

ORIGINAL RESEARCH

Open Access



Real-time artificial intelligence assisted insulin dosage titration system for glucose control in type 2 diabetic patients: a proof of concept study

Ying Chen^{1†}, Zhiwei Chen^{2†}, Lin Zhao^{1†}, Simin Li³, Zhen Ying¹, Peng Yu¹, Hongmei Yan¹, Hong Chen¹, Chun Yang⁴, Jiyang Zhang², Qingnan Meng¹, Yuchen Liu¹, Ling Cao¹, Yanting Shen¹, Chunyan Hu¹, Huiqun Huang¹, Xiaomu Li¹, Hua Bian¹ and Xiaoying Li^{1*}

Abstract

Objective This study aims to develop an insulin dosage adjustment model using machine learning of high quality electronic health records (EHRs) notes and then to form an artificial intelligence-based insulin clinical decision support workflow (iNCDS) implemented in the HIS system to give a real-time recommendation of insulin dosage titration. The efficacy and safety in clinical practice is evaluated in this proof-of-concept study.

Research design and methods We extracted patient-specific and time-varying features from the original EHRs data and performed machine learning analysis through 5-fold cross validation. In the patient-blind, single-arm interventional study, insulin dosage was titrated according to iNCDS in type 2 diabetic inpatients for up to 7 d or until hospital discharge. The primary end point of the trial was the difference in glycemic control as measured by mean daily blood glucose concentration during the intervention period.

Results A total of 3275 type 2 diabetic patients with 38,406 insulin counts were included for the model analysis. The XGBoost model presented the best performance with root mean square error (RMSE) of 1.06 unit and mean absolute relative difference (MARD) of 6.0% in the training dataset, and RMSE of 1.30 unit and MARD of 6.9% in the testing dataset. Twenty-three patients with T2DM (male 14, 60.9%; age 58.8 ± 10.7 years; duration of diabetes 11.8 ± 8.8 years, HbA1c $9.1 \pm 1.1\%$) were enrolled in the proof of concept trial. The duration of iNCDS intervention was 7.0 ± 0.1 d. The insulin dose recommended by iNCDS was accepted by physicians in 97.8%. The mean daily capillary blood glucose was markedly improved during the intervention period, with a reduction of mean daily capillary BG from 11.3 (8.0, 13.9) mmol/L in the first 24 h to 7.9 (6.5, 8.9) mmol/L in the last 24 h of the trial ($P < 0.001$). In addition, the time range below 3.9 mmol/L was decreased from 1.1% to 0.5%.

Conclusions The clinical decision support system of insulin dosage titration developed using a machine learning algorithm based on the EHRs data was effective and safe in glycemic control in type 2 diabetic inpatients.

Trial registrations [ClinicalTrials.gov](https://clinicaltrials.gov/ct2/show/study/NCT04053959) Identifier: NCT04053959.

Keywords Machine learning, Insulin dose titration, Artificial intelligence assisted clinical decision support, Glycemic control, Type 2 diabetes

[†]Ying Chen, Zhiwei Chen and Lin Zhao contributed equally to this work.

*Correspondence:

Xiaoying Li

li.xiaoying@zs-hospital.sh.cn

Full list of author information is available at the end of the article

1 Introduction

Type 2 diabetes is a common chronic disease, which affects more than 425 million people worldwide (Cho et al., 2018). Consistent hyperglycemia can lead to serious micro- and macrovascular complications, including blindness, kidney failure, lower limb amputation, increased risk of cardiovascular disease (Stratton et al., 2000). While management of glucose levels in well-targeted range can markedly improve diabetic complications and reduce mortality in diabetic patients (Holman et al., 2008; Hayward et al., 2015). Insulin is the most effective treatment for glycemia control. It is estimated that a large number of type 2 diabetic patients eventually require insulin therapy for glycemia control in the course of disease progression (Basu et al., 2019). However, two-thirds of insulin users had a poor control of glycemia with glycated hemoglobin (HbA1c) levels above 7.0% (53 mmol/L) and more than a third with HbA1c levels higher than 9.0% (75 mmol/L) mainly due to inaccurate and inertia titration of insulin dosage (Nawaz et al., 2017).

Type 2 diabetes is characterized by a failure of the tissues to appropriately respond to insulin and/or insufficient insulin production by pancreatic β cells (DeFronzo et al., 2015). Insulin therapy in people with type 2 diabetes is very different from that used to treat people with type 1 diabetes. In addition to basal bolus therapy, the basal and biphasic/premix insulin therapies are commonly used in clinical practice for type 2 diabetic patients. Meanwhile, in patients with type 2 diabetes, the response to insulin was individually different and varied by the progression of diabetes (Petersen & Shulman, 2018). Inappropriate therapeutic decision regarding insulin dosing may lead to significant hyperglycemia or hypoglycemia, both of which have poor and even have a life-threatening consequences (Akirov et al., 2017). Therefore, personalized and frequent dosage titrations are indispensable to overcome constant variations in insulin requirements. To achieve this goal, accurate and real-time insulin dosage titration delivered at an individual level could be meaningful. Current dosing tools are mostly based on HbA1c levels and body weight, however, insulin response is globally variable and thus HbA1c and weight-based dosing is suboptimal. In current clinical practice, insulin dosage relies significantly on expert knowledge. Given the mounting shortage of experienced physicians in titrating insulin dosage in both inpatients and outpatient, the current paradigm of clinical care lacks the ability to provide the timely insulin adjustments needed to improve the management of type 2 diabetes.

Over the past few years, a number of advanced machine learning techniques have been developed with potential priority in medicine (Van Calster & Wynants,

2019). Meanwhile, modern electronic health record (EHR) systems can provide readily accessible data to fuel scientific study and prediction modeling. Machine learning approaches have emerged as potentially powerful tools to mine EHRs data to aid in disease diagnosis and management, mimicking and perhaps even augmenting the physicians in clinical decision-making (Liang et al., 2019; Tomasev et al., 2019). These promising work on EHRs derived modelling suggests that incorporation of machine learning may enable clinical decision of insulin dosage titration. Some artificial intelligence decision support system have developed based on deep learning methods for type 1 diabetes. The automated advice of the DreaMed Advisor Pro analyzed data from various devices (CGM, glucometer and insulin pump) and provides recommendations for insulin-pump dosing adjustment (Nimri et al., 2018). And randomized controlled trial have demonstrated that use of the automated decision support tool for optimizing insulin pump settings was non-inferior to intensive insulin titration provided by physicians from specialized academic diabetes centers (Nimri et al., 2020). *k*-nearest neighbours (KNN) was also reported to develop an algorithm that provides weekly insulin dosage recommendations to adults with T1D using multiple daily injection (MDI) (Tyler et al., 2020). However, Insulin therapy in type 2 diabetes is more complicated. There have been some attempts to utilize large medical datasets for predicting insulin dosage adjustment (Ruan et al., 2020); however, none has been implemented in clinical settings.

