Original Paper

Toward a Clinically Actionable, Electronic Health Record–Based Machine Learning Model to Forecast 90-Day Change in Hemoglobin A_{1c} in Youth With Type 1 Diabetes: Feasibility and Model Development Study

Erin M Tallon[1,2,3*], PhD, RN; David D Williams[4*], MPH; Cintya Schweisberger[1,2], DO; Colin Mullaney[5], BS; Brent Lockee[1], MS; Diana Ferro[6], MSc, PhD; Craig A Vandervelden[1], PhD; Mitchell S Barnes[1], BS; Angelica Cristello Sarteau[7], MSPH, PhD; Anna R Kahkoska[7], MD, PhD; Susana R Patton[8], PhD, CDCES; Sanjeev Mehta[9], MD, MPH; Ryan McDonough[1,2,10], DO; Marcus Lind[11,12,13], MD, PhD; Leonard D'Avolio[5,14,15], PhD; Mark A Clements[1,2], MD, PhD

[1]Division of Pediatric Endocrinology and Diabetes, Children's Mercy Kansas City, Kansas City, MO, United States

[2]Department of Pediatrics, UMKC School of Medicine, Kansas City, MO, United States

[3]Institute for Data Science and Informatics, University of Missouri-Columbia, Columbia, MO, United States

[4]Division of Health Services and Outcomes Research, Children's Mercy Kansas City, Kansas City, MO, United States

[5]Blue Circle Health, Boston, MA, United States

[6]Preventive and Predictive Medicine, IRCCS, Bambino Gesù Children's Hospital, Rome, Italy

[7]Department of Nutrition, University of North Carolina at Chapel Hill, Chapel Hill, NC, United States

[8]Center for Healthcare Delivery Science, Nemours Children's Health, Jacksonville, FL, United States

[9]Joslin Diabetes Center, Boston, MA, United States

[10]Department of Pediatrics, Arkansas Children's Northwest, Springdale, AR, United States

[11]Department of Medicine, NU-Hospital Group, Uddevalla, Sweden

[12]Department of Molecular and Clinical Medicine, University of Gothenburg, Gothenburg, Sweden

[13]Department of Medicine, Sahlgrenska University Hospital, Gothenburg, Sweden

[14]Harvard Medical School, Boston, MA, United States

[15]Mass General Brigham, Boston, MA, United States

*these authors contributed equally

Corresponding Author:

Erin M Tallon, PhD, RN
Division of Pediatric Endocrinology and Diabetes
Children's Mercy Kansas City
2401 Gillham Road
Kansas City, MO
United States
Phone: 1 8166014023
Email: etallon@cmh.edu

Abstract

Background: Clinicians currently lack an effective means for identifying youth with type 1 diabetes (T1D) who are at risk for experiencing glycemic deterioration between diabetes clinic visits. As a result, their ability to identify youth who may optimally benefit from targeted interventions designed to address rising glycemic levels is limited. Although electronic health records (EHR)–based risk predictions have been used to forecast health outcomes in T1D, no study has investigated the potential for using EHR data to identify youth with T1D who will experience a clinically significant rise in glycated hemoglobin (HbA_{1c}) ≥0.3% (approximately 3 mmol/mol) between diabetes clinic visits.

Objective: We aimed to evaluate the feasibility of using routinely collected EHR data to develop a machine learning model to predict 90-day unit-change in HbA_{1c} (in % units) in youth (aged 9-18 y) with T1D. We assessed our model's ability to augment clinical decision-making by identifying a percent change cut point that optimized identification of youth who would experience a clinically significant rise in HbA_{1c}.

Methods: From a cohort of 2757 youth with T1D who received care from a network of pediatric diabetes clinics in the Midwestern United States (January 2012-August 2017), we identified 1743 youth with 9643 HbA_{1c} observation windows (ie, 2 HbA_{1c} measurements separated by 70-110 d, approximating the 90-day time interval between routine diabetes clinic visits). We used up to 5 years of youths' longitudinal EHR data to transform 17,466 features (demographics, laboratory results, vital signs, anthropometric measures, medications, diagnosis codes, procedure codes, and free-text data) for model training. We performed 3-fold cross-validation to train random forest regression models to predict 90-day unit-change in $HbA_{1c}(\%)$.

Results: Across all 3 folds of our cross-validation model, the average root-mean-square error was 0.88 (95% CI 0.85-0.90). Predicted $HbA_{1c}(\%)$ strongly correlated with true $HbA_{1c}(\%)$ (r=0.79; 95% CI 0.78-0.80). The top 10 features impacting model predictions included postal code, various metrics related to HbA_{1c}, and the frequency of a diagnosis code indicating difficulty with treatment engagement. At a clinically significant percent rise threshold of ≥0.3% (approximately 3 mmol/mol), our model's positive predictive value was 60.3%, indicating a 1.5-fold enrichment (relative to the observed frequency that youth experienced this outcome [3928/9643, 40.7%]). Model sensitivity and positive predictive value improved when thresholds for clinical significance included smaller changes in HbA_{1c}, whereas specificity and negative predictive value improved when thresholds required larger changes in HbA_{1c}.

Conclusions: Routinely collected EHR data can be used to create an ML model for predicting unit-change in HbA_{1c} between diabetes clinic visits among youth with T1D. Future work will focus on optimizing model performance and validating the model in additional cohorts and in other diabetes clinics.

