



Machine Learning Models for Inpatient Glucose Prediction

Andrew Zale¹ · Nestoras Mathioudakis¹

Accepted: 18 May 2022 / Published online: 27 June 2022

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2022

Abstract

Purpose of Review Glucose management in the hospital is difficult due to non-static factors such as antihyperglycemic and steroid doses, renal function, infection, surgical status, and diet. Given these complex and dynamic factors, machine learning approaches can be leveraged for prediction of glucose trends in the hospital to mitigate and prevent suboptimal hypoglycemic and hyperglycemic outcomes. Our aim was to review the clinical evidence for the role of machine learning–based models in predicting hospitalized patients’ glucose trajectory.

Recent Findings The published literature on machine learning algorithms has varied in terms of population studied, outcomes of interest, and validation methods. There have been tools developed that utilize data from both continuous glucose monitors and large electronic health records (EHRs). With increasing sample sizes, inclusion of a greater number of predictor variables, and use of more advanced machine learning algorithms, there has been a trend in recent years towards increasing predictive accuracy for glycemic outcomes in the hospital setting. While current models predicting glucose trajectory offer promising results, they have not been tested prospectively in the clinical setting.

Summary Accurate machine learning algorithms have been developed and validated for prediction of hypoglycemia and hyperglycemia in the hospital. Further work is needed in implementation/integration of machine learning models into EHR systems, with prospective studies to evaluate effectiveness and safety of such clinical decision support on glycemic and other clinical outcomes.

Keywords Machine learning · Artificial intelligence · Glucose · Insulin · Hospital · Diabetes

Introduction

Diabetes mellitus (DM) is a prevalent condition in the hospital, present in nearly 1 in 5 inpatients [1]. Considering how various dynamic factors (e.g., nutritional status, steroid and antihyperglycemic medications, renal function, and infection) may influence glucose homeostasis over a relatively short time frame, it can be difficult to attain or maintain glycemic

targets in hospitalized patients [2, 3]. Attaining and maintaining euglycemia is important in the hospital as both hyperglycemia and hypoglycemia have been associated with increased morbidity, mortality, and healthcare expenditures. [4–7]

To address the challenge of glycemic management throughout a patient’s evolving and often complex hospital course, there has been growing interest in and use of clinical decision support tools that provide insights into glucose trends. Machine learning using large electronic health record (EHR) data has been increasingly utilized to facilitate clinical decision-making for a broad range of clinical conditions [8, 9]. In the ambulatory setting, there is a large body of evidence for machine learning–based glucose and insulin dosing prediction, but there are fewer tested algorithms in hospitalized patients [10]. Tools that have been developed to predict glucose trajectory include continuous glucose monitoring (CGM) technology and machine learning models based on point-of-care and serum glucose measurements or CGM [11]. In recent years, there has been a growing interest in using CGM

This article is part of the Topical Collection on *Hospital Management of Diabetes*

✉ Nestoras Mathioudakis
nmathio1@jh.edu

Andrew Zale
azale2@jhmi.edu

¹ Division of Endocrinology, Diabetes & Metabolism,
Division of Biomedical Informatics and Data Science,
Johns Hopkins University School of Medicine, Baltimore,
MD 21287, USA

in the hospital setting as a decision support tool, akin to telemetry monitoring [12]. However, due to challenges associated with employing CGMs in the hospitalized setting, there has also been interest in employing non-CGM-based glycemic prediction algorithms using large EHR datasets. While glycemic prediction models do not necessarily recommend to a provider the best course of action to maintain euglycemia, they can offer insights into glycemic trends to guide therapeutic adjustments in the antihyperglycemic regimen.

The purpose of this review is to summarize machine learning-based models and clinical decision support tools for glycemic prediction in hospitalized patients. Our review focuses on the methodology used, hospital setting (ICU vs. non-ICU), model performance, and impact on clinical outcomes.

Machine Learning Overview

Machine learning is the computational process of taking an input of data (i.e., independent variables) and an output (i.e., dependent variable) and uses different statistical rules based on the parameters of each model to create an algorithm that represents the relationship between the raw data and the output. Machine learning has been used throughout medicine to find patterns in genome sequencing, improve diagnostic pathology, and filter relevant problems during a patient's admission [13]. EHR data is at the forefront of machine learning techniques due to the amount of raw information it possesses, and has been used to predict various disease phenotypes or diagnoses (e.g., congestive heart failure, chronic obstructive pulmonary disease). [14]

Within the field of diabetes, machine learning has been implemented broadly to improve detection of diabetes diagnosis [15] and complications [16], cardiovascular risk stratification [17], and insulin dosing [18]. An early application of machine learning in diabetes was a neural network model that achieved a sensitivity and specificity of 0.76 in predicting the onset of diabetes [19]. Since then, landmark studies in the field have included a gradient boosting model that predicts glucose response prediction based on personal and microbiome factors [20] and a deep learning model that identifies diabetic retinopathy [21]. Most notably, an initial closed-loop insulin delivery system was based on a rules-based model; [18] since then, bi-hormonal artificial pancreas technology has been associated with improved glycemic control when compared to sensor-augmented insulin pump therapy [22]. Over the past few years, there has been growing scientific interest in the use of machine learning with large EHR datasets to predict glucose trends or required insulin doses in hospitalized patients.

Phases of Machine Learning Research

Figure 1 outlines the development phases of machine learning-based research with application to the clinical topic of inpatient glycemic prediction [23]. These steps include problem selection, data collection, machine learning development, model validation, assessment of impact, and deployment and monitoring [24]. It is important to recognize that key differences in each of these phases of machine learning research will influence model results, comparability, and generalizability.

Problem Selection and Study Population

The first step in machine learning development is defining the study problem and population of interest. The study problem defines the dependent variable, specifically the outcome that is being predicted (i.e., hypoglycemia over the next 24 h, hypoglycemia during any portion of a patient's admission, etc.). For inpatient glycemic prediction, potential populations include patients with known diabetes or specific type of diabetes (type 1 vs. type 2), patients with acute/stress hyperglycemia, and exposure to specific anti-hyperglycemic medications; populations may be limited to ICU or general floor patients. Selection of the study population will directly affect the generalizability and clinical application of the model. For example, a machine learning model derived from patients with only diagnosed T2DM at the time of admission could not be universally applied to all patients with diabetes.