In the present study, we used machine learning methods of analyzing high quality EHRs notes to develop an insulin dosage adjustment model. Furthermore, it was implemented in the health information system (HIS) to form an artificial intelligence-based insulin clinical decision support workflow (iNCDSS), which could give a real time recommendation for insulin titration. And we further performed a proof-of-concept study to evaluate the efficacy and safety of iNCDSS system on glycemic control in type 2 diabetic patients who received insulin therapy.

2 Methods

2.1 Study design and participants

We built and validated a machine learning model based on the data retrospectively obtained from the EHRs of hospitalized type 2 diabetic patients who received subcutaneous insulin injection. Then, the model was implemented into the HIS system to form an insulin clinical decision support workflow, which could real-time give insulin titration recommendation. We conducted a prospective study to compare the performance of our iNCDSS model to that of human physicians.

Furthermore, we performed a proof-of-concept study to evaluate the efficacy and safety of the iNCDSS in clinical practice. The study was approved by the Ethical Committee of Zhongshan Hospital, Fudan University and conducted according to the Helsinki Declaration. Each participant provided a written informed consent.

3 Development and validation of the iNCDSS model

3.1 Data processing

In this study, EHRs of type 2 diabetes patients were collected from the wards of Department of Endocrinology and Metabolism, Zhongshan Hospital from January 2012 to September 2018. A total of 4037 patients receiving insulin for at least two consecutive days were included. After excluding the patients who used insulin pumps, intravenous injection, and glucocorticoids, a total of 3275 patients with 38,406 insulin counts were included for the model analysis.

3.2 Model development

Our goal is to predict the dose of insulin ordered for the specific patient at the specific timestamp. In this project, insulin was classified into three categories based on its onset time and length of action: long-acting insulin, short-acting insulin, and biphasic/premixed insulin. The long- (and intermediate-) acting category are usually injected before breakfast or bedtime, containing NPH, Detemir and Glargine. The short- (and rapid-) acting category is usually injected before meals, including insulin analogues, Lispro and Aspart. Biphasic/premixed insulin is a mixture of intermediate-acting and rapid-acting insulin, which is usually injected before meals (commonly before breakfast and dinner). Three typical insulin regimens are commonly used in clinic: (1) Basal insulin only: long-acting insulin once daily injected before breakfast or bedtime. (2) Biphasic or premixed insulin regimen: biphasic/premixed insulin twice daily injected before breakfast and dinner. (3) Basal bolus therapy: long-acting insulin once daily injected before breakfast or bedtime, combined with the short- (and rapid-) acting insulin daily injected before meals.

Therefore, the three categories of insulin are further generated into six predicted variables: the long- (and intermediate-) acting insulin injected before breakfast or bedtime, biphasic/premixed insulin injected before breakfast, biphasic/premixed insulin injected before dinner, the short- (and rapid-) acting insulin injected before breakfast, the short- (and rapid-) acting insulin injected before lunch and the short- (and rapid-) acting insulin injected before dinner.

In clinical practice, the current insulin dose is adjusted based on the previous insulin therapy. Here, we defined the insulin dose as below:

$$I_{k,t+1} = I_{k,t} + g_k(x_{t+1}), k \in \{b, p_6, p_{16}, r_6, r_{10}, r_{16}\},$$

Where k denotes the predicted variables; $I_{k,t}$ and $I_{k,t+1}$ denotes the dose of k insulin at the day t and the day $t+1$, respectively. g_k denotes the adjusted dose of k insulin at the day $t+1$. x_{t+1} indicates the features that could influence the insulin adjustment dose. The aim of this study is to train a machine learning model to predict the value of insulin adjustment dose g_k .

3.3 Feature extraction

The features were extracted from the original EHRs data. For predicting the specific timepoint adjusted dose of insulin, the data before the record of insulin order would be included for analysis. The features as candidates for analysis can be divided into two categories, patient-specific features and time-varying features. Patient-specific features were those specific for each patient, which mainly contained patients' demographic information, chief complaint, history of present illness, history of past illness, physical examination, laboratory testing, while time-varying features were those changed with time, which mainly consisted of doctor's order, and capillary glucose measurements (Supplementary Table 1). For each feature, we identified which category of EHR clinical data it was derived from and whether it was a binary classification or continuous variable or text format. Patient-specific features/variables with > 10% missing data were excluded.

Age, sex, physical examination, and laboratory testing items (except capillary glucose measurements) were extracted from the EHRs for analysis. And the latest values before the predicted timepoint were extracted as candidates for analysis when there were more than or equal to two values for the same item.

The text EHR notes (chief complaint, history of past illness and history of present illness) were manually annotated into the standardized lexicon and clinical features, allowing further processing of clinical information for analysis. Symptoms and durations of diabetes/hyperglycemia were extracted from the chief complaints section. The glucose control indicators, previous anti-hyperglycemic treatment regimen, and hypoglycemia occurrence were extracted from the history of present illness.

Given the unique and robust performance of the change of capillary blood glucose, insulin dose and anti-hyperglycemic drug, the characteristics for each patient at each timestamp were processed separately. The seven point blood glucose were commonly measured in the wards and the capillary blood glucose data were directly transmitted from the manufacturer's cloud to our hospital EHR. Therefore, we extracted seven blood glucose features a day. As mentioned above, the insulin could

be categorized into 6 features according to insulin type and injection timepoint, which could generate 6 insulin features a day. Anti-hyperglycemic agents were quantified by daily total dose. These features were calculated from the day of admission to the day of the predicted time. The values in past 5 d and the maximum/minimum/median/standard deviation values of each feature were included in the model analysis. Finally, 378 features were generated.

3.4 Model development and validation

We evaluated the prediction performance of commonly used machine learning models for regression tasks on datasets. The predicted dose of insulin at specific timepoint was used as output of the models. Variables mentioned above were used as inputs into the prediction models. The training models included XGBoost, support vector machine (SVM), neural network (NN), linear regression, lasso regression, and random forest, which have been demonstrated to be robust and applicable in big datasets (Ngiam & Khor, 2019). The results were compared with the same input data after the same pre-processing.

For internal model validation, we used a 5-fold cross validation. We randomly selected 80% of the dataset into the training dataset to develop the model and the remaining 20% into the testing dataset to validate the model. Missing values were left blank in the XGBoost model and filled with mean values for the rest of the models.

We used the metrics of root mean square error (RMSE), mean absolute relative difference (MARD) and Spearman's correlation to evaluate the performance of the models, which are basic evaluation indexes for regression tasks. RMSE represents the errors between predicted values and actual values. MARD is the average of relative absolute errors of predicted values. Spearman's correlation assesses rank similarity between predicted values and actual values.

The best algorithm was finally achieved and the identified optimized model was then applied in the testing dataset to assess its performance. Furthermore, the importance of each attribute was analyzed. If some attribute was assumed to be less important in the prediction, it would be dropped out in the final model.

