression model included people undergoing procedures, weight, type of diabetes, oxygen saturation level, use of medications (insulin, sulfonylurea, and metformin), and albumin levels. The machine learning model with the best performance was the XGBoost model (AUROC 0.96). This outperformed the logistic regression model, which had an AUROC of 0.75 for the estimation of the risk of clinically significant hypoglycemia.

CONCLUSIONS

Advanced machine learning models are superior to logistic regression models in predicting the risk of hypoglycemia in inpatients with diabetes. Trials of such models should be conducted in real time to evaluate their utility to reduce inpatient hypoglycemia.

[1]Oxford Centre for Diabetes, Endocrinology and Metabolism, Oxford University Hospitals National Health Service Foundation Trust, Oxford, U.K.
[2]Oxford National Institute for Health Research Biomedical Research Centre, Oxford, U.K.
[3]Oxford Centre for Diabetes, Endocrinology and Metabolism, University of Oxford, Oxford, U.K.
[4]Department of Mathematics, University of Cambridge, Cambridge, U.K.
[5]Alan Turing Institute, London, U.K.
[6]Big Data Institute, University of Oxford Li Ka Shing Centre for Health Information and Discovery, University of Oxford, Oxford, U.K.

Corresponding author: Rustam Rea, rustam.rea@nhs.net

Hypoglycemia is a common and serious complication affecting people with diabetes (1). It is an inappropriately low blood glucose (BG) that results in significant morbidity in people with type 1 diabetes and in many people with type 2 diabetes (2). A BG level of ≤3.9 mmol/L is defined as level 1 hypoglycemia. A BG level of 2.9 mmol/L and lower is defined as level 2 hypoglycemia requiring immediate action, as at that level, neurogenic and neuroglycopenic symptoms begin to occur (3). Hypoglycemia can lead

to permanent neurological damage if not treated promptly and can ultimately be fatal (1).

Hypoglycemia is an important and common clinical problem under inpatient settings. Retrospective analysis of a U.S. electronic medical records database showed a 20% incidence of hypoglycemia and a 7% incidence of severe hypoglycemia (4). There was an associated 66% increase in adjusted inpatient mortality risk and greater than 50% increase in length of hospitalization stay. In a cross-sectional national audit of over 200 hospitals in the U.K., the 2017 National Diabetes Inpatient Audit showed that almost one in five people with diabetes experience hypoglycemia during their hospital stay. Although only 7% experience a severe (level 2) hypoglycemic episode, this rises to 26.9% of all patients with type 1 diabetes with 185 people over the course of 1 week requiring injectable rescue treatment for their hypoglycemia. Inpatient hypoglycemia has been implicated in the development of adverse clinical and economic outcomes, including increased mortality (5–7), adverse cardiovascular outcomes (8,9), and increased duration of hospital stay (6,10,11).

A recent review article has highlighted the urgent need for evidence-based methodologies to reduce inpatient hypoglycemia. Several strategies have already been developed to predict and prevent the occurrence of inpatient hypoglycemia (12). One approach to reducing inpatient hypoglycemia is to retrospectively analyze historical clinical data and develop a prediction tool to determine the individualized risk of hypoglycemia during an inpatient admission. With such a prediction tool, prevention measures can be tested in inpatients with high hypoglycemia risk. The possibility of developing such a prediction tool lies in the growing availability of rich clinical data sets stored in a hospital's electronic patient records (EPR) system.

Previous studies have used clinical information from local healthcare systems to develop inpatient hypoglycemia risk prediction tools (13–15). In one study, the prediction model developed by the researchers was tested in a clinical trial. This demonstrated the feasibility of using such a model to reduce severe hypoglycemia (glucose <40 mg/dL or <2.2 mmol/L) in inpatients with diabetes (14).

However, previous studies have only applied multilinear or logistic regression models on these data sets resulting in only a modest predictive capability of the models. Over the last few years, a number of advanced machine learning techniques have been developed within the field of biomedical engineering (16). These can be used to create predictive models which can be tested and compared with traditional logistic regression models in order to determine the model with the best predictive performance. This is the first study to assess the performance of novel machine learning models in predicting the risk of inpatient hypoglycemia.

RESEARCH DESIGN AND METHODS

We compared the performance of 18 different machine learning algorithms in predicting the risk of hypoglycemia in inpatients with diabetes.