The method of evaluating model performance (i.e., how accurately the model predicts the outcome in either past or future patients) will also be driven by the problem of interest. For example, if the problem is predicting the next glucose measurement based on CGM, quantitative values for the error, such as mean average error, can be used to measure how far a prediction deviates from the true value; however, if clinical context is more valued than quantitative accuracy and the goal is to properly treat patients, Clarke error grid proportions can be reported [25]. If the problem is prediction of hypoglycemia or hyperglycemia, these binary outcomes can be reported as sensitivity/specificity, positive and negative predictive values, positive and negative likelihood ratios, or area under the receiver operating curve (AUC) [26]. AUC relates the tradeoff a model makes between sensitivity and specificity—a higher AUC indicates overall higher sensitivity and specificity. Generally, an AUC of 0.5 represents no discriminatory value, 0.7–0.8 is considered acceptable, and greater than 0.9 is considered outstanding. It is worth noting that for rare outcomes such as hypoglycemia, the AUC may not

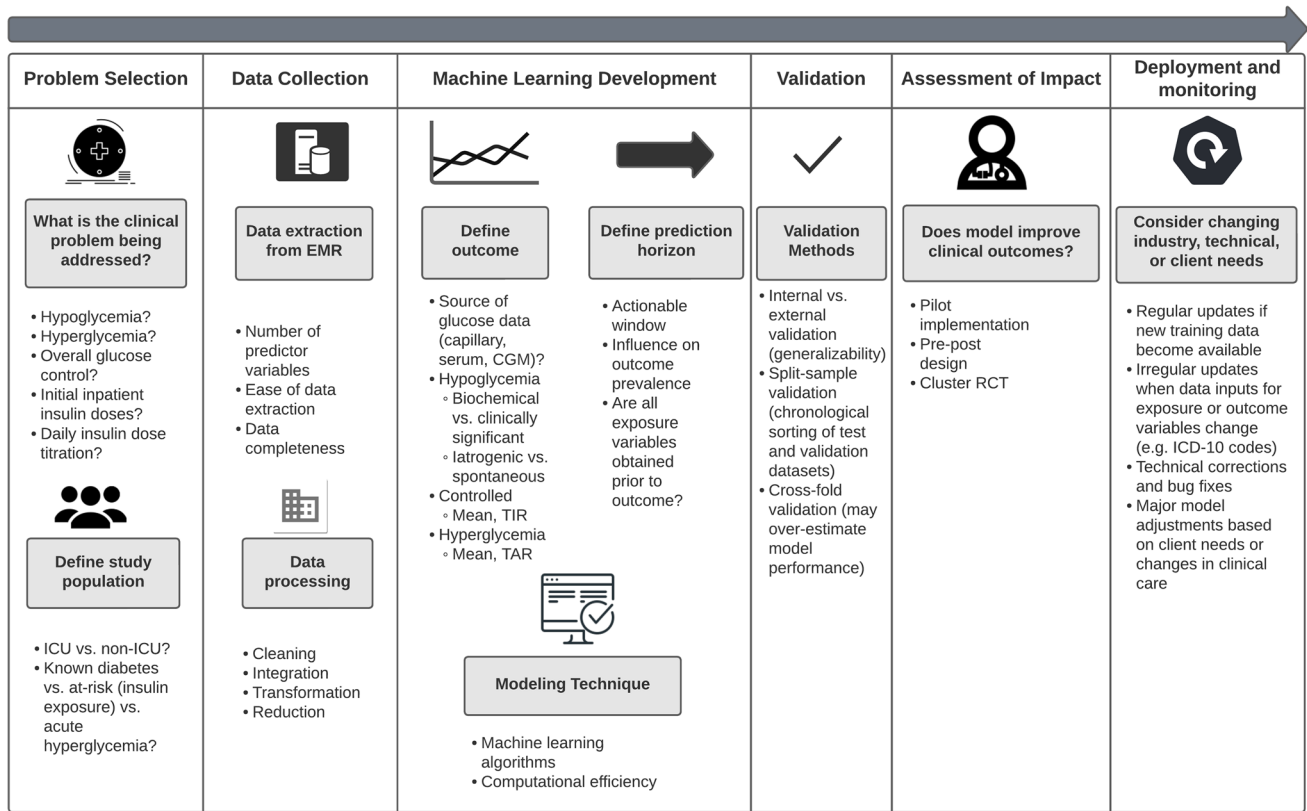


Fig. 1 Phases of research in machine learning models for inpatient glucose management. (*ICU* intensive care unit, *CGM* continuous glucose monitor, *TIR* time in range, *TAR* time above range, *RCT* randomized controlled trial, *ICD* International Classification of Diseases)

sufficiently quantify the accuracy of a model because of the imbalance between events (hypoglycemia) and non-events. In such cases, a precision-recall plot may be more informative. [27]

Data Extraction and Processing

After defining the study question and population, data related to clinical predictors and the outcome measure must be extracted from the EHR. Challenges at this stage include the availability of predictor variables and the ease of extracting relevant information from within the EHR. A collaborative and iterative approach between clinical content experts and EHR vendor certified analysts is recommended to identify and validate the data elements and sources. Since machine learning entails use of data from thousands of patients, another important consideration at this is ensuring the security of the data and minimizing breach of confidentiality by using secure virtual environments for data analysis and considering use of limited (deidentified) datasets when possible.

After a raw dataset is assembled, it must undergo “data tidying,” which consists of organizing the dataset in such a way that it is usable to answer the question at hand based on the defined exposure variables, index unit of observation

(e.g., calendar day, patient day, each glucose reading), and the prediction horizon (e.g., next glucose reading, next 24 h, remainder of admission). For example, if a machine learning algorithm attempts to predict hypoglycemia every calendar day but the dataset was extracted for every glucose observation, the dataset will need to be constructed longitudinally such that each row represents the index unit of observation (e.g., calendar day).