3.5 Performance of iNCDSS and human physicians

Furthermore, we compared the performance of our finalized model with that of human physicians. We allocated 24 physicians into 3 groups according to the levels of proficiency and clinical practice experience. Group A ($n=8$) consists of intern physicians with 1–2 year clinical practice, Group B ($n=8$) consists of resident physicians with 3–5 year clinical practice, and Group C ($n=8$)

consists of attending physicians with 6–10 year clinical practice. Each patient record was randomly assigned to three physicians (one in each group). The physicians in each group manually prescribed insulin dosage based on the peripheral capillary blood glucose measurements and clinical information in a pre-designed online electronic questionnaire. The actual insulin dosage was prescribed by the attending physicians, which was taken as the reference. Differences in RMSE, MARD and accuracy score were employed to evaluate the performance of iNCDSS system and physicians in each group.

4 Efficacy and safety of iNCDSS system: proof of concept trial

4.1 Trial design and participants

In February 2019, the finalized model was implemented in the doctor's order interface of the HIS to build a workflow for a clinical decision support system of insulin titration in the hospital setting, which could real-time read updated patient information and provide insulin dosage recommendation (Fig. 1). The iNCDSS recommendations are created once the features of EHRs transmitted to the model. Capillary glucose measurements were conducted by the nurse staff in charge and automatically uploaded into the HIS system. The recommendation of insulin dosage was sent to the healthcare provider for review in the Doctor's order interface. An example of the iNCDSS recommendation report for the healthcare provider is presented in Supplementary Fig. 1. The healthcare provider can review the current recommendation and choose "accept" or "reject." In the case of "reject," the healthcare providers can document the reason and feedback it to the log file.

We conducted a proof of concept trial to evaluate the efficacy and safety of iNCDSS in hospitalized patients with previously diagnosed T2DM from October, 2019 to January, 2020. This study was a patient-blind, single-arm interventional study and performed in the general ward of the Department of Endocrinology and Metabolism, Zhongshan Hospital, China.

Type 2 diabetic patients receiving subcutaneous insulin injection were recruited and screened for the inclusion and exclusion criteria. Inclusion criteria were type 2 diabetes, age of 18–75 years, and HbA1c between 7.0% and 11.0%. The following subjects were excluded: subjects with acute complications of diabetes, such as ketoacidosis or hyperglycemic hyperosmolar state; BMI ≥ 45 kg/m²; women who are pregnant or breast-feeding; subjects with severe cardiac, hepatic, renal diseases; subjects with any psychiatric or psychological diseases; subjects with severe edema, infections or peripheral circulation disorders, receiving surgery during hospitalization; and subjects who could not comply with the protocol.

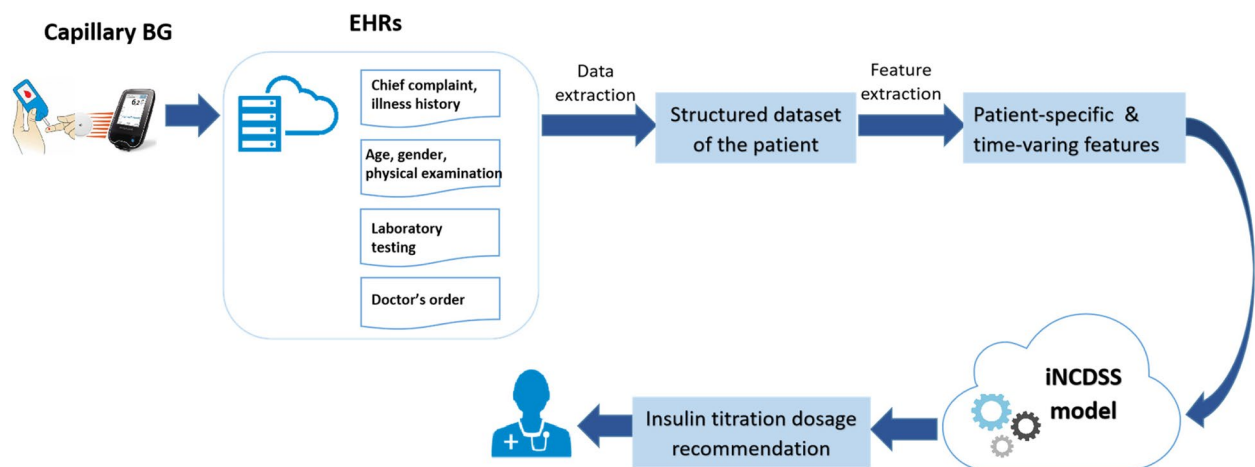


Fig. 1 Workflow of the iNCDSS system

4.2 Trial procedures

Body weight, height and total daily insulin dosage were recorded for each patient after enrollment. HbA1c level was measured on the first day of admission. The initial insulin dose was referred to the pre-study daily insulin dose. After the first cycle of insulin regimen, eligible patients received insulin dosage titration according to the iNCDSS system recommendation, which had to be confirmed twice daily by the physician in charge. The goal of insulin therapy was to achieve preprandial capillary blood glucose between 5.6–7.8 mmol/L and postprandial capillary glucose less than 10.0 mmol/L (Umpierrez et al., 2013). The treating physician could overrule the recommended insulin dosage if deemed necessary. All patients were studied for up to 7 d or until hospital discharge.

Throughout the trial, the patients were provided standard meals at usual meal time in the ward. Participants remained largely sedentary during the trial, and no physical activity was scheduled. Anti-diabetic medications were not changed during the period of the study.

4.3 Glucose monitoring

For each patient, capillary glucose concentration was measured at 7 time points of fasting, after breakfast, before and after lunch, before and after dinner, and before bedtime a day using Glucometer (Glupad, Sinomedisite, China). Capillary glucose measurements were performed by the nurse staff according to standard procedures with a point-of-care testing device, which is integrated into the HIS system. And continuous glucose monitoring (CGM) was performed using flash glucose monitoring (Abbott Freestyle Libre, USA) placed on the upper left arm. The CGM data were analyzed retrospectively, and the treatment was not influenced by data gained by CGM.

4.4 Efficacy and safety outcomes

The primary end point of the study was difference in glycemic control as measured by mean daily BG concentration. The secondary end points included glucose concentration in target range (TIR) of 3.9–10.0 mmol/L, glucose concentration above range (10.1–13.9 mmol/L or > 13.9 mmol/L) or below range (3.0–3.8 mmol/L or < 3.0 mmol/L), and glycemic variability (American, 2021). Glycemic variability was determined by the coefficient of variation (CV) of glucose values.

4.5 Adverse events and safety

Safety was evaluated as number of hypoglycemic events and the physician's adherence to iNCDSS recommendation was assessed. The reportable serious adverse events included severe hypoglycemia defined as a capillary glucose level of less than 2.2 mmol/L or an episode that required the assistance of another person and hyperglycemia (> 20 mmol/L) with ketonemia or hyperosmolar coma, as determined by capillary measurements, along with other adverse events and serious adverse events (American, 2021).