Introduction

Background

Type 1 diabetes (T1D), an immune-mediated chronic disease that affects more than 1 in 300 youth in the United States, is characterized by significant to near-total loss of endogenous insulin production [1,2]. Given insulin's critical role in maintaining glucose homeostasis, the most immediate and pervasive downstream effect of insulin deficiency is persistent, life-threatening hyperglycemia that must be identified through frequent glucose monitoring and managed with lifelong administration of exogenous insulin [1].

Youth with T1D attend routine (often quarterly) diabetes clinic visits where clinicians use glycated hemoglobin (HbA_{1c}) testing to assess glycemic status [3,4]. Considered the gold standard for monitoring long-term glycemia in diabetes, HbA_{1c} testing provides an objective measure of an individual's mean blood glucose during the previous 2-3 months [3,5]. To achieve glycemic goals, youth with T1D are increasingly being encouraged to adopt sophisticated diabetes technologies, such as hybrid closed-loop insulin pumps and continuous glucose monitoring (CGM) systems [6,7]. Concurrent with the rising availability of these technologies and a strong research base linking HbA_{1c} with the development of diabetes complications, the American Diabetes Association has incrementally lowered its recommended HbA_{1c} goals for youth with diabetes [4,8].

Despite increased adoption of advanced diabetes technologies over time, data from the T1D Exchange indicated that between 2010-2012 and 2016-2018, mean HbA_{1c} in US youth with T1D rose from 7.8% (62 mmol/mol) to 8.4% (68 mmol/mol); and in 2016-2018, only 16% (686/4346) of youth were meeting the American Diabetes Association's (then)

recommended HbA_{1c} goal of <7.5% (<58 mmol/mol) [6]. A separate analysis of 2015-2016 data indicated that fewer than 20% (1817/9685) of US youth with T1D less than the age of 18 years had an HbA_{1c}<7.5% (58 mmol/mol); and fewer than 10% (690/9685) of youth had an HbA_{1c}<7% (53 mmol/mol) [9]. Previous research has shown that 1 in 5 youth with T1D experience an increasing HbA_{1c} trajectory between the ages of 8 and 18 years [10].

Through a phenomenon known as "metabolic memory," periods of hyperglycemia are known to increase risk for diabetes-related microvascular and macrovascular complications for >10 years following initial exposure [11]. A similar—but beneficial—legacy effect is observed in individuals with T1D who are exposed to near-normal glycemia and later experience more favorable long-term diabetes outcomes, even when glycemic levels later rise [11,12]. These findings point to a critical need to optimize the early identification of youth who are candidates for targeted interventions to improve deteriorating glycemia.

The increasing availability of real-world clinical data housed in electronic health records (EHR) is generating opportunities to investigate population-level health outcomes, develop classification and risk prediction models to augment clinical decision-making, and accelerate diagnostic and therapeutic discovery [13-15]. Machine learning (ML) has been used to meaningfully advance understanding of numerous clinical outcomes in individuals with diabetes [16-18], and EHR-based risk predictions have been leveraged to generate insights across the health-disease spectrum, including T1D [19-21].

Given the multifactorial etiology of rising glycemic levels in youth with T1D, it remains difficult to identify youth who are at the highest risk of experiencing increased HbA_{1c} between routine diabetes clinic visits. To date, no study has

investigated the feasibility of or potential for using EHR data to develop a predictive model to identify youth with T1D who will experience a clinically significant rise in HbA_{1c} between clinic visits. Such a model could augment clinical decision-making and facilitate initiation of interventions that increase behaviors known to improve glycemia in high-risk youth.

Objective

We sought to evaluate the feasibility of using ML to identify youth (aged 9-18 y) with T1D who were candidates for behavioral and care delivery interventions designed to reduce or prevent a predicted rise in HbA_{1c}. To do so, we used routinely collected EHR data to develop an interpretable and clinically actionable ML model to forecast unit-change (ie, increase or decrease, in % units) in HbA_{1c} in 90 days. We then evaluated the ability of our model to augment clinical decision-making by identifying a percent-change cut point that optimized identification of youth who experienced a clinically significant rise in HbA_{1c} at their subsequent diabetes clinic encounter.

Methods

Study Design

We applied the random forest (RF) regression algorithm to longitudinal EHR data to develop a model to forecast 90-day unit-change in HbA_{1c} (in % units). We used RF due to its utility for constructing accurate, noise-resilient ML models from high-dimensional data [22,23]. To evaluate our model's ability to identify youth who, based on predicted rise in HbA_{1c}, were true candidates for intervention, we evaluated the sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) of predicted versus actual change in HbA_{1c} at several cut points: ≥0.3%, ≥0.4%, ≥0.5%, and ≥0.6% (approximately 3 mmol/mol, 4 mmol/mol, 5 mmol/mol, and 7 mmol/mol, respectively).

Source Data and Study Cohort

Using data extracted from Oracle Health EHR (formerly Cerner Millenium Electronic Medical Record System; Nashville, Tennessee) [24], we used diagnosis code and laboratory data to identify a cohort of 2757 youth with T1D who received care from a network of pediatric diabetes clinics in the Midwestern United States between January 2012 and August 2017. Criteria used to identify this T1D cohort are provided in Multimedia Appendix 1.