Data Set

The study was approved by the Oxford University Hospitals National Health Service Foundation Trust Clinical Data Warehouse Program Board following completion of a Data Protection Impact Assessment. Data from Oxford University Hospitals National Health Service Foundation Trust was used which included the Cerner EPR system, the laboratory information management system and the point-of-care testing system. All the data used was collected for routine patient management with no additional data input required for the modeling. The data set contains hospital admission data from 1 September 2014 to 30 June 2018 for qualified patients with diabetes. The qualified patients are defined as meeting the following criteria: *1*) being an inpatient as coded in the EPR; *2*) having one diagnosis code among E10 (type 1 diabetes), E11 (non–type 1 diabetes), E13 (other specified diabetes), E14 (unspecified diabetes), or O24 (diabetes in pregnancy) as defined in the World Health Organization ICD-10 (17); and *3*) having at least one BG test performed during the hospital admissions. Hospital admission data for qualified patients, including patient demographics, procedures undertaken, diagnosis, laboratory tests, medication administration details, and vital signs, were extracted from different source data systems and pooled into a final data table for use by the machine learning prediction models.

A schematic representation of the data flow from the source data systems to the final data set used in the current study is depicted in Supplementary Fig. 1.

Predictors and Outcome

The outcome of interest in the current study is the risk of inpatient hypoglycemia during a hospital admission. We prepared two binary outcome variables $Hypo^{<4.0}$ and $Hypo^{<3.0}$ for each hospital admission representing two different severities of the potential hypoglycemic episodes since the degree of hypoglycemia may be influenced by different clinical predictors. We put value 1 to $Hypo^{<4.0}$ for any level 1 hypoglycemic episode (any BG measurement <4 mmol/L) and value 1 to $Hypo^{<3.0}$ for any level 2 hypoglycemic episode (any BG measurement <3 mmol/L) detected during the hospital admission, and value 0 to the two variables if no hypoglycemic episode detected (all BG measurements >4 mmol/L).

We preprocessed the integrated data set from the EPR and prepared 42 candidate predictors of interest based on clinical knowledge and previous studies. The predictors cover patient demographics (age, sex, and so on), procedures (value 1 for at least one procedure undertaken), laboratory test results (sodium, potassium levels, and so on), medication administration details (names and doses of medication delivered including different types of insulin), and vital signs (temperature, heart rate, and so on). Additional variables were added to the data set to improve the performance of the prediction algorithm. These included an episode of hypoglycemia in a previous admission within 6 months. Table 1 shows the full list of predictors and how they were represented in the source data systems and in the prediction models with units of the predictors provided.

Prediction Models

We evaluated the prediction performance of 18 different machine learning models on the data set. The models were used to predict the risk of hypoglycemia (either BG <4.0 mmol/L [$Hypo^{<4.0}$] or <3.0 mmol/L [$Hypo^{<3.0}$]). A total of 42 different variables were used as inputs into the prediction model. The models cover a wide range of commonly used classification algorithms including logistic regression, random forests, and artificial neural networks that have been previously

Table 1—Potential predictors and how they are represented in the EPR and in the prediction models