4.6 Statistical analysis

Machine learning was performed using Python 3.6. XGBoost was implemented with the open source package provided by the XGBoost community (Chen & Guestrin, 2016). Neural network algorithm was implemented using Keras API with Tensorflow as backend (Abadi & Agarwal, 2015). The remaining algorithms were implemented with the machine learning library "Scikit-learn" (Pedregosa et al., 2011). Internal validation was obtained via 5-fold stratified cross-validation. RMSE and MARD are used to measure the prediction performance of

insulin dose. The gain was used to evaluate the importance of features. The gain of the feature indicates the average gain across all splits the feature is used in. Since we used 5-fold cross validation, we calculated the mean gain for models.

Clinical studies were analyzed using SAS 9.3 and two side *P* value less than 0.05 was considered statistically significant. RMSE, MARD and accuracy score were used to evaluate the performance of iNCDSS system and physicians in each group. The matched *t* test was used to compare the performance of iNCDSS and physicians. The change from baseline measurements to the end of the study was analyzed by two-sided paired *t* test and a Wilcoxon signed-rank test for the continuous measurements.

The seven-point blood glucose profiles were analyzed using generalized linear mixed model.

5 Results

5.1 Development and validation of the iNCDSS model

A total of 3275 type 2 diabetic patients with 38,406 insulin dose counts were included in the model analysis. The demographics and clinical characteristics of patients are present in Supplementary Table 2. The mean age was 60.6 ± 12.5 years old and 1863 (56.9%) subjects were men.

Table 1 showed the performance metrics of the machine learning algorithms in the training and testing datasets, respectively. The XGBoost algorithm exhibited a stable performance in both training and testing sets,

Table 1 Performance of the machine learning algorithms in training sets and testing sets

Model	Training Sets				Testing Sets			
	Total	Biphasic or Premixed Insulin	Long-acting Insulin	Basal Bolus Therapy	Total	Biphasic or Premixed Insulin	Long-acting Insulin	Basal Bolus Therapy
Count	30,725	16,977	655	13,093	7681	4244	164	3273
XGBoost								
RMSE (Unit)	1.059	1.094	1.771	0.961	1.304	1.263	1.976	1.305
MARD (%)	0.060	0.056	0.079	0.064	0.069	0.062	0.090	0.078
Spearman's correlation coefficient	0.987	0.984	0.953	0.980	0.982	0.980	0.938	0.970
RandomForest								
RMSE (Unit)	0.492	0.481	0.808	0.485	1.324	1.286	2.010	1.320
MARD (%)	0.027	0.024	0.034	0.030	0.072	0.064	0.092	0.082
Spearman's correlation coefficient	0.996	0.996	0.987	0.993	0.981	0.979	0.937	0.968
SVM								
RMSE (Unit)	1.279	1.266	2.028	1.247	1.392	1.373	2.041	1.367
MARD (%)	0.057	0.054	0.072	0.061	0.068	0.062	0.080	0.074
Spearman's correlation coefficient	0.984	0.981	0.946	0.975	0.980	0.977	0.933	0.967
NN								
RMSE (Unit)	1.198	1.217	1.830	1.131	1.388	1.342	2.339	1.370
MARD (%)	0.069	0.062	0.084	0.077	0.079	0.069	0.110	0.091
Spearman's correlation coefficient	0.984	0.981	0.951	0.973	0.979	0.977	0.917	0.963
Linear regression								
RMSE (Unit)	1.467	1.485	2.113	1.403	1.469	1.485	2.053	1.401
MARD (%)	0.070	0.068	0.082	0.072	0.070	0.068	0.083	0.072
Spearman's correlation coefficient	0.977	0.974	0.940	0.962	0.977	0.974	0.936	0.961
Lasso								
RMSE (Unit)	1.499	1.514	2.171	1.436	1.499	1.513	2.114	1.433
MARD (%)	0.070	0.068	0.081	0.072	0.070	0.068	0.081	0.072
Spearman's correlation coefficient	0.975	0.971	0.931	0.961	0.975	0.971	0.928	0.960
Ridge								
RMSE (Unit)	1.322	1.315	2.018	1.286	1.347	1.331	1.980	1.319
MARD (%)	0.072	0.066	0.082	0.078	0.073	0.067	0.084	0.080
Spearman's correlation coefficient	0.981	0.978	0.941	0.970	0.980	0.978	0.934	0.968

RMSE root mean square error, MARD mean absolute relative difference

SVM support vector machine, NN neural network

and demonstrated the best performance in the testing sets, with RMSE of 1.30 unit and Spearman's correlation coefficient of 0.982 in the testing dataset. Meanwhile, the XGBoost model also reported relatively low standard deviations with MARD of 6.0% in the training dataset and 6.9% in the testing dataset, which is the best among all algorithms.

To gain insight into how the model generated an insulin dose prediction, we identified key clinical features driving the dose prediction. The relatively important features of the XGBoost model were shown in Supplementary Table 3. Following previous findings on the important features, we tried to construct a simplified and clinically operable decision model. XGBoost algorithms are based on recursive decision tree building from past residuals and can identify those trees that contribute the most to the decision of the predictive model. The number of key features was set to the most 10 important features. Scatter plots of predictive dosage of insulin by simplified XGBoost model against the actual insulin dosage were shown in Fig. 2. The results showed that the simplified XGBoost models well performed for RMSE and MARD.

5.2 The performance of iNCDSS compared with human physicians

We also compared the performance of our iNCDSS system with human physicians in type 2 diabetic inpatients. Human physicians were categorized into three groups according to the levels of proficiency and years of clinical practice. Our iNCDSS model achieved lower RMSE and MARD, and higher accuracy score than the intern and resident physicians, and similar with or slightly lower than the attending physicians (Supplementary Table 4). These results indicated that our iNCDSS system could be capable of potentially assisting junior physicians in titrating insulin dosage.

5.3 Efficacy and safety of iNCDSS system on glycemic control in type 2 diabetic inpatients

A total of 23 type 2 diabetic patients were included in the study. The baseline clinical characteristics of the participants were shown in the Supplementary Table 5. The combined antidiabetic medications included metformin, acarbose, and DPP-4 inhibitors. The run-in period after admission was 1.1 ± 0.2 d. The duration of intervention with iNCDSS system was 7.0 ± 0.1 d. The mean blood glucose level at baseline was 13.6 ± 3.0 mmol/L, and the mean baseline HbA1c was $9.1 \pm 1.1\%$. 97.8% of the total insulin recommendations was adhered by physicians.