HbA_{1c} Measurements and Observation Windows

For youth with T1D, we identified health encounters that were associated with HbA_{1c} measurements (ie, laboratory and point-of-care HbA_{1c} measurements) and HbA_{1c} observation windows that met inclusion criteria. Each HbA_{1c} observation window comprised 2 documented HbA_{1c} measurements (from a single individual) separated by a time interval of 70-110 days. The 70- to 110-day time interval was selected to approximate the 3-month (ie, 90-day) time interval between regularly scheduled diabetes clinic visits.

Certain encounters with HbA_{1c} data were excluded from consideration and therefore not included in any HbA_{1c} observation windows. HbA_{1c} values documented at or shortly after T1D diagnosis tend to be more extreme than those documented at subsequent time points (ie, after an individual with T1D begins receiving regular insulin injections) [25, 26]. As such, each youth's first-documented encounter with an HbA_{1c} value was excluded under the assumption that a youth's first HbA_{1c} measurement may have been obtained at the time of T1D diagnosis. We also excluded data from encounters where youth were <9 years old, as the incidence of clinically significant rise in HbA_{1c} is less common in this age group [6,27].

We excluded observation windows associated with HbA_{1c} measurements that were separated by <70 days or >110 days, as well as those where the first encounter for a given HbA_{1c} observation window (ie, the index encounter) was associated with an HbA_{1c} of >12% (>108 mmol/mol). The latter exclusion criterion was used because individuals with an HbA_{1c} of >12% (>108 mmol/mol) were already considered ideal candidates for intervention. Encounter-level data from all HbA_{1c} observation windows that met inclusion criteria were included in our final dataset, which could include data from multiple HbA_{1c} observation windows per individual.

Outcome Definition

The forecasted outcome was unit-change in HbA_{1c} (in % units) at the end of 90 days. After predicting each youth's percent change in HbA_{1c} in 90 days (ie, at the time of the follow-up encounter), we used various thresholds to determine an HbA_{1c} percent rise cut point that optimized identification of individuals who were true candidates for intervention at the time of their index encounter: ≥0.3%, ≥0.4%, ≥0.5%, and ≥0.6% (approximately 3 mmol/mol, 4 mmol/mol, 5 mmol/mol, and 7 mmol/mol, respectively). We considered these cut points to be clinically relevant and actionable, given that a long-term decrease of ≥0.3% (3 mmol/mol) in HbA_{1c} is associated with reduced risk of long-term diabetes complications [28].

Data Extraction

We used SQL queries to comprehensively extract up to approximately 5 years (January 2012-August 2017) of structured and unstructured EHR data for each youth with index and follow-up encounter data for at least 1 qualifying HbA_{1c} observation window. These data included demographics, laboratory results, vital signs, anthropometric measures, encounter locations, medications, diagnosis codes, procedure codes, structured clinical vocabulary codes, and free-text data from diabetes- and non–diabetes-related clinical notes, messages, and reports.

Demographic data included sex (female, male), age, ethnicity (non-Hispanic, Hispanic), race (White, Black or African American, Asian, American Indian or Alaska Native, Native Hawaiian or Pacific Islander, and other), primary language (eg, English or Spanish), health plan type; and

postal code (3- and 4-digit postal code prefixes). Additional extracted data included up to approximately 5 years of all available laboratory test results, clinical event and observation data, vital signs (heart rate, respiratory rate, oxygen saturation, and blood pressure), anthropometric measures (weight, height, and BMI), and medications (mapped to standard generic drug names [29]). We also extracted diagnosis codes (ie, *ICD-9* [*International Classification of Diseases, Ninth Revision*], ICD-10 [*International Statistical Classification of Diseases, Tenth Revision*], and Systematized Nomenclature of Medicine Clinical Terms [SNOMED CT] codes), procedure codes (ie, Current Procedural Terminology [CPT] codes); and other structured clinical vocabulary codes (ie, SNOMED CT).

We chose not to include data generated by diabetes devices (eg, automated insulin delivery and CGM systems). Early on, we observed that HbA_{1c} was easiest to predict in youth who used diabetes devices that generate diabetes data (eg, glucose levels) in real time. However, since most diabetes centers do not have broad or ready access to device data in near-real time, we sought to evaluate the potential of using only EHR data to predict HbA_{1c}.

Feature Engineering

We engineered features using data documented during all available historical encounters, as well as during HbA_{1c} observation window index and follow-up encounters. Processes used to transform variables into features for ML varied by data type. In all, our feature engineering processes generated 17,466 input features for model fitting.

Numeric Variables

For numeric variables (eg, laboratory results, weight, and vital signs), we created features by calculating summary metrics (ie, mean, slope, and SD). In general, we created 2 sets of features for each numeric variable, based on proximity of the measurements to the HbA_{1c} observation window's index encounter. One set of features was created using data documented during the 12 months preceding (and at) the index encounter. A second set was created using all available EHR data documented before (and at) the index encounter. For example, we created 2 features for mean HbA_{1c}: one calculated using the previous 12 months of HbA_{1c} data (up to and including the index encounter), and the other calculated using all available HbA_{1c} data (up to and including the index encounter). Given the intrinsic insensitivity of RF to numerical outliers, we did not alter or drop outliers from the data. Once all numerical features were created, missing numerical values were imputed using the population median.