Category	Predictor	Data in EPR	Data in models	Unit	Completeness (%)	Data sets
Demographics	Age	Year of birth	Computed based on the year of admission	years	100	IH
	Sex	Male/female	Binary variable (1/0)	NA	100	IH
	Ethnicity	Ethnicity (categorical value)	Categorical variable (white British, African, etc.)	NA	100	IH
	Weight	Weight measured at time of admission	Weight value	kg	71	IH
	Height	Height measured at time of admission	Height value	cm	59	IH
	Type of diabetes	Type of diabetes (categorical value)	Categorical variable (T1D/T2D/other)	NA	100	IH
Vital signs	Diastolic blood pressure	Multiple measurements	Average value throughout the admission	mmHg	73	IH
	Systolic blood pressure	Multiple measurements	Average value throughout the admission	mmHg	73	IH
	Heart rate	Multiple measurements	Average value throughout the admission	/min	71	IH
	Oxygen saturation	Multiple measurements	Average value throughout the admission	%	73	IH
	Temperature	Multiple measurements	Average value throughout the admission	Celsius	72	IH
Laboratory tests	Albumin	Multiple measurements	Average value throughout the admission	g/L	81	IH
	Amylase	Multiple measurements	Average value throughout the admission	IU/L	15	IH
	C-peptide	Multiple measurements	Average value throughout the admission	pmol/L	17	IH
	Cortisol	Multiple measurements	Average value throughout the admission	nmol/L	26	IH
	Creatinine	Multiple measurements	Average value throughout the admission	µmol/L	80	IH
	C-reactive protein	Multiple measurements	Average value throughout the admission	mg/L	78	IH
	eGFR	Multiple measurements	Average value throughout the admission	mL/min/ 1.73 m^2	80	IH
	Hemoglobin	Multiple measurements	Average value throughout the admission	g/L	80	IH
	HbA$_{1c}$	Multiple measurements	Average value throughout the admission	%	42	IH
	Potassium	Multiple measurements	Average value throughout the admission	mmol/L	80	IH
	Sodium	Multiple measurements	Average value throughout the admission	mmol/L	80	IH
	White cells	Multiple measurements	Average value throughout the admission	× 10^9/L	79	IH
Medications	Sulfonylurea	Drug dose and time	Binary variable (1 for on drug and 0 for not)	NA	100	IH
	DPP-4	Drug dose and time	Binary variable (1 for on drug and 0 for not)	NA	100	IH
	GLP-1	Drug dose and time	Binary variable (1 for on drug and 0 for not)	NA	100	IH
	Metformin	Drug dose and time	Binary variable (1 for on drug and 0 for not)	NA	100	IH
	Morphine	Drug dose and time	Binary variable (1 for on drug and 0 for not)	NA	100	IH
	Pioglitazone	Drug dose and time	Binary variable (1 for on drug and 0 for not)	NA	100	IH
	Bisoprolol	Drug dose and time	Binary variable (1 for on drug and 0 for not)	NA	100	IH
	Amitriptyline	Drug dose and time	Binary variable (1 for on drug and 0 for not)	NA	100	IH
	Pregabalin	Drug dose and time	Binary variable (1 for on drug and 0 for not)	NA	100	IH
	Dexamethasone	Drug dose and time	Binary variable (1 for on drug and 0 for not)	NA	100	IH
	Prednisolone	Drug dose and time	Binary variable (1 for on drug and 0 for not)	NA	100	IH
	Intravenous insulin	Multiple rates of insulin infusion	Binary variable (1 for on i.v. insulin and 0 for not)	NA	100	IH+
	Insulin (rapid-acting analog)	Multiple doses of different amount	Average total daily insulin dose	unit	100	IH+
	Insulin (mixed analog)	Multiple doses of different amount	Average total daily insulin dose	unit	100	IH+
	Insulin (long-acting analog)	Multiple doses of different amount	Average total daily insulin dose	unit	100	IH+
	Insulin (short-acting human)	Multiple doses of different amount	Average total daily insulin dose	unit	100	IH+
	Insulin (mixed human)	Multiple doses of different amount	Average total daily insulin dose	unit	100	IH+
	Insulin (intermediate-acting human)	Multiple doses of different amount	Average total daily insulin dose	unit	100	IH+
Procedures	Procedure indication	Procedure name and time	Binary variable (1 for had at least one procedure during the admission and 0 for not)	NA	100	IH+
Previous hypoglycemia	Previous biochemical hypoglycemia	Blood glucose measurements	Binary variable (1 for had at least one blood glucose <4 mmol/L)	NA	63	PH
	Previous clinically significant hypoglycemia	Blood glucose measurements	Binary variable (1 for had at least one blood glucose <3 mmol/L)	NA	63	PH

DPP-4, dipeptidyl peptidase 4; GLP-1, glucagon-like peptide 1; NA, not applicable; T1D, type 1 diabetes; T2D, type 2 diabetes. Insulin (rapid-acting analog): "Insulin aspart," "Insulin lispro," "Insulin glulisine," "Insulin faster acting aspart." Insulin (mixed analog): "Insulin aspart biphasic (Novomix 30)," "Insulin lispro biphasic (Humalog Mix 25 and Humalog Mix 50)." Insulin (long-acting analog): "Insulin glargine," "Insulin detemir," "Insulin degludec." Insulin (short-acting human): "Insulin Actrapid," "Insulin Humulin S." Insulin (mixed human): "Insulin Humulin M3." Insulin (intermediate-acting human): "Insulin Insulatard," "Insulin Humulin I."

demonstrated to be robust and applicable to big data sets.

Model Validation and Comparison
For internal model validation, we used a 10-fold cross-validation. We randomly selected nine-tenths of the data set to be the training data set (developing the model) and the remaining one-tenth to be used to validate the model.

The outcome variables indicate whether or not a hypoglycemic episode occurred during an admission. The model was constructed to predict the probability of at least one hypoglycemic episode occurring. We measured the area under the receiver operating characteristic curve (AUROC), which shows the probability that the model correctly ranks the risk of hypoglycemia) higher no hypoglycemia. The AUROC is not sufficient on its own to use as a hypoglycemia prognostic model as it does not take into account the prevalence of hypoglycemia in the population. It assumes that positive and negative predictions are equally important. A detailed technical analysis of the shortcomings of the AUROC was recently conducted by Saito et al. (18). To ensure a comprehensive assessment of predictive performance, we used additional metrics. We used the terminology TP to represent the number of true positive predictions; FP to represent the number of false positives; TN to represent the number of true negatives; and FN to represent the number of false negatives. We defined precision (positive predictive value) as the ratio TP/(TP + FP). This is a

measure of the ability of the model to correctly predict a patient as having hypoglycemia. We defined recall (sensitivity) as the ratio TP/(TP + FN). This is a measure of the ability of the model to label as hypoglycemic all of patients who did indeed develop hypoglycemia. The precision and the recall were calculated for each machine learning model (Table 3).