The mean daily capillary blood glucose, pre-prandial capillary blood glucose, and postprandial capillary blood glucose were improved during the treatment period (Fig. 3A, B, C). The mean daily blood glucose decreased significantly from 11.3 (8.0, 13.9) mmol/L in the first 24 h to 7.9 (6.5, 8.9) mmol/L in the last 24 h of the trial. The mean pre-prandial blood glucose decreased significantly from 10.3 (7.1, 12.0) mmol/L in the first 24 h to 7.2 (5.9, 8.8) mmol/L in the last 24 h of the trial. The mean postprandial blood glucose decreased significantly from 12.8 (9.7, 14.7) mmol/L in the first 24 h to 8.3 (7.5, 10.2) mmol/L in the last 24 h of the trial. And the percentage of pre-prandial blood glucose in target range (5.6–7.8) was from 31.8% at the first 24 h to 56% in the last 24 h, which was comparable to previous best-practice study (Umpierrez et al., 2013; Umpierrez et al., 2007). Meanwhile, the mean daily total insulin dose increased over time and remained constant for the rest of the study (Fig. 3D). Significant improvement in seven-point blood glucose profile was seen across all time points in the last 24 h of treatment period compared with the first 24 h: fasting -1.6 (95%CI, $-2.9, -0.4$; $P=0.013$), post-breakfast -3.1 (95%CI, $-5.0, -1.3$; $P=0.0013$), pre-lunch

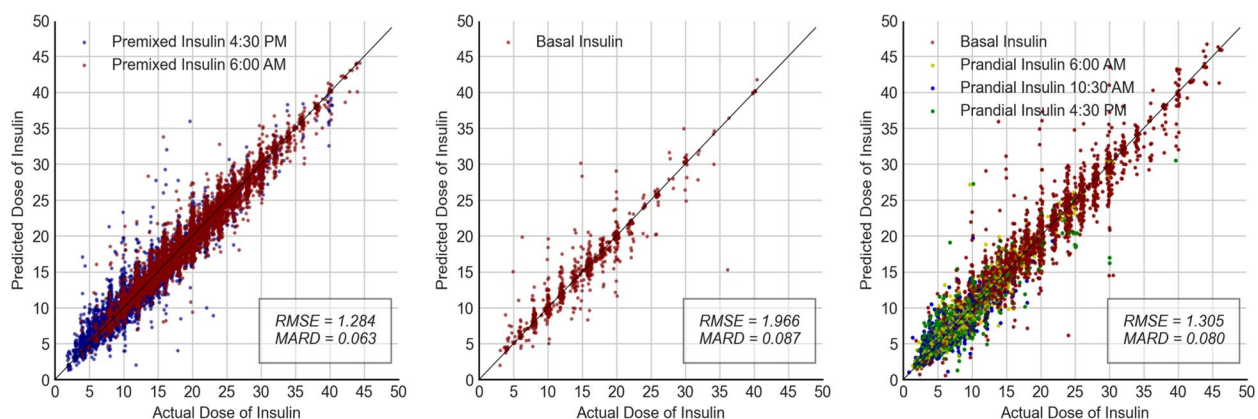


Fig. 2 Distribution of the predicted dose of insulin and the actual dose of insulin

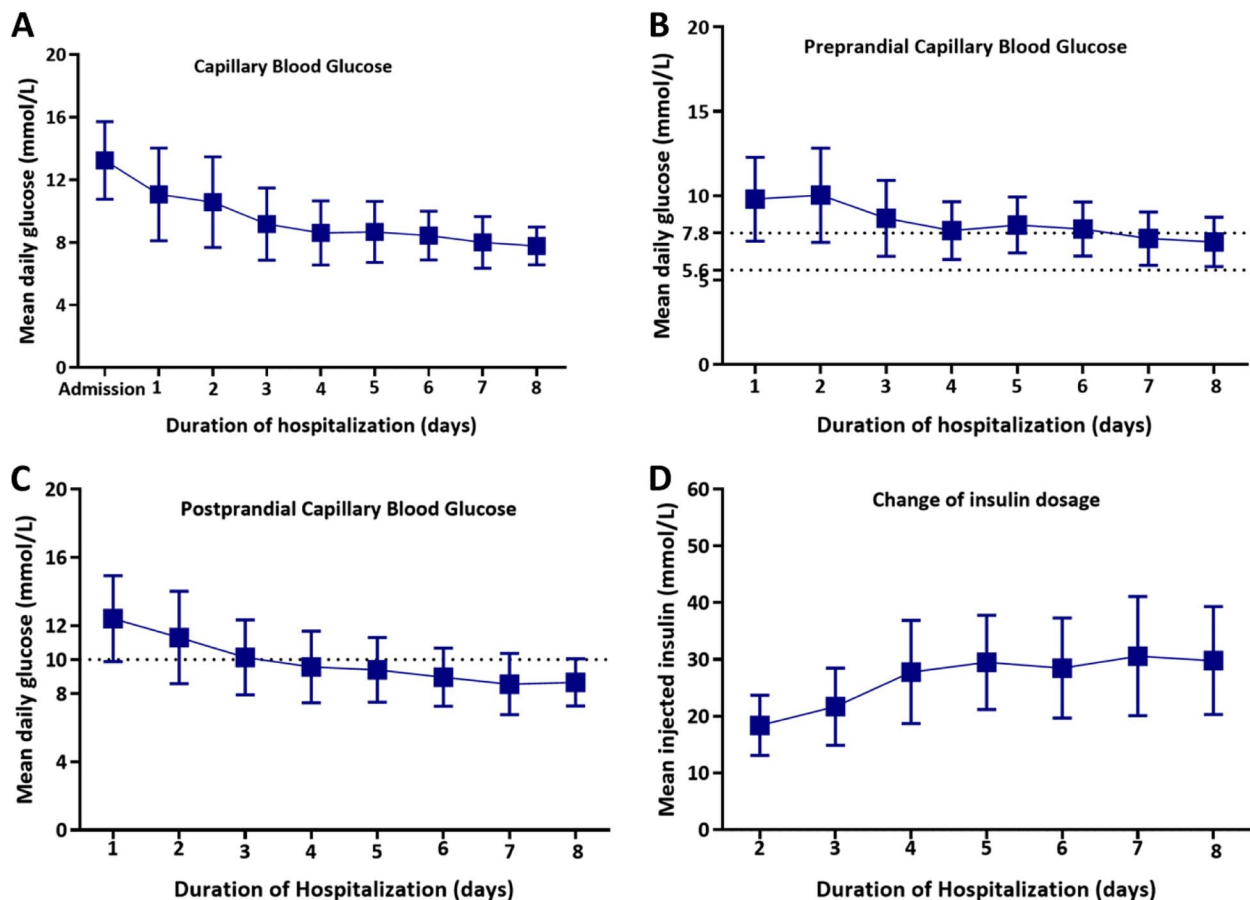


Fig. 3 **A** Mean daily blood glucose. **B** Mean preprandial capillary blood glucose. Preprandial capillary blood glucose demonstrated the fasting, pre-lunch, pre-dinner and pre-bedtime blood glucose. The preprandial blood glucose target was 5.6–7.8 mmol/L. **C** Mean postprandial capillary blood glucose. Postprandial capillary blood glucose demonstrated the post-breakfast, post-lunch, and post-dinner blood glucose. The postprandial blood glucose target was < 10.0 mmol/L. **D** Mean injected total daily insulin. Data are presented as median and interquartiles. Day 2 of hospitalization refers to the first day of the intervention trial

– 2.5 (95%CI, – 4.0, – 0.9; $P=0.0021$), post-lunch – 3.5 (95%CI, – 5.2, – 1.9; $P=0.0001$), pre-dinner – 3.4 (95%CI, – 5.2, – 1.6; $P=0.0004$), post-dinner – 2.3 (95%CI, – 4.2, – 0.4; $P=0.019$), pre-bedtime – 1.4 (95%CI, – 3.1, – 0.3; $P=0.098$) (Fig. 4).