Each youth's diagnostic (ie, first) HbA_{1c} result was included as a separate feature, as was the HbA_{1c} result documented at the observation window's index encounter. Because research suggests that youth with T1D can be grouped into one of several HbA_{1c} trajectory clusters [10], we created an HbA_{1c} trajectory feature by using k-means clustering [30] to assign youth to 1 of 4 clusters based on their quarterly HbA_{1c} measurements.

Categorical Variables

We used data documented at the observation window's index encounter to create features from demographic data (eg, age, race, ethnicity, primary language, health plan type, and postal code). For each categorical demographic variable, we used the StringIndexer feature transformer to convert the categories associated with each variable into numeric indices, thus creating a single feature for each of these variables [31].

We used Clinical Classification Software Revised (CCSR), developed by the Agency for Healthcare Research and Quality, to group *ICD-10* codes into meaningful categories [32]. Thereafter, each CCSR category and each *ICD-9*, *ICD-10*, SNOMED CT, and CPT code was treated as a separate variable. We created 2 sets of features for each of these separate variables, based on how many times each had been assigned to the individual relative to the observation window's index encounter. One set of features was created by calculating the frequency that each had been assigned to the individual during the 12 months preceding (and at) the index encounter. The second set was created using all available EHR data documented before (and at) the index encounter. Absence of diagnosis, procedure, or structured clinical vocabulary codes was presumed to reflect true absence, rather than missingness, of these data variables.

Medication variables were similarly transformed into 2 sets of features based on how often each medication had been prescribed relative to the index encounter. One set of features was created by calculating the frequency that each medication had been prescribed to the individual during the 12 months preceding (and at) the index encounter. The second set was created using all available medication data documented before (and at) the index encounter. Encounter frequencies were similarly calculated and included as separate features. Absence of medication and encounter data was presumed to reflect true absence of these data.

Natural Language Processing

We used term frequency–inverse document frequency (TF-IDF) vectorization, a natural language processing technique, to process free-text data from clinical notes, messages, and reports. In TF-IDF vectorization, words (ie, tokens) are first converted into a matrix of token counts [33]. The matrix is then transformed into a normalized TF-IDF representation that most heavily weights tokens that occur infrequently across the entire corpus of available text [33]. As such, TF-IDF is used to assign the highest weight to words that have the most discriminating power. After ranking by weight, we constrained the total number of features generated via TF-IDF vectorization to 250 single-word terms and 250 two-word terms, each of which had to be present in at least 5 documents.

Model Development and Evaluation

RF uses bootstrap aggregation and random feature sampling to independently train a series of uncorrelated decision tree regressors, known as "weak learners" [22,23,34-36]. Predictions from this ensemble of weak learners are averaged

to produce a single "strong learner" with improved prediction accuracy [23]. Relative to many other ML methods, the RF algorithm presents several key advantages, including decreased risk of overfitting, straightforward calculation of the degree to which individual input features contribute to model predictions, and robustness to missing data [22,34].

After randomly splitting the entire dataset into 3 nonoverlapping data subsets, we used 3-fold cross-validation to recursively fit RF regressors to 2 of the 3 subsets and then evaluate model performance on the third, held-out subset. We used 3-fold (rather than 5- or 10-fold) cross-validation due to the large number of HbA_{1c} observation windows included in our analysis, as well as our desire to reduce variance in the estimated performance of our model. Hyperparameters used for model fitting are presented in Table 1. Model performance was evaluated by averaging the mean absolute error (MAE) and the root-mean-square error (RMSE)—the SD of the residuals [37]—across all 3 cross-validation models.

Table 1. Hyperparameter values used for random forest regressor model training. A complete list of hyperparameter keys accepted by the random forest regressor algorithm and definitions of each can be found on the web [38]. Hyperparameters not listed below were set to default values.

Hyperparameter	Value used	Default value
NumTrees	40	20
MaxDepth	7	5
MaxBins	128	32
MinInstancesPerNode	8	1
FeatureSubsetStrategy	"onethird"	"onethird"
Impurity	"variance"	"variance"
MinInfoGain	0.0	0.0
MinWeightFractionPerNode	0.0	0.0
SubsamplingRate	1.0	1.0

Decision tree regressors are grown by recursively splitting on features to maximize impurity reduction [39]. Feature splits that reduce impurity by maximally reducing variance are considered important; thus, the features that are split to maximize reduction in variance are also deemed important [39,40]. We evaluated feature importance by calculating and ranking the mean reduction in variance associated with only those features that were used by all 3 of our cross-validation models to forecast HbA_{1c}.

We used Python (version 3) and Scala (version 2; Programming Methods Laboratory at École Polytechnique Fédérale de Lausanne) to clean and transform the data. ML analyses were conducted using the Apache Spark MLlib (version 2) ML library [41].

Statistical Analysis

Pearson r correlations were used to assess the strength and direction of the relationship between actual and predicted HbA_{1c} values. We also used sensitivity, specificity, PPV, and NPV as clinical performance metrics to aid in identifying a predicted HbA_{1c} percent rise threshold that would facilitate optimal capture of youth who would experience a clinically significant rise in HbA_{1c} in 90 days.