Model Development

Following the development of the initial model (inpatient hypo [IH]), two further iterations of the data set were carried out. In the first change to the data set, the dose of insulin was added and the variable of intravenous insulin was indicated separately (IH+). In the second change, any previous admission in the last months containing a low glucose (<3 and <4 mmol/L) were added (previous hypo [PH]). The 18 different machine learning models were then rerun on each of these new data sets, and the model outputs were compared with the original data set. The last column in Table 1 shows which variable is used in each data set.

Variable Ranking

We sought to understand how the different variables contributed to the predictions by the XGBoost model (the best predictive model among the 18 tested models). We evaluated the predictive power of each individual variable by providing XGBoost with one variable at a time and assessing the diagnostic accuracy of the model that it constructed using only that variable. We evaluated the AUROC (using 10-fold cross-validation) to get a full picture of each variable's predictive power.

Statistical Analysis

All statistical analyses were performed using Python 3.6 and R version 3.3. All algorithms were implemented with the machine learning library "sklearn" (19) that contains all the algorithms and data science utilities used in this report. Internal validation was obtained via 10-fold stratified cross-validation. Performance comparisons were made with t tests from which we specified P values <0.001 to be considered statistically significant.

RESULTS

Baseline Characteristics

We analyzed data obtained from 17,658 inpatients with diabetes (9,277 males,

mean [SD], age 66 [18] years) who underwent 32,758 hospital admissions between July 2014 and August 2018. We identified all the biochemical (level 1) and clinically significant (level 2) hypoglycemic episodes during these admissions. The incidence of biochemical hypoglycemia during a hospital admission was 21.5% and that of clinically significant hypoglycemia was 9.6%. This is in keeping with data from the National Diabetes Inpatient Audit (20).

A selection of the baseline characteristics of the inpatient cohort and the glycemic outcomes are reported in Table 2.

Model Performance

The performance metrics of the machine learning models tested on the PH data set are presented in Table 3. The AUROC varied between 0.62 and 0.96 for different machine learning models. The estimation performance was better when predicting the risk of Hypo$^{<3.0}$ compared with predicting that of Hypo$^{<4.0}$. The AUROC for the logistic regression model was acceptable with 0.73 and 0.75 for estimation of the risk of Hypo$^{<4.0}$ and Hypo$^{<3.0}$, respectively. However, the best performing model for predicting the risk of Hypo$^{<4.0}$ and Hypo$^{<3.0}$ was the XGBoost model, which had the highest AUROC (0.96 for both), the highest precision value (0.88), as well as a high recall value (0.70) among all the models. Figure 1 shows the ROC curves contrasting the logistic regression, gradient boosting, decision tree, and XGBoost models. Supplementary Table 4 shows the normalized confusion matrix for these four models with the true positive and true negative rates on the upper left and lower right in the matrices, respectively. The XGBoost model was again the best performing model with a true positive rate of 0.98 and a true negative rate of 0.71 (Supplementary Table 4).

Logistic Regression Model

Estimated regression coefficients with SEs and P values from the logistic regression model with the PH data set are presented in Supplementary Table 3. The variables that are significant predictors of hypoglycemia are shown in Table 4. Similar predictors for Hypo$^{<4.0}$ and Hypo$^{<3.0}$ were found. Significant predictors for both Hypo$^{<4.0}$ and Hypo$^{<3.0}$ included weight, type of diabetes, oxygen saturation,

albumin level, sulfonylurea use, metformin use, intravenous insulin titration, long-acting human insulin use, procedures undertaken, and previous hypoglycemic episode.

Model Development

Two iterations of the original data set were carried out during the development of the prediction model. Supplementary Table 1 shows that the addition of a variable to discriminate patients on intravenous insulin compared with those on subcutaneous insulin and a variable for the dose of subcutaneous insulin (IH + data set) increased the best performing model (XGBoost) AUROC by 3 percentage points for Hypo$^{<4.0}$. When the PH data set was used, all models showed higher AUROC values while the XGBoost and gradient boosting models stood out with a significant increase of AUROC by 15 percentage points.

Variable Ranking

Supplementary Table 2 demonstrates the relative importance of the variables with the top three most important variables being previous hypoglycemic episodes, albumin levels, and type 2 diabetes. A number of novel predictive variables were identified from the machine learning method. These include several vital signs and medications that have logical clinical rationale underlying their importance to hypoglycemia. Further studies will be required to confirm their importance in the development of hypoglycemia.

CONCLUSIONS

To our knowledge, this is the first study comparing the performance of advanced machine learning models in predicting the risk of inpatient hypoglycemia. With the rich inpatient data set collected from the EPR system in a large university hospital, the 18 machine learning models showed high predicting power with an average AUROC at 0.85 for the detection of clinically significant hypoglycemia. The model with the highest AUROC (0.96) was the XGBoost model, which outperformed the logistic regression model. This model performed equally well in predicting both level 1 hypoglycemia (BG <4 mmol/L) and level 2 hypoglycemia (BG <3 mmol/L).