The percentage of glucose concentration in time in range (TIR) (3.9–10.0 mmol/L) was constantly improved for CGM and capillary blood glucose (Fig. 5A and B). TIR (3.9–10.0 mmol/L) was improved from 64.8% in the first 24 h to 92.4% in the last 24 h of treatment period ($P<0.001$). Time spent above 10.0 mmol/L was decreased from 27.6% to 7.1% ($P<0.001$). Time spent below 3.9 mmol/L was decreased from 1.1% to 0.5%. Meanwhile, glycemic variability was improved during treatment period (CV: 27.6% at day 1 VS 25.2% at day 4 VS 20.1% at day 7, respectively) (Supplementary Table 6).

No episodes of severe hypoglycemia (ie, requiring assistance of someone else) or hyperglycemia with ketonemia occurred during the trial. No other

serious adverse events occurred during the period of intervention.

6 Conclusions

In this study, we constructed a novel machine learning model to safely and efficiently recommend insulin dosage for type 2 diabetic patients. The model was built based on accessible variables within the EHR, and was integrated into a decision support system (iNCDSS) under the EHR framework. In practice, the model demonstrated superior performance in recommending insulin dosage titration to the intern and resident physicians and is comparable to attending physicians. In the proof of concept trial, our data demonstrated that subcutaneous insulin dosage titrated according to the iNCDSS system was efficacious and safe for glycemic control. To our knowledge, this is the first time to implement a machine learning algorithm into the EHRs to real time read patient's information and give

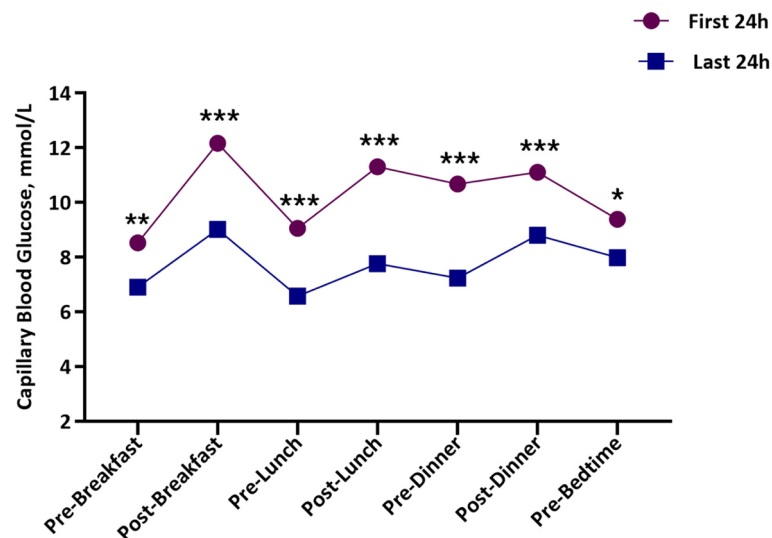


Fig. 4 The seven-point monitored blood glucose profile at the last 24 h and at the first 24 h of treatment period. * $P < 0.10$; ** $P < 0.05$; *** $P < 0.001$

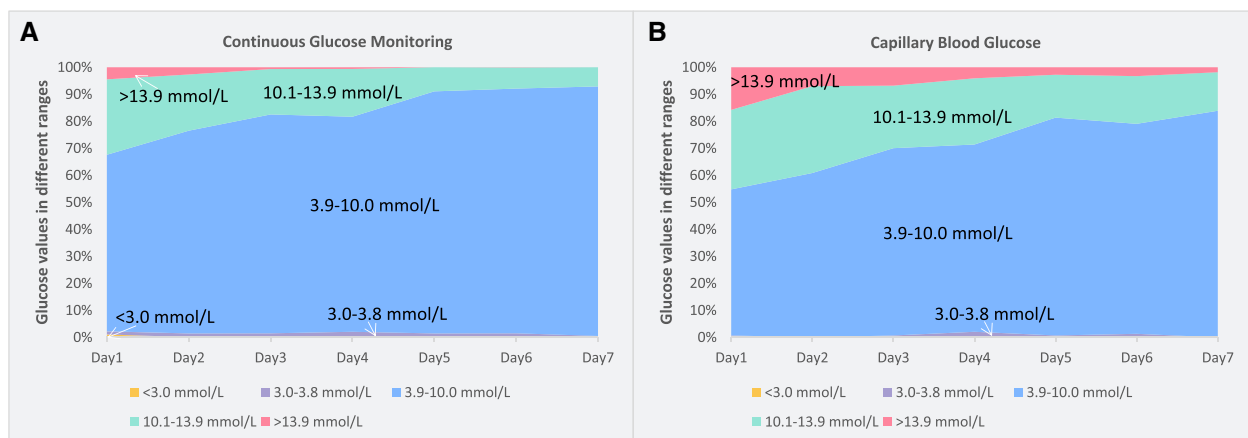


Fig. 5 **A** Mean percentage of continuous glucose monitoring data in different glycaemic ranges during the treatment period. **B** Mean percentage of blood glucose values in different glycaemic ranges during the treatment period

treatment recommendation, and to explore its efficacy and safety in clinical practice by a proof of concept study.

In current clinical practice, insulin dosage titration follows the guidance and to a great extent relies on the physician skill and experience. In order to transform the insulin dosage decision made by experienced physicians into general clinical practice, we built the iNCDSS on machine learning from the experienced physicians, which continuously and real-time recommends insulin dosage for type 2 diabetic patients who receive subcutaneous insulin injection. Among the machine learning models, XGBoost is a highly flexible nonparametric model that integrates a large number of other machine

learning models (decision trees) and has great potential in solving experimental problems. The XGBoost model has shown superiorities in many data mining competitions in recent years due to its advantages (Chen & Guestrin, 2016), which include minimal requirements for attribute normalization, processing missing values intelligently, offering solutions to avoid overfitting (Friedman, 2001). In our current study, XGBoost presented the best performance with the lowest RMSE and MARD and the highest Spearman's correlation coefficient in the testing dataset. Of note, XGBoost performed significantly better than the SVM model, which has been used to predict inpatient insulin dosage (Ruan et al., 2020).