Summary statistics, correlations, RMSE, MAE, and sensitivity, specificity, PPV, and NPV metrics were assessed using Stata/SE (Stata standard edition) software (version 18.5; StataCorp) [42].

Ethical Considerations

Clinical and model output data were collected and coded in an institutional review board–approved research data repository at Children's Mercy Kansas City (Kansas City, Missouri; IRB #11120355) that met the requirements for a waiver of written informed consent as outlined in 45 CFR 46.116.

Results

Overview

Out of 2757 youth with T1D, 1743 youth (63.2%) had one or more HbA_{1c} observation windows (n=9643) that met inclusion criteria (Figure 1).

Figure 1. Flowchart depicting inclusion and exclusion criteria for the study cohort and for glycated hemoglobin observation windows. Abbreviations: HbA_{1c}: glycated hemoglobin; T1D: type 1 diabetes.
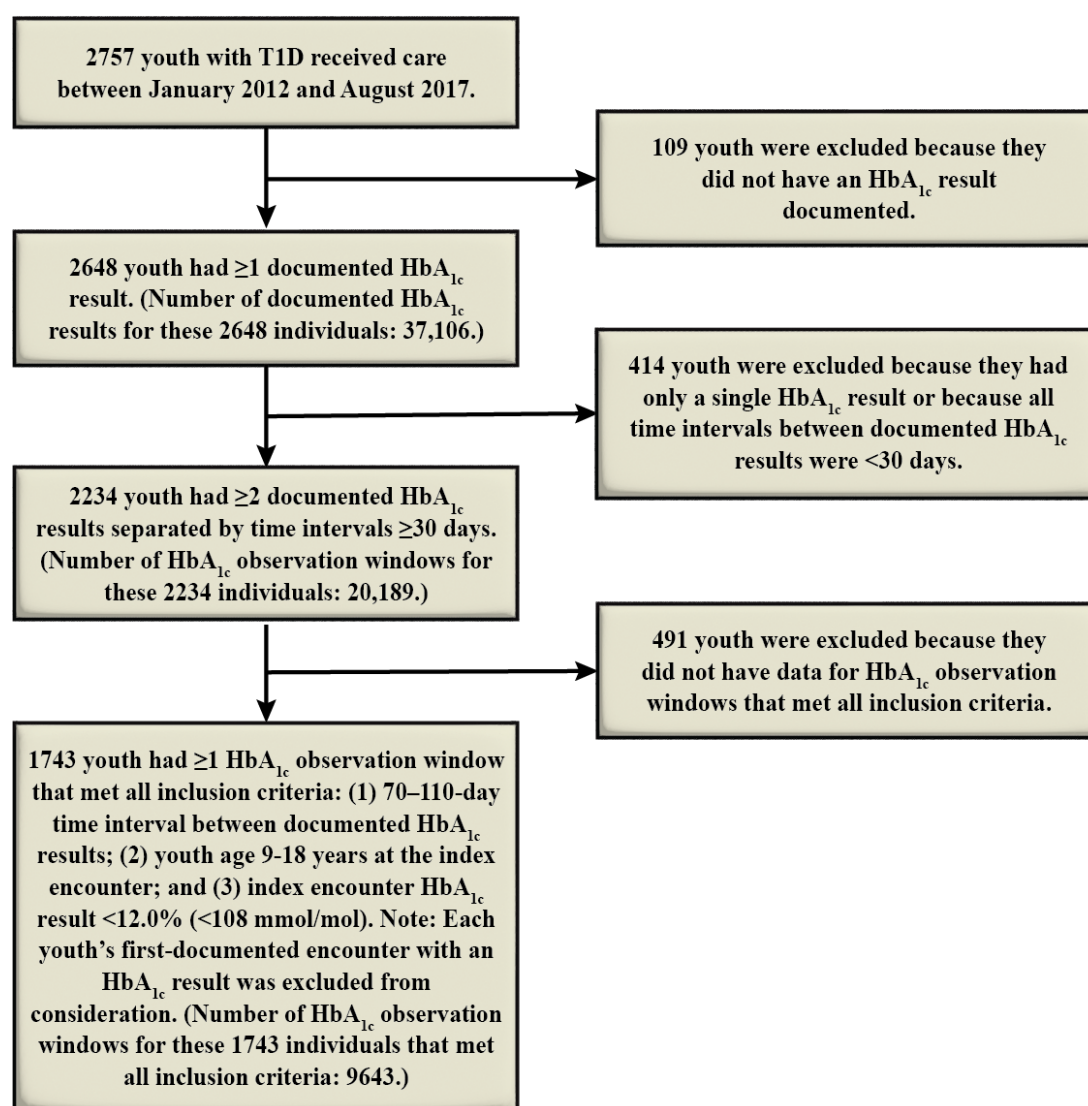
Characteristics of the entire cohort that met inclusion criteria are summarized in Table 2. The observed frequencies that youth experienced a rise in HbA_{1c} that exceeded each percent change cut points ($\geq 0.3\%$, $\geq 0.4\%$, $\geq 0.5\%$, and $\geq 0.6\%$ [approximately 3 mmol/mol, 4 mmol/mol, 5 mmol/mol, 7 mmol/mol]) were 40.7%, 35.6%, 30.8%, and 26.5%, respectively. Characteristics of observations included in each nonoverlapping K-fold are summarized in Table 3.