Among the 18 evaluated machine learning models, most machine learning

Table 2—Baseline characteristics and glycemic outcomes of the inpatients cohorts

	Inpatients with diabetes (N = 17,658)
Predictors	Inpatient hospital admissions (n = 32,758)
Sex, N (%)	
Female	8,381 (47)
Male	9,277 (53)
Age, mean (SD)	66 (18)
Ethnicity, N (%)	
White British	12,511 (70.8)
African	116 (0.7)
Pakistani	331 (1.9)
Chinese	53 (0.3)
Indian	254 (1.4)
Not stated	2,869 (16.2)
Other	1,524 (8.6)
Type of diabetes, N (%)	
Insulin-dependent diabetes	1,696 (9.6)
Non–insulin-dependent diabetes	14,006 (79.3)
Other forms	1,956 (11.1)
Systolic blood pressure, mean (SD)	132.5 (18.2)
eGFR, mean (SD)	29.8 (6.4)
Medication use	
Sulfonylurea, n (%)	6,435 (19.6)
DPP-4, n (%)	1,415 (4.3)
GLP-1, n (%)	349 (1.1)
Metformin, n (%)	10,756 (32.8)
Insulin, n (%)	
Intravenous insulin	4,678 (14.3)
Rapid-acting analog	3,954 (12.1)
Mixed-acting analog	1,553 (4.7)
Long-acting analog	5,118 (15.6)
Short-acting human	3,561 (10.9)
Mixed-acting human	1,388 (4.2)
Intermediate-acting human	2,394 (7.3)
Procedures, n (%)	22,931 (70.0)
Glycemic outcomes	
Hypoglycemia, n (%)	
Biochemical hypoglycemia	7,030 (21.5)
Clinically significant hypoglycemia	3,154 (9.6)
BG level, mean (SD)	10.1 (4.7)

N (%), number of patients and percentage over the total number of patients; n (%), number of admissions and percentage over the total number of admissions. DPP-4, dipeptidyl peptidase 4; GLP-1, glucagon-like peptide 1. Insulin (rapid-acting analog): "Insulin aspart, " "Insulin lispro, " "Insulin glulisine, " "Insulin faster acting aspart. " Insulin (mixed analog): "Insulin aspart biphasic (Novomix 30), " "Insulin lispro biphasic (Humalog Mix 25 and Humalog Mix 50). " Insulin (long-acting analog): "Insulin glargine, " "Insulin detemir, " "Insulin degludec. " Insulin (short-acting human): "Insulin Actrapid, " "Insulin Humulin S. " Insulin (mixed human): "Insulin Humulin M3. " Insulin (intermediate-acting human): "Insulin Insulatard, " "Insulin Humulin I. "

prognostic models performed markedly better than the predictions of the linear regression model. XGBoost is a highly flexible nonparametric model that integrates a large number of other machine learning models (decision trees). It was consistently the best performer with the highest AUROC, the highest precision, and good recall. It is important to note that XGBoost performed significantly better than the linear regression model, which has been used to predict inpatient hypoglycemia. There was an improvement of over 20 percentage points in terms of AUROC for prediction of both biochemical and clinically significant hypoglycemia. The high predictive capability of the XGBoost model also came with a high precision and recall showing low levels of overestimation of risk and low levels of missed events.

With the stepwise iteration of the data set, the predictive ability of the XGBoost model and some other models improved significantly. This was in contrast to the logistic regression model, which showed no significant improvement. This emphasizes the importance of developing a clinically relevant and comprehensive data set on which to base any machine learning in order to optimize the capability of the learning algorithm.

Our data set covers a wide range of potential predictors of inpatient hypoglycemia. The logistic regression model detected a number of significant predictors for clinically significant hypoglycemia including weight, type of diabetes, diastolic blood pressure, oxygen saturation, temperature, albumin levels, sulfonylurea use, metformin use, intravenous insulin use, high dose of long-acting human insulin, and people undertaking procedures (see Supplementary Table 3). Previous studies also found similar predictors, such as the albumin levels and glucose-lowering drugs (13,15,21). In our data set, undertaking any kind of procedure was also found to be a significant predictor. This is clinically understandable as procedures may disrupt the daily hospital routine of food intake and drug administration and, thereby, cause increased variability in glycemic levels.

Our machine learning models outperform other inpatient hypoglycemia prediction models that have been published using logistic or multivariate regression techniques. These have shown a discrimination of between 0.70 and 0.80 (15,21). Our model development also includes a variable for previous hypoglycemia, which has not previously been included in other prediction models.

Machine learning models have been widely used within hospital information systems to predict the risk of emergency admission, sepsis in the intensive care unit, and identifying type 2 diabetes using electronic health records (22–24). The performance of the current study also compares favorably to these other predictive models that have an AUROC of between 0.75 and 0.85.