We limited the length of the observation window to 5 days to ensure that the model was based on clinical and time-relevant features. The key features of insulin dosage on specified time points and insulin regimen of our model are not only included in the insulin titration guideline, but also outnumber the guidance. A number of significant variables were detected, which are associated with clinical decision made by physicians. However, carbohydrate intake/meal content was not included in the final model since the size of meal was stable during the hospitalized period. Integration of real-time prediction model for insulin dosage was difficult due to limited implementation and large model complexity. We employed a simplified model using readily accessible variables within the EHR which demonstrated good performance. The model was integrated into a decision support system (iNCDSS) under the EHR framework. Practically, the iNCDSS system could read the clinical information of patients and feed the required information to the model, which then calculates the next insulin dosage.

We evaluated the feasibility and safety of the iNCDSS system for glycemic control in type 2 diabetic patients. 52% diabetic patients achieved the target of 5.6–7.8 mmol/L for premeal and pre-bedtime blood glucose and 75% diabetic patients achieved the target of less than 10 mmol/L for post meal blood glucose at the last 24 h of the trial, which is well acceptable (Umpierrez et al., 2013; Umpierrez et al., 2007). In addition, we have used CGM to closely monitor blood glucose levels to assess the iNCDSS system for efficacy and safety. Time in range assessed by CGM (3.9–10.0 mmol/L) was greatly improved from 64.8% in the first 24 h to 92.4% in the final 24 h of the trial, while time in hypoglycaemia substantially decreased. The low occurrence of hypoglycaemia whilst achieving glycemic control might be attributed to the algorithm of real time insulin dosage titration once a decreased trend was detected. Glycemic variability is associated with increase in cardiovascular morbidity and mortality (Ceriello et al., 2019). In our study, coefficient of variation of blood glucose was improved to 25.2% at day 4 and to 20.1% at day 7, which was lower than the threshold recommended by the ADA guideline (American, 2021). Bally et al. (2018) has tried to apply artificial pancreas technology to establish euglycemia in type 2 diabetes inpatients on the general ward. The performance of iNCDSS system was comparable to the closed-loop insulin delivery system with regard to mean glucose (7.9 mmol/L VS 8.5 mmol/L). In addition, there was a high level of satisfaction among the physicians who used the iNCDSS in the trial. Physicians did reject the recommendations by the iNCDSS system in 9 times of 3 patients. However, the overriding of algorithm was very low and was unlikely to influence the outcomes.

Our work has the following clinical applications. First, it can facilitate to improve glycemic control for inpatients with type 2 diabetes in other wards. One of the challenges for inpatient glucose control in other wards is lack of experienced physician in dealing with insulin dosage titration. Therefore, our iNCDSS system is more likely to be needed by non-endocrinologists in clinical practice. Second, it could assist primary care physicians in optimizing insulin dosage titration. Given the increasing prevalence of diabetes and the mounting shortage of endocrinologists, primary care physicians would take care more and more diabetic patients. Innovative and easily accessible iNCDSS would definitely facilitate in managing blood glucose in more patients. Moreover, the iNCDSS presented the ability to standardize the health care, ensuring that all people with type 2 diabetes would receive high quality care, even if they live in remote areas and taken care by primary care physicians. However, the utility of this approach should be tested in designated prospective clinical trials. Our current study therefore paved the way for future clinical trials to evaluate the feasibility and safety of the iNCDSS system for glucose control in primary care clinics.

There are some strengths in our current study. First, we evaluated a wide range of machine learning models and compared their predicting capacity against the most commonly used statistical model, which we used as a benchmark model. Furthermore, we constructed a simplified and well performed model, which holds benefits for ready implementation into the EHR. Second, we incorporated the model into the existing EHR to develop a workflow that could real time access the patient's information and continuously give insulin dosage recommendation. Third, the efficacy and safety of iNCDSS system were validated in a proof of concept trial. Finally, the iNCDSS system was personalized data-driven and could be consistently updated as more data were available.

Although iNCDSS achieved good performance in insulin dosage titration, some challenges remained. As designed, our current model attempts to predict insulin dosage recommended by senior physicians. The predictive capability achievable is thus far from a fully automated system. Given the proposed machine learning method is purely data-driven, our model may vary whenever different datasets are learned. We have chosen the data from the top hospital in which the EHRs are strictly required to be standardized and comprehensively documented, and the doctor's order originated from experienced attending physicians. Considering the insulin dosing variance in different levels of hospitals and physicians, we only included EHRs from our hospital. Meanwhile in the hospital, the seven point capillary blood glucose levels were commonly

measured and were integrated into the EHRs for a long period of time, which warranted the dataset development and paved the way for building the workflow. Second, although we had a pool of more than 300 clinical features, our modelling principle is a trade-off between having a minimal number of features and the capacity of good prediction, therefore avoiding overfitting. Thirdly, the trial only provides a preliminary assessment of the clinical course and outcome of patients. A large and multi-center clinical trial is needed.

In conclusion, we developed an insulin dosage titration model using advanced machine learning methods. The model has been implemented in the EHR workflow to form clinical decision support system of insulin dosage titration. And the efficacy and safety of the iNCDSS system has been preliminarily evaluated in type 2 diabetes inpatients. The iNCDSS provides opportunities for a new modality for intensive insulin management in type 2 inpatients.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1007/s44194-023-00020-7>.

Additional file 1: Supplementary Table 1. Features extraction. **Supplementary Figure 1.** The insulin dosage recommendation interface. **Supplementary Table 2.** Clinical characteristics of type 2 diabetic patients in training and testing datasets. **Supplementary Table 3.** Features importance in the artificial intelligence policy. **Supplementary Table 4.** Illustration of insulin dosage titration performance of the iNCDSS model and physicians. **Supplementary Table 5.** Baseline clinical characteristics of the patients in the trial ($N=23$). **Supplementary Table 6.** Glucose control based on sensor glucose measurements at the first 24 h and the last 24 h of the trial

Acknowledgements

We would like to thank the physicians and nurses at the Department of Endocrinology and Metabolism for their help in the conduction of the clinical trial, and SL in the Yidu Cloud Technology Inc. and the network engineer at the Computer Network Center and Big Data and Artificial Intelligence Center for their help in establishing and maintaining of the clinical decision support platform. The authors are grateful to the field investigators for their work and the participants for their cooperation.

Authors' contributions

XL designed the study and interpreted the data. YC, LZ, PY, HY, HC, CY, JZ, QM, YL, LC, YS, CH, HH, XL, HB collected the data. YC, CZ and SL analyzed and interpreted the data and YC wrote the manuscript. XL revised the manuscript. All authors approved the final submitted version of the manuscript.