Table 2. Demographic and clinical characteristics of 1743 youth with glycated hemoglobin observation windows that met inclusion criteria.

Demographic and clinical characteristics	All HbA_{1c}[a] observation windows (n=9643)	Index encounter of each youth's first HbA_{1c} observation window (n=1743)
Age (y), mean (SD)	13.8 (2.6)	12.9 (2.7)
Sex, n (%)		
Female	4599 (47.7)	844 (48.4)
Male	5044 (52.3)	899 (51.6)
Unknown	0 (0)	0 (0)
Race, n (%)		
White	8196 (85)	1449 (83.1)
Black or African American	616 (6.4)	133 (7.6)
Asian	53 (0.5)	12 (0.7)
American Indian or Alaska Native	42 (0.4)	8 (0.5)

Demographic and clinical characteristics	All HbA$_{1c}$[a] observation windows (n=9643)	Index encounter of each youth's first HbA$_{1c}$ observation window (n=1743)
Native Hawaiian or Pacific Islander	8 (0.1)	3 (0.2)
Other	63 (0.7)	10 (0.6)
Unknown	665 (6.9)	128 (7.3)
Ethnicity, n (%)		
Non-Hispanic or non-Latino	8978 (93.1)	1620 (93.0)
Hispanic or Latino	656 (6.8)	121 (6.9)
Unknown	9 (0.1)	2 (0.1)
HbA$_{1c}$ at index encounter (%), mean (SD)	8.6 (1.3)	8.5 (1.5)
HbA$_{1c}$ at index encounter (mmol/mol), mean (SD)	70 (14.2)	69 (16.4)
Change in HbA$_{1c}$[b] (%), median (IQR)	0.1 (–0.4 to 0.6)	0.1 (–0.4 to 0.7)
Change in HbA$_{1c}$[b] (mmol/mol), median (IQR)	1 (–4 to 7)	1 (–4 to 8)
HbA$_{1c}$ increase, n (%)		
≥0.3%	3928 (40.7)	763 (43.8)
≥0.4%	3435 (35.6)	662 (38)
≥0.5%	2966 (30.8)	580 (33.3)
≥0.6%	2552 (26.5)	498 (28.6)

[a]HbA$_{1c}$: glycated hemoglobin.
[b]Change in HbA$_{1c}$: (HbA$_{1c}$ at the observation window's follow-up encounter)–(HbA$_{1c}$ at the observation window's index encounter).

Table 3. Demographic and clinical characteristics of youth with glycated hemoglobin observation windows included in each K-fold subset.

| Demographic and clinical characteristics	HbA$_{1c}$[a] observation windows: fold 1 (n=3151)	HbA$_{1c}$ observation windows: fold 2 (n=3129)	HbA$_{1c}$ observation windows: fold 3 (n=3363)
Youth, n (%)	1291 (41.0)	1288 (41.2)	1381 (41.1)
Age (y), mean (SD)	13.9 (2.6)	13.8 (2.6)	13.8 (2.6)
Sex, n (%)			
Female	1534 (48.7)	1488 (47.6)	1577 (46.9)
Male	1617 (51.3)	1641 (52.4)	1786 (53.1)
Unknown	0 (0)	0 (0)	0 (0)
Race, n (%)			
White	2690 (85.4)	2658 (85.0)	2848 (84.7)
Black or African American	174 (5.6)	206 (6.6)	236 (7.0)
Asian	14 (0.4)	16 (0.5)	23 (0.7)
American Indian or Alaska Native	17 (0.5)	13 (0.4)	12 (0.4)
Native Hawaiian or Pacific Islander	3 (0.1)	3 (0.1)	2 (0.1)
Other	17 (0.5)	21 (0.6)	25 (0.7)
Unknown	236 (7.5)	212 (6.8)	217 (6.4)
Ethnicity, n (%)			
Non-Hispanic or non-Latino	2911 (92.4)	2928 (93.6)	3139 (93.3)
Hispanic or Latino	236 (7.5)	197 (6.3)	223 (6.6)
Unknown	4 (0.1)	4 (0.1)	1 (0.1)
HbA$_{1c}$[a] at index encounter (%), mean (SD)	8.6 (1.3)	8.6 (1.3)	8.6 (1.3)
HbA$_{1c}$ at index encounter (mmol/mol), mean (SD)	70 (14)	70 (14)	70 (14)
HbA$_{1c}$ increase ≥0.3%, n (%)	1255 (39.8)	1293 (41.3)	1380 (41)

[a]HbA$_{1c}$: glycated hemoglobin.

Model Performance

Across all 3 folds of our cross-validation model, average RMSE was 0.88 (Figure 2). Thus, in 68% (6557/9643) of cases (representing one SD), our predictions were within ±0.88% (95% CI 0.85-0.90) of the true percent change in HbA$_{1c}$. The average MAE across all 3 folds was 0.64 (95%

CI 0.63-0.65). Predicted HbA_{1c}(%) strongly correlated with true HbA_{1c}(%; r=0.79; 95% CI 0.78-0.80).