Moreover, exposure variables must be temporally related to the outcome variable such that all exposures occurred prior to the outcome (i.e., insulin doses administered in the previous 24 h to predict glucose values in the next 24 h). In some cases, machine learning models have used aggregate data across an entire patient stay to predict glycemic outcomes occurring at any time during admission, which is in fact more of a cross-sectional design (association) rather than longitudinal design (prediction). When insulin or other medications are being used as relevant exposure variables, one should consider the pharmacokinetic profile of the different insulin types to determine insulin dose on board at the time of the prediction. Similarly, glucocorticoids, which can have a significant influence on glucose homeostasis, may need to be normalized to a common equivalent dose (e.g., hydrocortisone or prednisone equivalents) and their

pharmacokinetic profile would need to be accounted for to determine the active glucocorticoid dose at the time of the index prediction. These data processing steps can be labor intensive, requiring a series of data merges, collapsing of data into summary measures, and computations.

Machine Learning Development

Defining the Outcome

The next key step is defining the outcome of interest, and variation among different research groups in defining glucose-related outcomes will influence model performance and inter-model comparability. For example, studies exploring hypoglycemia may use different cutoffs (glucose ≤ 70 , < 70 , < 54 , < 50 , < 40 mg/dl, etc.), and may apply different criteria to distinguish iatrogenic from spontaneous hypoglycemia. Similarly, various definitions of controlled glucose or hyperglycemia could be considered based on mean glucose values, peak glucose values, or proportion of readings within a defined range (e.g., 70–180 mg/dl). Differences in glycemic outcome definitions will have a direct effect on outcome prevalence and, accordingly, model performance (sensitivity and specificity). [28] For example, choosing more restrictive hypoglycemia cutoffs (< 54 mg/dL vs. < 70 mg/dL) or narrowing the prediction horizon (hypoglycemia with next glucose measurement vs. at any time during admission) will decrease the prevalence of the outcome, thereby affecting the sensitivity and specificity of the model (i.e., it can be harder to accurately predict a more rare outcome). [29]

Defining the Prediction Horizon

The prediction horizon (i.e., how far ahead the model attempts to predict the outcome) will have important implications on clinical relevance, usability, and model performance. For example, a shorter prediction window (e.g., next patient day) may be more actionable (i.e., provide a definite time period for clinician intervention), whereas a larger prediction window (e.g., patient admission) increases the prevalence of an outcome and thus model performance, but may have diminished value as a decision support tool throughout the entire hospitalization. Models that attempt to predict risk of a dysglycemic outcome at any time during hospitalization may be useful to flag patients at risk at the time of admission but would have less clinical utility for inpatient clinicians for day-to-day adjustments of the antihyperglycemic regimen.

Modeling Technique

Once the outcome and prediction window have been defined, a tidied dataset can be used to create a novel machine learning algorithm. There are many different machine learning

techniques—the simplest is linear regression, which infers that there is a linear relationship between exposure variables and the outcome variable and fits slopes for each exposure variable [30]. Although complex modeling techniques can be more robust in creating non-linear models, they are at risk of being computationally inefficient.

Modeling techniques require different amounts of computational storage since they employ unique methodologies to create the algorithm. The simplest of models for quantitative prediction is simple linear regression, which creates a linear model that relates the predictor to the outcome with a slope and y-intercept. For two-class classification (i.e., yes or no), the equivalent would be logistic regression, where each predictor has an associated slope and a y-intercept to output a probability that a prediction is either “yes” or “no.”

The primary limitation to linear and logistic regression is that as relationships between predictors and outcome become less linear, the predictive accuracy of the models worsen [31]. Thus, there has been growing interest in utilizing modeling techniques that do not impose a given relationship on outcome and predictors to leverage information from big datasets. Other modeling techniques used for prediction include *k*-nearest neighbors, tree-based methods such as random forests, and neural networks.

In *k*-nearest neighbors, observations in the training set are plotted in as many dimensions as there are predictors. Then, a prediction is made based on the average of the *k* nearest points. Simply, *k*-nearest neighbors make a prediction based on the outcome of the observations most similar to the point of interest. Tree-based methods create cutoffs of different predictor values that are set as rules, and a prediction is then made based on those rules. To improve model performance, multiple trees can be used with the resulting prediction based on the average prediction from all the trees in the random forest. In neural networks, the predictors are used to build new functions that are then related to the outcome. The term neural network comes from the new intermediate functions built between predictors and outcome, which were considered analogous to neurons either activating or inhibiting a downstream neuron (the prediction).

In addition to logistic regression, most published models in inpatient glycemic prediction use a derivative of tree-based models. The primary derivatives of tree-based models are bagging, random forests, and boosting [30]. In bagging, repeated random samples are taken from the dataset and predicted upon to allow for multiple trees to be created and reduce the variance of the overall model. Random forests go beyond the bagging technique by limiting the predictors that could be chosen from when creating each rule to decorrelate the trees and make them more reliable. The notion behind random forests is that there are a set of variables that are most correlated to the outcome of interest, but by forcing the model to randomly choose from less correlated predictors,

the model will better capture the nuances between predictors. Boosting is unique in that rather than creating multiple trees attempting to predict the same outcome, each sequential tree in boosting is built off the residual, or errors, in the previous tree such that each tree is built sequentially. Different machine learning methods such as stochastic gradient boosting and XGBoost modify the boosting technique in an attempt to maximize model performance or improve processing efficiency. [32]

Model Validation

A developed machine learning model can then be tested on a set of patients from the same hospital (internal validation) or different hospitals (external validation) who are naïve to the model to estimate model performance. Previous research has highlighted that external validation is crucial to algorithm development, as internal validation can overestimate model performance on a future cohort of patients [33]. Differences in validation methods using retrospective data may affect how a model performs after prospective evaluation, as random cross-fold validation may not account for secular trends in the hospital setting that could affect the prevalence of the outcome. This technique of blocking observations is based on the concern that in structured data models may perform worse if observations are randomly assigned to the training and test set [34]. For inpatient glucose prediction, data are inherently structured by time: increasing understanding of dysglycemia and new protocols that are developed over time play a role in mitigating hypoglycemia and hyperglycemia, but the secular trends are poorly captured in machine learning models. By dividing earlier observations into the training set and more recent observations into the test set, researchers can report model performance statistics that reflect changes in the healthcare system not directly accounted for by included predictor variables.