There are several key strengths of the current study. First, we evaluated a wide range of machine learning models and compared their predicting ability against the most commonly used statistical model, which we used as a benchmark model. Second, we used an iterative approach to develop the model with additional variables that revealed how significant improvements to the model could be achieved. Third, this was the largest data set used to predict inpatient hypoglycemia containing the data for

Table 3—Performance metrics of the machine learning models based on the PH data set

Machine learning algorithm	Biochemical hypoglycemia (BG <4 mmol/L)			Clinically significant hypoglycemia (BG <3 mmol/L)		
	AUROC	Precision	Recall	AUROC	Precision	Recall
Logistic regression	0.73	0.48	0.10	0.75	0.39	0.10
SGD	0.74	0.12	0.10	0.77	0.10	0.10
kNN	0.62	0.40	0.18	0.62	0.30	0.15
Decision tree	0.81	0.70	0.71	0.84	0.68	0.73
Gaussian-naive Bayes	0.81	0.47	0.68	0.86	0.33	0.81
Bernoulli-naive Bayes	0.82	0.60	0.60	0.86	0.47	0.67
Multinomial-naive Bayes	0.75	0.10	0.10	0.79	0.10	0.10
SVM	0.79	0.73	0.10	0.83	0.41	0.10
QDA	0.77	0.23	0.96	0.89	0.15	0.97
Random forest	0.94	0.86	0.67	0.93	0.96	0.66
Extra trees	0.93	0.85	0.68	0.93	0.94	0.66
LDA	0.88	0.69	0.75	0.90	0.72	0.72
Passive aggressive	0.76	0.46	0.25	0.77	0.33	0.10
AdaBoost	0.89	0.68	0.60	0.93	0.63	0.46
Bagging	0.93	0.84	0.70	0.92	0.93	0.67
Gradient boosting	0.96	0.87	0.70	0.96	0.96	0.67
XGBoost	0.96	0.88	0.70	0.96	0.97	0.67
MLP	0.74	0.57	0.17	0.78	0.47	0.14
Mean (SD)	0.82 (0.10)	0.59 (0.25)	0.49 (0.29)	0.85 (0.10)	0.55 (0.31)	0.48 (0.31)

kNN, k-nearest neighbor; LDA, linear discriminant analysis; MLP, multilayer perceptron (artificial neural network); QDA, quadratic discriminant analysis; SGD, stochastic gradient descent; SVM, support vector machine.

32,758 hospital admissions. It also integrated clinical information that previous studies have not considered before, such as the medication dosage information and hypoglycemia in previous admissions.

However, as with all modeling efforts, there are also limitations. One limitation with the data set is the unavailability of carbohydrate intake/meal content information from the EPR during the hospital admissions. Carbohydrate intake has a direct impact on the postprandial glucose excursions and, consequently, the prandial insulin doses titrated to individuals and could be an important predictor for hypoglycemic events. This information is unavailable because it is not routinely recorded electronically in hospitals. Second, the current electronic health record does not record the level of hypoglycemia unawareness, prior continuous glucose monitoring (CGM) data, or prior self-monitoring of BG. These data are likely to be an important factor in developing hypoglycemia while in hospital, although for acutely unwell patients, these data may not be directly applicable. Third, our data set is derived from a single organization, and the generalizability of our best performing machine learning model needs to be tested in other data sets and evaluated in other centers. Finally, although we have developed a highly predictive model for inpatient hypoglycemia, the feasibility of using this model needs to be tested within a live EPR to confirm the ability of the model to receive data in real time and ensure that the model performs as strongly as the current data suggests.

One of the advantages of the current prediction model is that it uses variables that are readily accessible within the EPR. As a result, the model can be integrated into a decision support system under the EPR framework. In practice, the decision support system would access the clinical information of a new inpatient and feed the required information to the prediction model, which would then calculate the risk of the patient experiencing either biochemical or clinically significant hypoglycemia during his or her hospital admission. This would enable the decision support system to suggest appropriate treatment options based on individual risk levels, thereby reducing hypoglycemia and its consequent associated morbidity and potentially reduce the economic burden of prolonged hospital stay due to
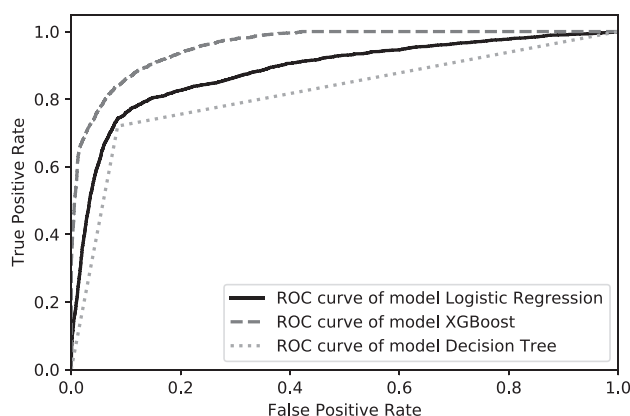
Figure 1—ROC curves for logistic regression, XGBoost, and decision tree model when predicting biochemical hypoglycemia.