Figure 2. Distribution of the prediction error (ie, residuals) across all 3 cross-validation K-folds. Root-mean-square error is equal to the SD of the prediction error. RMSE: root-mean-square error.
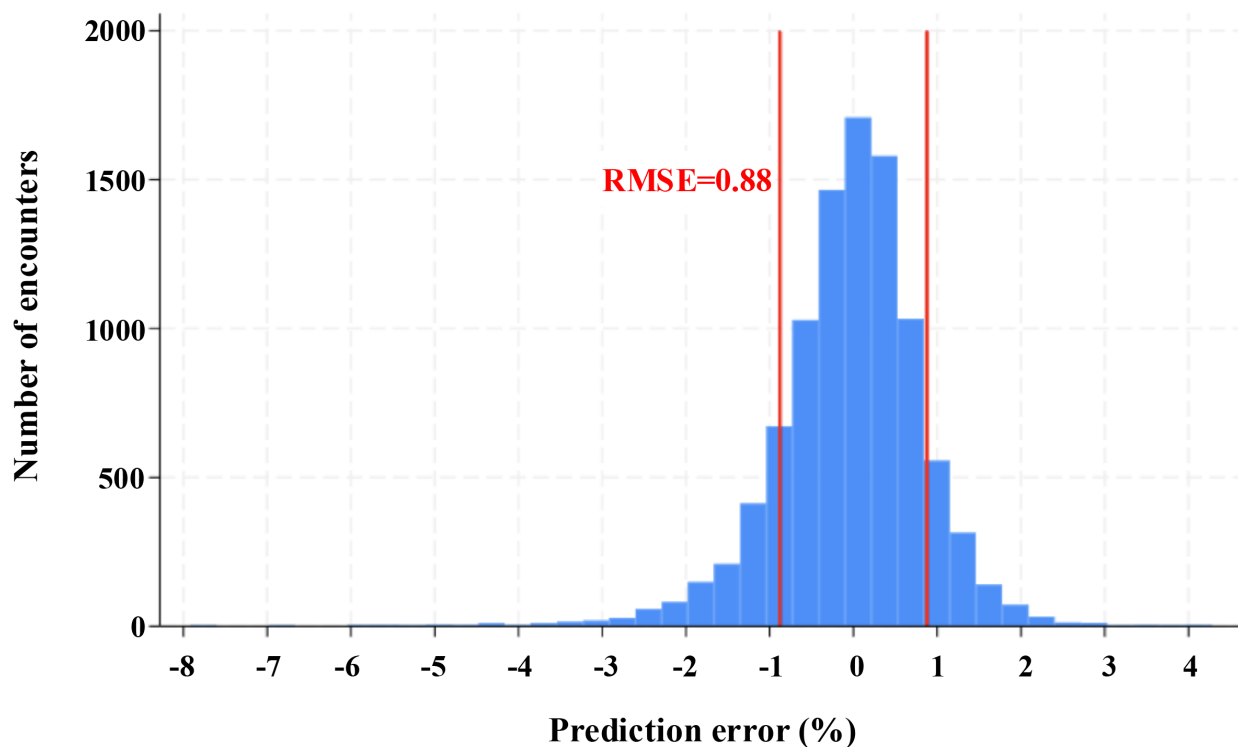

Feature Importance

Across all 3 folds of our cross-validation model, the top 10 features identified as having the greatest impact on model predictions included postal code, various metrics related to HbA_{1c}, and the number of times that the individual had been assigned a diagnosis code indicating difficulty with treatment engagement (Figure 3). The top 30 most important features used to predict percent change in HbA_{1c} are in Multimedia Appendix 2.

Figure 3. Top 10 most important features for predicting 90-day percent change in glycated hemoglobin, assessed via gain-based feature importance. In random forest regression, gain is a feature importance measure that reflects, for a given feature, the mean increase in node purity (ie, mean reduction in variance) that the feature contributes across all splits in which it is used. Z91.19 is a diagnosis code from the *ICD-10* (*International Classification of Diseases, Tenth Revision*), that is used to code for nonadherence to, or noncompliance with, medical treatment. Dx: diagnosis; HbA_{1c}: hemoglobin A_{1c}.
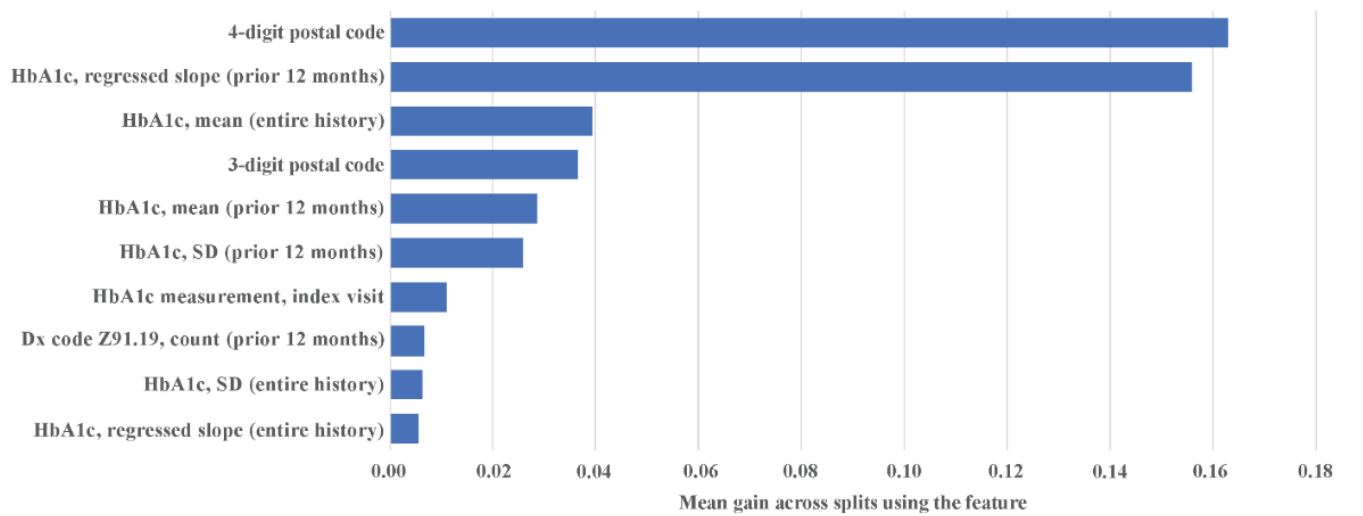
Percent Change Cut Points