Assessment of Impact and Deployment and Monitoring

After a model has been developed and validated, it can be integrated into the EHR and evaluated prospectively for effectiveness in reducing the desired outcome of interest. This process requires a near real-time data feed from the EHR into the algorithm, with output returned to the end user through the EHR. Some EHR vendors, such as Epic Systems and Cerner, have built-in tools for predictive analytics to simplify the process of getting data out and back into the EHR in real time. For example, the Epic Cognitive Computing platform (EpicCare, Verona), which leverages Microsoft Azure cloud software, can be used to configure prediction models with EHR data, create and modify workflows based on the results of the predictive models, and report outcomes of the models and expected workflows [35]. Recently, the

Epic Cognitive Computing platform was used to quickly develop a model for COVID-19-related outcomes [35]. Similarly, Cerner has partnered with Amazon Web Services to create a cloud-based health platform for predictive analytics called “Project Apollo.” [35] It is expected that these embedded tools for predictive analytics will accelerate the deployment of machine-learning based prediction models by streamlining the process for clients and by using native functionality of the EHR. Various formats for the output of the model could be considered, such as a best practice advisory/alert (either intrusive or non-intrusive) or a report embedded with the glucose/insulin management section of the EHR.

After deployment of an EHR-based prediction model, fixed updates may be needed to revise prediction models as new data become available, while ad hoc updates may be needed if there are new therapies (e.g., new insulins or oral antihyperglycemic medications). In addition, feedback from end users about clinical utility of the EHR-based model should be considered to adjust the model based on end user needs or any changes in clinical care practices.

Machine Learning Models for Inpatient Glucose Prediction

Due to the established adverse effects of poor glycemic control on patient outcomes, there has been growing research over the past decade in predicting dysglycemia using machine learning approaches in the hospital setting [4]. As studies in the field of inpatient glucose prediction continue to accrue, there has been a gradual progression from traditional regression modeling to more advanced machine learning models, with studies using larger datasets and much larger numbers of clinical predictors showing continually improving predictive accuracy. Table 1 summarizes validated machine learning models for glucose prediction in the hospital. We review these key studies with respect to the study populations, outcome definitions, key exposure variables, prediction horizon, validation method, and model performance.

Logistic Regression Models for Inpatient Glucose Prediction

Early machine learning methods relied on relatively small sample sizes and used logistic regression to predict the risk of hypoglycemia occurring during hospital admission. In 2012, Elliott et al. used a logistic regression model with nine predictors that included prandial insulin, sulfonylurea, basal insulin, weight, and renal function [36]. They employed a training set of 172 patients and presented model performance on a test set of 3,028 patients; notably, their inclusion criteria for the test set included patients with glucose < 90 mg/

Table 1 Validated models for glucose prediction in hospitalized patients

Author, year	Number of patients	Modeling technique	Clinical predictors	Prediction horizon	Outcome definition	Validation approach	Model performance
Elliott, 2012 [36]	Training set = 172, test set = 3028	Logistic regression	9 total, including prandial insulin, sulfonyleurea, basal insulin, weight, and renal function	Hospital stay	Hypoglycemia < 70 mg/dL, < 60 mg/dL, and < 40 mg/dL	Internal validation	50% sensitivity cutoff for hypoglycemia < 70 mg/dL, < 60 mg/dL, and < 40 mg/dL identified 71%, 70%, and 55% of hypoglycemic events respectively Model AUC was 0.733 (95% CI: 0.719–0.747)
Stuart, 2017 [37]	9584 admissions	Logistic regression	13 total including demographics, insulin and sulfonyleureas, and electrolytes	Hospital stay	Hypoglycemia < 4.0 mmol/L	Bootstrapping	Model AUC in validation cohort was 0.71 (95% CI: 0.63–0.79)
Ena, 2018 [38]	Training set = 839, test set = 561	Logistic regression	4 total: GFR < 30 mL/min/1.73 m ² , insulin dose > 0.3 units/kg/day, length of stay, previous episode of hypoglycemia during 3 months before admission	Hospital stay	Hypoglycemia < 70 mg/dL	External validation	Model AUC in validation cohort was 0.71 (95% CI: 0.63–0.79)
Mathioudakis, 2018 [39]	Training set = 13,360, test set = 5902	Logistic regression	44 variables in model predicting ≤ 70 mg/dL and 35 variables in model predicting < 54 mg/dL	24 h	Hypoglycemia ≤ 70 mg/dL and < 54 mg/dL	Internal validation	C-statistic of 0.77 (95% CI: 0.75–0.78 and 0.80 (95% CI: 0.78–0.82) were achieved for models predicting hypoglycemia ≤ 70 mg/dL and < 54 mg/dL respectively
Winterstein, 2018 [40]	21,840 patients	Logistic regression	38 variables	24 h	Hypoglycemia < 50 mg/dL not followed by glucose value > 80 mg/dL within 10 min	Bootstrapping	C-statistic of 0.887 (95% CI: 0.874–0.899) was achieved for predicting hypoglycemia on days 3–5
Shah, 2019 [41]	Training set = 300 patients, test set = 300 patients	Logistic regression	5 variables: age, ED visit 6 months prior, insulin use, use of oral agents that do not induce hypoglycemia, and severe CKD	Hospital stay	Hypoglycemia ≤ 70 mg/dl	External validation	Validation cohort c-statistic was 0.642 with a sensitivity of 0.77 and specificity of 0.28 when using a risk score cutoff of ≥ 9

Table 1 (continued)