Table 4—Most significant predictors from the logistic regression model

Predictors	Biochemical hypoglycemia (BG <4 mmol/L)			Clinically significant hypoglycemia (BG <3 mmol/L)		
	Coefficient	P value	z score	Coefficient	P value	z score
PrevLowGlucose3 (+)	3.842	<0.001	28.42	4.021	<0.001	20.39
Albumin level (−)	−0.078	<0.001	−27.22	−0.074	<0.001	−19.51
Intravenous insulin (+)	0.639	<0.001	15.43	0.501	<0.001	9.82
Procedure indication (+)	0.485	<0.001	14.87	0.339	<0.001	6.81
Sulfonylurea (+)	0.572	<0.001	14.24	0.311	<0.001	5.35
Type 2 diabetes (−)	−0.820	<0.001	−13.68	−0.656	<0.001	−7.88
Weight (−)	−0.010	<0.001	−7.42	−0.012	<0.001	−6.38
Oxygen saturation (+)	0.059	<0.001	6.31	0.067	<0.001	5.30
Metformin (−)	−0.212	<0.001	−6.02	−0.258	<0.001	−5.02
Long-acting human insulin (+)	0.011	<0.00	4.77	0.010	<0.001	5.35
Rapid-acting human insulin (+)	0.023	<0.001	4.11		NS	
Mixed insulin analog (+)	0.007	<0.001	3.67		NS	

Factors with a positive coefficient value increase the risk of hypoglycemia, and factors with a negative coefficient value decrease the risk of hypoglycemia. The factors are listed in order of effect size on the logistic regression model, e.g., an increase in albumin value reduces the risk of hypoglycemia, people with type 2 diabetes have an decreased risk of hypoglycemia. A "+" or "−" sign is given to each of the factors to indicate the effect direction. NS, not significant.

hypoglycemia. In a previous single-center study, a linear regression prediction model with a sensitivity of 50% and specificity of 71% was used to detect patients at risk for hypoglycemia. Clinician education resulted in medication change in 40% of patients and a reduction of 68% in the rate of severe hypoglycemia in alerted high-risk patients versus nonalerted high-risk patients (14). The benefits of a significantly more powerful prediction model based on machine learning need to be evaluated in a large multicenter randomized controlled trial.

Another potential application of the prediction model is the selection of high-risk patients who may benefit from an advanced treatment option, such as CGM or closed-loop insulin delivery. The use of CGM in the inpatient setting was considered at a recent symposium where the trials in both the intensive care unit and non–intensive care unit settings were reviewed (25). While there was some evidence that CGM may reduce the rates of severe hypoglycemia, it was recognized that there were limited data on clinical outcomes and that such technology may be most suitable for "populations.. at high risk for glucose variability and hypoglycemia." Closed-loop insulin delivery, or artificial pancreas, is a novel treatment option for people with diabetes who require exogenous insulin administration (26). The system titrates insulin based on real-time glucose monitoring and a titration algorithm. Previous clinical studies have shown promising glycemic results of the closed-loop systems under inpatient settings (27,28). However, the system is costly and cannot be applied in every inpatient with diabetes. Prediction models, such as the one described in this study, could be used as a preselection tool to determine which patients would benefit the most from CGM or automated insulin delivery.

In conclusion, this study demonstrates for the first time, the utility of advanced machine learning models in predicting the risk of hypoglycemia for inpatients with diabetes. We have shown that these models are significantly better in predicting inpatient hypoglycemia than the traditional logistic regression model. However further trials are needed to determine if this prediction model provides a significant clinical advantage over traditional logistic regression analysis or more simple risk factor prediction models, e.g., insulin use alone. Such machine learning models need to be evaluated within a real-time clinical setting to demonstrate their ability to predict hypoglycemia following admission. The use of new technological methods, such as machine learning and artificial intelligence, are not a substitute for clinicians, but they should be used to enhance clinical judgement and support everyday decisions. Multicenter clinical trials are now needed to evaluate their utility within a clinical decision support system and reduce the burden of hypoglycemia in hospital.