Funding

This work is supported by the grants from the National Nature Science Foundation (No. 82000822, 31830041). The funders of the study did not contribute to design, data collection, data analysis, data interpretation, writing of the report, or the decision to submit the report for publication. The corresponding author had full access to all the data in the study and had final responsibility for the decision to submit for publication.

Declarations

Competing interests

Author Xiaoying Li is a member of the Editorial Board for *Current Medicine*. The paper was handled by the other Editor and has undergone rigorous peer

review process. Author Xiaoying Li was not involved in the journal's review of, or decisions related to, this manuscript.

Author details

¹Ministry of Education Key Laboratory of Metabolism and Molecular Medicine, Department of Endocrinology and Metabolism, Zhongshan Hospital, Fudan University, Shanghai, China. ²Big Data and Artificial Intelligence Center, Zhongshan Hospital, Fudan University, Shanghai, China. ³Yidu Cloud Technology Inc., Beijing, China. ⁴Computer Network Center, Zhongshan Hospital, Fudan University, Shanghai, China.

Received: 30 September 2022 Accepted: 8 January 2023

Published online: 03 February 2023

References

- Abadi M, Agarwal A, Barham P, et al. Large-scale machine learning on heterogeneous systems: TensorFlow; 2015.
- Akirov A, Grossman A, Shochat T, Shimon I. Mortality among hospitalized patients with hypoglycemia: insulin related and noninsulin related. *J Clin Endocrinol Metab*. 2017;102:416–24.
- American DA. 6. Glycemic targets: standards of medical Care in Diabetes-2021. *Diabetes Care*. 2021;44:S73–84.
- Bally L, Thabit H, Hartnell S, Andereggen E, Ruan Y, Wilinska ME, et al. Closed-loop insulin delivery for glycemic control in noncritical care. *N Engl J Med*. 2018;379:547–56.
- Basu S, Yudkin JS, Kehlenbrink S, Davies JI, Wild SH, Lipska KJ, et al. Estimation of global insulin use for type 2 diabetes, 2018–30: a microsimulation analysis. *Lancet Diabetes Endocrinol*. 2019;7:25–33.
- Ceriello A, Monnier L, Owens D. Glycaemic variability in diabetes: clinical and therapeutic implications. *Lancet Diabetes Endocrinol*. 2019;7:221–30.
- Chen T, Guestrin C. XGBoost: a scalable tree boosting system. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 2016. p. 785–94.
- Cho NH, Shaw JE, Karuranga S, Huang Y, da Rocha Fernandes JD, Ohlrogge AW, et al. IDF Diabetes Atlas: Global estimates of diabetes prevalence for 2017 and projections for 2045. *Diabetes Res Clin Pract*. 2018;138:271–81.
- International Diabetes Federation. IDF Diabetes Atlas, 8th edn. Brussels, 2017. <http://diabetesatlas.org/resources/2017-atlas.html>.
- DeFronzo RA, Ferrannini E, Groop L, Henry RR, Herman WH, Holst JJ, et al. Type 2 diabetes mellitus. *Nat Rev Dis Primers*. 2015;1:15019.
- Friedman JH. Greedy function approximation: a gradient boosting machine. *Ann Stat*. 2001;29(5):1189–232.
- Hayward RA, Reaven PD, Wiitala WL, Bahn GD, Reda DJ, Ge L, et al. Follow-up of glycemic control and cardiovascular outcomes in type 2 diabetes. *N Engl J Med*. 2015;372:2197–206.
- Holman RR, Paul SK, Bethel MA, Matthews DR, Neil HA. 10-year follow-up of intensive glucose control in type 2 diabetes. *N Engl J Med*. 2008;359:1577–89.
- Liang H, Tsui BY, Ni H, Valentim CCS, Baxter SL, Liu G, et al. Evaluation and accurate diagnoses of pediatric diseases using artificial intelligence. *Nat Med*. 2019;25:433–8.
- Nawaz MS, Shah KU, Khan TM, Rehman AU, Rashid HU, Mahmood S, et al. Evaluation of current trends and recent development in insulin therapy for management of diabetes mellitus. *Diabetes Metab Syndr*. 2017;11(Suppl 2):S833–9.
- Ngiam KY, Khor IW. Big data and machine learning algorithms for health-care delivery. *Lancet Oncol*. 2019;20:e262–73.
- Nimri R, Dassau E, Segall T, Muller I, Bratina N, Kordonouri O, et al. Adjusting insulin doses in patients with type 1 diabetes who use insulin pump and continuous glucose monitoring: variations among countries and physicians. *Diabetes Obes Metab*. 2018;20:2458–66.
- Nimri R, Battelino T, Laffel LM, Slover RH, Schatz D, Weinzier SA, et al. Insulin dose optimization using an automated artificial intelligence-based decision support system in youths with type 1 diabetes. *Nat Med*. 2020;26:1380–4.
- Pedregosa F, Varoquaux G, Gramfort A, et al. Scikit-learn: machine learning in Python. *J Mach Learn Res*. 2011;12:2825–30.
- Petersen MC, Shulman GI. Mechanisms of insulin action and insulin resistance. *Physiol Rev*. 2018;98:2133–223.

- Ruan Y, Bellot A, Moysova Z, Tan GD, Lumb A, Davies J, van der Schaar M, Rea R. Predicting the risk of inpatient hypoglycemia with machine learning using electronic health records. *Diabetes Care*. 2020;43:1504–11. <https://doi.org/10.1101/2020.03.02.20029017>.
- Stratton IM, Adler AI, Neil HA, Matthews DR, Manley SE, Cull CA, et al. Association of glycaemia with macrovascular and microvascular complications of type 2 diabetes (ukpds 35): prospective observational study. *BMJ*. 2000;321:405–12.
- Tomasev N, Glorot X, Rae JW, Zielinski M, Askham H, Saraiva A, et al. A clinically applicable approach to continuous prediction of future acute kidney injury. *Nature*. 2019;572:116–9.
- Tyler NS, Mosquera-Lopez CM, Wilson LM, Dodier RH, Branigan DL, Gabo VB, et al. An artificial intelligence decision support system for the management of type 1 diabetes. *Nat Metab*. 2020;2:612–9.
- Umpierrez GE, Smiley D, Zisman A, Prieto LM, Palacio A, Ceron M, et al. Randomized study of basal-bolus insulin therapy in the inpatient management of patients with type 2 diabetes (RABBIT 2 trial). *Diabetes Care*. 2007;30:2181–6.
- Umpierrez GE, Smiley D, Hermayer K, Khan A, Olson DE, Newton C, et al. Randomized study comparing a basal-bolus with a basal plus correction insulin regimen for the hospital management of medical and surgical patients with type 2 diabetes: basal plus trial. *Diabetes Care*. 2013;36:2169–74.
- Van Calster B, Wynants L. Machine learning in medicine. *N Engl J Med*. 2019;380:2588.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.