Author, year	Number of patients	Modeling technique	Clinical predictors	Prediction horizon	Outcome definition	Validation approach	Model performance
Kim, 2020 [42]	20 patients	Recurrent neural network	CGM data	30 min	Quantitative glucose value	Internal validation	Average root mean squared error of prediction was 21.5 mg/dL and mean absolute percentage error was 11.1%
Kyi, 2020 [43]	594 patients	Logistic regression	10 variables	Hospital stay	At least 2 days with capillary glucose < 72 or > 270 mg/dl	Internal validation	Early identification of persistent adverse glycemia had ROC of 0.806, with sensitivity, specificity and PPV of 0.84, 0.66, and 0.53 respectively
Ruan, 2020 [8]	17,658 patients	XGBoost	42 variables	Hospital stay	Hypoglycemia < 4.0 mmol/L and < 3.0 mmol/L	Internal validation	AUROC was achieved of 0.96 for the XGBoost model in predicting hypoglycemia < 4.0 mmol/L and < 3.0 mmol/L
Elbaz, 2021 [44]	3,605 patients in training set, 2,425 patients in validation set I and 3,635 patients in validation set II	Logistic regression	10 variables	1st week of admission	Hypoglycemia ≤ 70 mg/dL	Internal validation and external validation	AUC in the two validation sets was 0.72 and 0.71
Fitzgerald, 2021 [45]	10,938 patients in training set and 5,172 patients in test set	Boosted tree	Time series data including glucose, nutrition and insulin dosing	2 h	Quantitative glucose value	Internal validation	Mean absolute percentage error estimated 16.5–16.8% with 97% of predictions clinically acceptable
Mathioudakis, 2021 [9]	35,147 patients	Stochastic gradient boosting	43 variables	24 h after each glucose measurement	Hypoglycemia ≤ 70 mg/dL	Internal validation and external validation	Internal validation C-statistic was 0.90 (95% CI: 0.89–0.90) and external validation C-statistic ranged from 0.86 to 0.88
van den Boorn, 2021 [46]	94 patients	First and second derivatives	CGM data	30 min	Quantitative glucose value	Internal validation	Mean squared difference was 7.39 mg/dL
Horton, 2022 [47]	11,847 patients	Logistic regression	41 variables	Impending during ICU stay	Hypoglycemia < 70 mg/dL requiring 50% dextrose within 1 h	External validation	AUC on external validation was 0.79

Table 1 (continued)

Author, year	Number of patients	Modeling technique	Clinical predictors	Prediction horizon	Outcome definition	Validation approach	Model performance
Zale, 2022 [29]	46,142 patients from Hospital 1 (development and internal validation), 23,042, 18,827, 10,204, and 20,519 patients from hospitals 2–5 (external validation) respectively	Random forest	59 variables	Next glucose measurement	Hypoglycemia ≤ 70 mg/dL and hyperglycemia > 180 mg/dL	Internal validation and external validation	External validation set sensitivity ranged from 0.64–0.70, 0.75–0.80, and 0.76–0.78 for controlled, hyperglycemia, and hypoglycemia respectively. External validation set specificity ranged from 0.80–0.87, 0.82–0.84, and 0.87–0.90 for controlled, hyperglycemia, and hypoglycemia respectively

dL, as they were predicting hypoglycemic events based on antecedent mild hypoglycemia. Their model, which included predictor variables of insulin doses, weight, creatinine clearance, and sulfonylurea use, achieved a sensitivity of 71% for a hypoglycemic event < 70 mg/dL. This predictive model was coupled with therapeutic recommendations and, to our knowledge, is the only published model that has been deployed in the EHR as an informatics alert and evaluated prospectively. [48]

Ena et al. also used a logistic regression model that included four predictors: GFR < 30 mL/min/1.73 m², insulin dose > 0.3 units/kg/day, length of stay, and previous episode of hypoglycemia during the 3 months prior to admission [38]. Their training and test sets were 839 patients and 561 patients, respectively, from two different nationwide retrospective cohort studies. The authors achieved a validation set AUC of 0.71 for prediction of a hypoglycemic event < 70 mg/dL during a hospital admission. Stuart et al. utilized logistic regression that included thirteen predictors to predict a hypoglycemic event < 4.0 mmol/L (< 72 mg/dL) during hospital admission. [37] Their model achieved an AUC of 0.733 using a cohort of 9,584 admissions. Shah et al. developed a risk prediction tool based on five variables (age, ED visit 6 months prior to admission, insulin use, use of oral agent that does not induce hypoglycemia, and severe CKD) from 300 patients using logistic regression. Using a risk score cutoff of at least 9, their model achieved a sensitivity of 0.77 and specificity of 0.28.

The next phase in machine learning for prediction of hypoglycemia in hospitalized patients used logistic regression to make predictions using a shorter prediction horizon. In 2018, Mathioudakis et al. predicted hypoglycemia occurring on the next calendar day, and achieved a C-statistic of 0.77 when predicting a hypoglycemic event ≤ 70 mg/dL using a logistic regression model based on 44 predictors from 13,360 hospitalized patients in the training set and 5,902 patients in the test set [39]. Similarly, in 2018, Winterstein et al. published a logistic regression model that predicted a hypoglycemic episode defined as a glucose < 50 mg/dL not followed by a glucose value > 80 mg/dL within 10 min in a prediction horizon of 24 h; their model, which used a large cohort of 21,840 patients and 38 predictor variables, achieved a C-statistic of 0.887 when limited to admission days 3–5. [40]

Since the publication of those two studies in 2018, researchers have employed logistic regression in large datasets to answer different questions about predicting inpatient hypoglycemia. Elbaz et al. asked if they could predict glucose ≤ 70 mg/dL in the first week of a patient's admission [44]. Their dataset included a training set and two validation sets of 3,605, 2,425, and 3,635 patients, respectively. Their externally validated logistic regression achieved AUCs of 0.72 and 0.71 when predicting hypoglycemia during the

first week of a patient's admission. Horton et al. published a logistic regression model that included 41 predictors for impending hypoglycemia during a patient's ICU stay [47]. They trained their model using physiologic data up to 12 h prior to a hypoglycemic episode that required treatment with 50% dextrose and excluded any subsequent episodes of hypoglycemia in that patient's admission. Their training set included 11,847 patients. Upon externally validating their model on the MIMIC-III data, they achieved an AUC of 0.79. Another approach to incorporating both hypoglycemia and hyperglycemia into glucose prediction was presented by Kyi et al. [43] Their model sought to predict persistent dysglycemia (glucose < 72 mg/dL or > 270 mg/dL on two admission days) at the time of admission. Their dataset included 594 patients and ten predictors such as admission dysglycemia, HbA1c \geq 8.1%, sulfonylurea or insulin use, glucocorticoid use, Charlson Comorbidity Index score, and admission days. Their logistic regression model achieved an ROC of 0.806.