References

1. Cryer PE, Davis SN, Shamoon H. Hypoglycemia in diabetes. Diabetes Care 2003;26:1902–1912
2. UK Hypoglycaemia Study Group. Risk of hypoglycaemia in types 1 and 2 diabetes: effects of treatment modalities and their duration. Diabetologia 2007;50:1140–1147
3. Agiostratidou G, Anhalt H, Ball D, et al. Standardizing clinically meaningful outcome measures beyond HbA1c for type 1 diabetes: a consensus report of the American Association of Clinical Endocrinologists, the American Association of Diabetes Educators, the American Diabetes Association, the Endocrine Society, JDRF International, The Leona M. and Harry B. Helmsley Charitable Trust, the Pediatric Endocrine Society, and the T1D Exchange. Diabetes Care 2017;40:1622–1630
4. Brodovicz KG, Mehta V, Zhang Q, et al. Association between hypoglycemia and inpatient mortality and length of hospital stay in hospitalized, insulin-treated patients. Curr Med Res Opin 2013;29:101–107
5. Borzi V, Fontanella A. The clinical impact of hypoglycemia in hospitalized patients. Ital J Med 2015;9:11–19
6. Gómez-Huelgas R, Guijarro-Merino R, Zapatero A, et al. The frequency and impact of hypoglycemia among hospitalized patients with diabetes: a population-based study. J Diabetes Complications 2015;29:1050–1055
7. Turchin A, Matheny ME, Shubina M, Scanlon JV, Greenwood B, Pendergrass ML. Hypoglycemia and clinical outcomes in patients with diabetes hospitalized in the general ward. Diabetes Care 2009;32:1153–1157
8. Akhavan P, Aghili R, Malek M, Ebrahim Valojerdi A, Khamseh ME. Hypoglycemia: adverse cardiovascular outcomes in non-critically ill people with type 2 diabetes. Arch Iran Med 2016;19:82–86
9. Carey M, Boucai L, Zonszein J. Impact of hypoglycemia in hospitalized patients. Curr Diab Rep 2013;13:107–113
10. Evans M, Wolden ML, Thorsted BL, McEwan PC, Jacobsen JL. Inpatient hypoglycaemia increases length of hospital stay and all-cause mortality risk. Diabet Med 2015;32:23
11. Nirantharakumar K, Marshall T, Kennedy A, Narendran P, Hemming K, Coleman JJ. Hypoglycaemia

is associated with increased length of stay and mortality in people with diabetes who are hospitalized. Diabet Med 2012;29:e445–e448

12. Ruan Y, Tan GD, Lumb A, Rea RD. Importance of inpatient hypoglycaemia: impact, prediction and prevention. Diabet Med 2019;36: 434–443

13. Stuart K, Adderley NJ, Marshall T, et al. Predicting inpatient hypoglycaemia in hospitalized patients with diabetes: a retrospective analysis of 9584 admissions with diabetes. Diabet Med 2017; 34:1385–1391

14. Kilpatrick CR, Elliott MB, Pratt E, et al. Prevention of inpatient hypoglycemia with a real-time informatics alert. J Hosp Med 2014;9: 621–626

15. Mathioudakis NN, Everett E, Routh S, et al. Development and validation of a prediction model for insulin-associated hypoglycemia in non-critically ill hospitalized adults. BMJ Open Diabetes Res Care 2018;6:e000499

16. Park C, Took CC, Seong JK. Machine learning in biomedical engineering. Biomed Eng Lett 2018;8:1–3

17. WHO. World Health Organization. ICD-10 version:2010 [Internet], 2010. Available from https://icd.who.int/browse10/2019/en. Accessed 20 March 2019

18. Saito T, Rehmsmiere M. The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. Plos One 2015;10:e0118432

19. Pedregosa F, Varoquaux G, Gramfort A, et al. Scikit-learn: machine learning in Python. J Mach Learn Res 2011;12:2825–2830

20. NHS Digital. National Diabetes Inpatient Audit England and Wales, 2017 [Internet], 2017. Available from https://files.digital.nhs.uk/pdf/s/7/nadia-17-rep.pdf

21. Elliott MB, Schafers SJ, McGill JB, Tobin GS. Prediction and prevention of treatment-related inpatient hypoglycemia. J Diabetes Sci Technol 2012;6:302–309

22. Rahimian F, Salimi-Khorshidi G, Payberah AH, et al. Predicting the risk of emergency admission with machine learning: development and validation using linked electronic health records. PLoS Med 2018;15:e1002695

23. Desautels T, Calvert J, Hoffman J, et al. Prediction of sepsis in the intensive care unit with minimal electronic Health record data: a machine learning approach. JMIR Med Inform 2016; 4:e28

24. Zheng T, Xie W, Xu L, et al. A machine learning-based framework to identify type 2 diabetes through electronic health records. Int J Med Inform 2017;97:120–127

25. Wallia A, Umpierrez GE, Rushakoff RJ, et al. Consensus statement on inpatient use of continuous glucose monitoring. J Diabetes Sci Technol 2017;11:1036–1044

26. Boughton CK, Hovorka R. Advances in artificial pancreas systems. Sci Transl Med 2019;11: eaaw4949

27. Bally L, Thabit H, Hartnell S, et al. Closed-loop insulin delivery for glycemic control in noncritical care. N Engl J Med 2018;379:547–556

28. Thabit H, Hartnell S, Allen JM, et al. Closed-loop insulin delivery in inpatients with type 2 diabetes: a randomised, parallel-group trial. Lancet Diabetes Endocrinol 2017;5:117–124