Tree-Based Models for Inpatient Glucose Prediction

Model performance vastly improved when more advanced machine learning techniques were employed, since a patient's glucose is not simply linearly related to factors such as antihyperglycemic medication dosing, kidney function, and carbohydrate intake. In 2020, Ruan et al. published an XGBoost model that predicted whether a patient would be hypoglycemic at any time during their admission [8]. Their model, which included 42 clinical covariates and a total sample size of 17,658 patients, achieved an AUROC of 0.96, outperforming a multivariate logistic regression AUROC of 0.75. Seeking to decrease the prediction interval for potentially increased clinical utility, Mathioudakis et al. published in 2021 a stochastic gradient boosting model using 43 static and time-varying clinical covariates that predicted whether or not a patient would have at least one hypoglycemic reading in a rolling window of 24 h from each glucose measurement, which could act as a warning for a provider to adjust a patient's insulin regimen [9]. The total sample size across internal and external validation hospitals was 35,147 patients, and the stochastic gradient boosting model achieved a C-statistic of 0.90 upon internal validation and ranged from 0.86 to 0.88 upon external validation in four different hospitals.

A recent trend in the field of inpatient glucose prediction has been to shorten the prediction horizon while incorporating as much data present in the EHR as possible. One of the shortcomings of predicting hypoglycemia at the admission level is that the only data that can be included in the model are those that are known at the time of admission, which prevents changes in labs, vital signs, and medications from being incorporated. Risk of hypoglycemia is inherently

dynamic within a patient's hospital course and being able to incorporate new data with a shorter prediction horizon allows for more clinically meaningful predictions. Zale et al. published this year a random forest model employing 59 predictors that classified a patient as hypoglycemic, controlled, and hyperglycemic using a large dataset with over 100,000 patients from five hospitals [29]. Unlike earlier glycemic prediction models, this model predicted the glycemic outcome for the next glucose observation at the time of each index glucose measurement, with the interquartile range for time to next glucose measurement of 1.63 to 4.37 h. Four of those hospitals were used for external validation and achieved sensitivities for controlled, hyperglycemic, and hypoglycemic of 0.64–0.70, 0.75–0.80, and 0.76–0.78, respectively; specificities for controlled, hyperglycemic, and hypoglycemic ranged from 0.80 to 0.87, 0.82 to 0.84, and 0.87 to 0.90, respectively, across the external validation hospitals.

Predicting Glucose as a Quantitative Variable

With the increase in dataset sizes, several studies have been published in the last 2 years using CGM data and time series analyses to predict glucose over a shorter prediction horizon and to treat glucose as a quantitative (discrete value) rather than a binary outcome. Kim et al. published a recurrent neural network model to predict a patient's glucose with a prediction horizon of 30 min and achieved an average root mean squared error of 21.5 mg/dL [42]. Their sample size was 20 patients and included 99.4 patient-days of CGM data. Van den Boorn et al. utilized a CGM dataset with a larger sample of ICU patients ($N=94$) that totaled 134,673 glucose measurements. Their model attempted to predict the glucose reading with a prediction horizon of 30 min based on the first and second derivatives of the CGM data, and achieved a mean squared difference of 7.39 mg/dL [46]. The time series approach to predicting glucose in the short-term horizon has not been limited to CGM data, but has also been applied to the MIMIC-III dataset from critical care unit patients. Fitzgerald et al. applied a boosted tree model to this dataset of ICU patients, dividing the patients into a training set of 10,938 patients and a test set of 5,172 patients. The authors achieved 97% clinically acceptable predictions with mean absolute percentage error estimated between 16.5 and 16.8% with a 2-h prediction horizon. [45]

Building upon CGM Prediction

Commercially available CGM devices offer real-time alerting for both hypoglycemia and hyperglycemia and could therefore be considered an alternative strategy to EHR-derived machine learning prediction models in the hospital. Although a CGM-based glucose telemetry system

holds potential as an innovative clinical decision support strategy [49, 50], it is worth noting that real-time alerts for hypoglycemia may not provide an adequate lag time for clinicians to proactively reduce insulin doses (e.g., basal insulin) to prevent the event from occurring in the first place. Certainly, CGM-based systems may provide a safety check for hospital-based clinicians as detection of impending hypoglycemia could be addressed by providing the patient with supplemental carbohydrates; however, we suggest that there is still a need for machine learning–based personalized insulin dosing algorithms that would make recommendations in real time for insulin dosing to achieve desired glycemic targets without hypoglycemia. Integrating CGM with insulin doses and other clinical data from hospitalized patients may allow for improved machine learning models for prediction of insulin doses throughout a patient’s hospitalization. Finally, there are financial and human factor considerations (staff training, time, resources) for broad implementation of CGM devices in hospitalized patients, and cost-effectiveness studies are needed prior to adoption in specific inpatient populations.

Future Directions

Despite the growing body of research in the past decade on machine learning–based tools for glucose prediction in the hospital setting, this field is still in a relatively nascent state, with most prediction models having been validated using retrospective data, but only one published model [51] (developed using a relatively small sample) having been integrated into EHR systems and tested prospectively for clinical effectiveness. Although machine learning holds potential for glycemic prediction in the hospital, it is unknown whether providing hospital-based clinicians more accurate predictions in real time will modify prescribing practices and improve glycemic outcomes. We previously found that there is a high degree of clinical inertia in the hospital even after an overt hypoglycemic episode [52], so the clinical value of machine learning–based glycemic prediction remains to be demonstrated. In the meantime, additional published studies from different healthcare systems using EHR data are needed to refine our understanding of parameters that affect insulin dosing in the hospital so that glycemic predictions can ultimately be coupled with actionable insulin dosing recommendations [53]. Considering that many of the insulin dosing algorithms are proprietary and unpublished, publication of methods and results will allow the scientific community to advance our understanding of factors involved in glucose homeostasis in the hospital and to continually improve the predictive accuracy of insulin dosing algorithms.

Conclusion

Advanced machine learning models using large EHR datasets with large numbers of clinical predictors achieve greater predictive accuracy for glucose than more traditional regression modeling techniques in hospitalized patients. Inferences and comparisons of different machine learning algorithms for inpatient glucose prediction require careful consideration of the study population, outcome and exposure variables, and prediction horizon used to derive the model, as well as the validation approach used to describe model performance. Further prospective studies would be needed to demonstrate the added value of deploying machine learning derived predictive analytic support on improving glycemic management and other outcomes in hospitalized patients.

Declarations

Conflict of Interest The authors declare no competing interests.

References

1. Ghosh S, Manley SE, Nightingale PG, et al. Prevalence of admission plasma glucose in “diabetes” or “at risk” ranges in hospital emergencies with no prior diagnosis of diabetes by gender, age and ethnicity. *Endocrinol Diabetes Metab.* 2020;3(3): e00140. <https://doi.org/10.1002/edm2.140>.
2. Lemieux I, Houde I, Pascot A, et al. Effects of prednisone withdrawal on the new metabolic triad in cyclosporine-treated kidney transplant patients. *Kidney Int.* 2002;62(5):1839–47. <https://doi.org/10.1046/j.1523-1755.2002.00611.x>.
3. Gosmanov AR, Umpierrez GE. Management of hyperglycemia during enteral and parenteral nutrition therapy. *Curr Diab Rep.* 2013;13(1):155–62. <https://doi.org/10.1007/s11892-012-0335-y>.
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(2):101–7. <https://doi.org/10.1185/03007995.2012.754744>.
5. Varlamov EV, Kulaga ME, Khosla A, Prime DL, Rennert NJ. Hypoglycemia in the hospital: systems-based approach to recognition, treatment, and prevention. *Hosp Pract (1995).* 2014;42(4):163–72. <https://doi.org/10.3810/hp.2014.10.1153>.
6. Li DB, Hua Q, Guo J, Li HW, Chen H, Zhao SM. Admission glucose level and in-hospital outcomes in diabetic and non-diabetic patients with ST-elevation acute myocardial infarction. *Intern Med.* 2011;50(21):2471–5. <https://doi.org/10.2169/internalmedicine.50.5750>.
7. Mendez CE, Mok KT, Ata A, Tanenberg RJ, Calles-Escandon J, Umpierrez GE. Increased glycemic variability is independently associated with length of stay and mortality in noncritically ill hospitalized patients. *Diabetes Care.* 2013;36(12):4091–7. <https://doi.org/10.2337/dc12-2430>.
8. Ruan Y, Bellot A, Moysova Z, et al. Predicting the risk of inpatient hypoglycemia with machine learning using electronic health records. *Diabetes Care.* 2020;43(7):1504–11. <https://doi.org/10.2337/dc19-1743>.

9. Mathioudakis NN, Abusamaan MS, Shakarchi AF, et al. Development and validation of a machine learning model to predict near-term risk of iatrogenic hypoglycemia in hospitalized patients. *JAMA Netw Open*. 2021;4(1): e2030913. <https://doi.org/10.1001/jamanetworkopen.2020.30913>.
10. Klonoff DC, Ahn D, Drincic A. Continuous glucose monitoring: a review of the technology and clinical use. *Diabetes Res Clin Pract*. 2017;133:178–92. <https://doi.org/10.1016/j.diabres.2017.08.005>.
11. Wallia A, Umpierrez GE, Rushakoff RJ, et al. Consensus statement on inpatient use of continuous glucose monitoring. *J Diabetes Sci Technol*. 2017;11(5):1036–44. <https://doi.org/10.1177/1932296817706151>.
12. Rodbard D. Continuous glucose monitoring: a review of successes, challenges, and opportunities. *Diabetes Technol Ther*. 2016;18(Suppl 2):S3–13. <https://doi.org/10.1089/dia.2015.0417>.
13. Esteva A, Robicquet A, Ramsundar B, et al. A guide to deep learning in healthcare. *Nat Med*. 2019;25(1):24–9. <https://doi.org/10.1038/s41591-018-0316-z>.
14. Miotto R, Wang F, Wang S, Jiang X, Dudley JT. Deep learning for healthcare: review, opportunities and challenges. *Brief Bioinform*. 2018;19(6):1236–46. <https://doi.org/10.1093/bib/bbx044>.
15. Wu YT, Zhang CJ, Mol BW, et al. Early prediction of gestational diabetes mellitus in the Chinese population via advanced machine learning. *J Clin Endocrinol Metab*. 2021;106(3):e1191–205. <https://doi.org/10.1210/clinem/dgaa899>.
16. Dagliati A, Marini S, Sacchi L, et al. Machine learning methods to predict diabetes complications. *J Diabetes Sci Technol*. 2018;12(2):295–302. <https://doi.org/10.1177/1932296817706375>.
17. Hathaway QA, Roth SM, Pinti MV, et al. Machine-learning to stratify diabetic patients using novel cardiac biomarkers and integrative genomics. *Cardiovasc Diabetol*. 2019;18(1):78. <https://doi.org/10.1186/s12933-019-0879-0>.
18. Broome DT, Hilton CB, Mehta N. Policy Implications of artificial intelligence and machine learning in diabetes management. *Curr Diab Rep*. 2020;20(2):5. <https://doi.org/10.1007/s11892-020-1287-2>.
19. Smith JW, Everhart JE, Dickson W, Knowler WC, Johannes RS. Using the ADAP learning algorithm to forecast the onset of diabetes mellitus. *Am Med Inform Assoc* 1988:261.
20. Zeevi D, Korem T, Zmora N, et al. Personalized nutrition by prediction of glycemic responses. *Cell*. 2015;163(5):1079–94. <https://doi.org/10.1016/j.cell.2015.11.001>.
21. Ting DSW, Cheung CY, Lim G, et al. Development and validation of a deep learning system for diabetic retinopathy and related eye diseases using retinal images from multiethnic populations with diabetes. *JAMA*. 2017;318(22):2211–23. <https://doi.org/10.1001/jama.2017.18152>.
22. El-Khatib FH, Balliro C, Hillard MA, et al. Home use of a bihormonal bionic pancreas versus insulin pump therapy in adults with type 1 diabetes: a multicentre randomised crossover trial. *Lancet*. 2017;389(10067):369–80. [https://doi.org/10.1016/S0140-6736\(16\)32567-3](https://doi.org/10.1016/S0140-6736(16)32567-3).
23. Berner ES. *Clinical decision support systems*. vol 233. Springer; 2007.
24. Chen PC, Liu Y, Peng L. How to develop machine learning models for healthcare. *Nat Mater*. 2019;18(5):410–4. <https://doi.org/10.1038/s41563-019-0345-0>.
25. Clarke WL, Cox D, Gonder-Frederick LA, Carter W, Pohl SL. Evaluating clinical accuracy of systems for self-monitoring of blood glucose. *Diabetes Care*. 1987;10(5):622–8. <https://doi.org/10.2337/diacare.10.5.622>.
26. Mandrekar JN. Receiver operating characteristic curve in diagnostic test assessment. *J Thorac Oncol*. 2010;5(9):1315–6. <https://doi.org/10.1097/JTO.0b013e3181ec173d>.
27. Saito T, Rehmsmeier M. The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. *PLoS ONE*. 2015;10(3): e0118432. <https://doi.org/10.1371/journal.pone.0118432>.
28. Brenner H, Gefeller O. Variation of sensitivity, specificity, likelihood ratios and predictive values with disease prevalence. *Stat Med*. 1997;16(9):981–91.
29. Zale AD, Abusamaan MS, McGready J, Mathioudakis N. Development and validation of a machine learning model for classification of next glucose measurement in hospitalized patients. *EClinicalMedicine*. 2022;44: 101290. <https://doi.org/10.1016/j.eclinm.2022.101290>.
30. James G, Witten D, Hastie T, Tibshirani R. *An introduction to statistical learning*. vol 112. Springer; 2013.
31. Emmert-Streib F, Dehmer M. Evaluation of regression models: model assessment, model selection and generalization error. *Machine Learn Knowledge Extract*. 2019;1(1):521–51.
32. Friedman JH. Stochastic gradient boosting. *Comput Stat Data Anal*. 2002;38(4):367–78.
33. Bleeker S, Moll H, Steyerberg EA, et al. External validation is necessary in prediction research: a clinical example. *J Clin Epidemiol*. 2003;56(9):826–32.
34. Roberts DR, Bahn V, Ciuti S, et al. Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. *Ecography*. 2017;40(8):913–29.
35. Siwicki B. An Epic cognitive computing platform primer. <https://www.healthcareitnews.com/news/epic-cognitive-computing-platform-primer>
36. Elliott MB, Schafers SJ, McGill JB, Tobin GS. Prediction and prevention of treatment-related inpatient hypoglycemia. *J Diabetes Sci Technol*. 2012;6(2):302–9. <https://doi.org/10.1177/193229681200600213>.
37. 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(10):1385–91. <https://doi.org/10.1111/dme.13409>.
38. Ena J, Gaviria AZ, Romero-Sánchez M, et al. Derivation and validation model for hospital hypoglycemia. *Eur J Intern Med*. 2018;47:43–8. <https://doi.org/10.1016/j.ejim.2017.08.024>.
39. 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(1): e000499. <https://doi.org/10.1136/bmjdr-2017-000499>.
40. Winterstein AG, Jeon N, Staley B, Xu D, Henriksen C, Lipori GP. Development and validation of an automated algorithm for identifying patients at high risk for drug-induced hypoglycemia. *Am J Health Syst Pharm*. 2018;75(21):1714–28. <https://doi.org/10.2146/ajhp180071>.
41. Shah BR, Walji S, Kiss A, James JE, Lowe JM. Derivation and validation of a risk-prediction tool for hypoglycemia in hospitalized adults with diabetes: the hypoglycemia during hospitalization (HyDHo) score. *Can J Diabetes*. 2019;43(4):278–282. e1. <https://doi.org/10.1016/j.jcjd.2018.08.061>.
42. Kim DY, Choi DS, Kim J, et al. Developing an individual glucose prediction model using recurrent neural network. *Sensors (Basel)*. 2020;20(22). doi:<https://doi.org/10.3390/s20226460>
43. Kyi M, Gorelik A, Reid J, et al. Clinical prediction tool to identify adults with type 2 diabetes at risk for persistent adverse glycemia in hospital. *Can J Diabetes*. 2020. <https://doi.org/10.1016/j.jcjd.2020.06.006>.
44. Elbaz M, Nashashibi J, Kushnir S, Leibovici L. Predicting hypoglycemia in hospitalized patients with diabetes: a derivation and validation study. *Diabetes Res Clin Pract*. 2021;171: 108611. <https://doi.org/10.1016/j.diabres.2020.108611>.
45. Fitzgerald O, Perez-Concha O, Gallego B, et al. Incorporating real-world evidence into the development of patient blood glucose prediction algorithms for the ICU. *J Am Med Inform*

- Assoc. 2021;28(8):1642–50. <https://doi.org/10.1093/jamia/ocab060>.
46. van den Boorn M, Lagerburg V, van Steen SCJ, Wedzinga R, Bosman RJ, van der Voort PHJ. The development of a glucose prediction model in critically ill patients. *Comput Methods Programs Biomed.* 2021;206: 106105. <https://doi.org/10.1016/j.cmpb.2021.106105>.
 47. Horton WB, Barros AJ, Andris RT, Clark MT, Moorman JR. Pathophysiologic signature of impending ICU hypoglycemia in bedside monitoring and electronic health record data: model development and external validation. *Crit Care Med.* 2022;50(3):e221–30. <https://doi.org/10.1097/CCM.00000000000005171>.
 48. Cruz P, Blackburn MC, Tobin GS. A systematic approach for the prevention and reduction of hypoglycemia in hospitalized patients. *Curr Diab Rep.* 2017;17(11):117. <https://doi.org/10.1007/s11892-017-0934-8>.
 49. Singh LG, Satyarengga M, Marcano I, et al. Reducing inpatient hypoglycemia in the general wards using real-time continuous glucose monitoring: the Glucose Telemetry System, a randomized clinical trial. *Diabetes Care.* 2020. <https://doi.org/10.2337/dc20-0840>.
 50. Spanakis EK, Levitt DL, Siddiqui T, et al. The effect of continuous glucose monitoring in preventing inpatient hypoglycemia in general wards: the Glucose Telemetry System. *J Diabetes Sci Technol.* 2018;12(1):20–5. <https://doi.org/10.1177/1932296817748964>.
 51. Kilpatrick CR, Elliott MB, Pratt E, et al. Prevention of inpatient hypoglycemia with a real-time informatics alert. *J Hosp Med.* 2014;9(10):621–6. <https://doi.org/10.1002/jhm.2221>.
 52. Mathioudakis N, Everett E, Golden SH. Prevention and management of insulin-associated hypoglycemia in hospitalized patients. *Endocr Pract.* 2016;22(8):959–69. <https://doi.org/10.4158/EP151119.OR>.
 53. Nguyen M, Jankovic I, Kalesinskas L, Baiocchi M, Chen JH. Machine learning for initial insulin estimation in hospitalized patients. *J Am Med Inform Assoc.* 2021. <https://doi.org/10.1093/jamia/ocab099>.

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