Our cross-validation model's ability to accurately predict change in HbA_{1c} at various percent change cut points is illustrated in Table 4. At each percent change cut point (≥0.3%, ≥0.4%, ≥0.5%, and ≥0.6% [approximately 3 mmol/mol, 4 mmol/mol, 5 mmol/mol, 7 mmol/mol]), PPV was 60.3%, 56.4%, 52.7%, and 53.1%, respectively, indicating an approximately 1.5- to 2-fold enrichment (relative to the observed frequency of each outcome [Table 1]) for identifying youth who would experience a clinically significant rise in HbA_{1c}. Sensitivity and PPV improved when predictions involved smaller changes in HbA_{1c}, whereas specificity and NPV improved when predictions involved larger changes in HbA_{1c}. Sensitivity, specificity, PPV, and NPV metrics for each K-fold are in Multimedia Appendix 3.

Table 4. Sensitivity, specificity, positive predictive value, and negative predictive value of predicted versus true percent change in HbA_{1c} across all 3 cross-validation K-folds.

Model metrics at each percent change cut point	Estimate, % (95% CI)
Predicted HbA_{1c}[a] % change: ≥0.3%	
Sensitivity (True HbA_{1c}% change: ≥0.3%)	28.7 (27.3-30.2)
Specificity (True HbA_{1c}% change: ≥0.3%)	87 (86.1-87.9)
PPV[b] (True HbA_{1c}% change: ≥0.3%)	60.3 (58.1-62.5)
NPV[c] (True HbA_{1c}% change: ≥0.3%)	64 (62.9-65)
Predicted HbA_{1c}% change: ≥0.4%	
Sensitivity (True HbA_{1c}% change: ≥0.4%)	17.4 (16.1-18.7)
Specificity (True HbA_{1c}% change: ≥0.4%)	92.6 (91.9-93.2)
PPV (True HbA_{1c}% change: ≥0.4%)	56.4 (53.3-59.4)
NPV (True HbA_{1c}% change: ≥0.4%)	66.9 (65.9-67.9)
Predicted HbA_{1c}% change: ≥0.5%	
Sensitivity (True HbA_{1c}% change: ≥0.5%)	10 (8.9-11.1)
Specificity (True HbA_{1c}% change: ≥0.5%)	96 (95.5-96.5)
PPV (True HbA_{1c}% change: ≥0.5%)	52.7 (48.4-56.9)
NPV (True HbA_{1c}% change: ≥0.5%)	70.6 (69.6-71.5)
Predicted HbA_{1c}% change: ≥0.6%	
Sensitivity (True HbA_{1c}% change: ≥0.6%)	6.1 (5.2-7.1)
Specificity (True HbA_{1c}% change: ≥0.6%)	98.1 (97.7-98.4)

Model metrics at each percent change cut point	Estimate, % (95% CI)
PPV (True HbA$_{1c}$% change: ≥0.6%)	53.1 (47.2-58.9)
NPV (True HbA1% change: ≥0.6%)	74.4 (73.5-75.3)

[a]HbA$_{1c}$: glycated hemoglobin.

[b]PPV: positive predictive value (it is the probability that the cases predicted to experience clinically significant rise in HbA$_{1c}$ [at or above each percent rise threshold] did experience that outcome).

[c]NPV: negative predictive value (it is the probability that the cases not predicted to experience clinically significant rise in HbA$_{1c}$ [at or above each percent rise threshold] did not experience that outcome).

Discussion

Principal Findings

We used routinely collected EHR data, including both structured and unstructured data, to establish the feasibility of constructing an interpretable ML model for predicting unit-change in HbA$_{1c}$ (in % units) between quarterly diabetes clinic visits among youth (aged 9-18 y) with T1D. For those predicted to experience a ≥0.3% (approximately 3 mmol/mol) rise in HbA$_{1c}$ during the following 3 months, PPV was 60.3%, indicating a 1.5-fold enrichment (relative to the observed frequency [40.7%] of this outcome) for identifying youth who would experience a clinically significant rise in HbA$_{1c}$. This finding, which suggests that EHR data may be useful for identifying youth who will experience rising glycemic levels, is clinically relevant given that a long-term increase of ≥0.3% (3 mmol/mol) in HbA$_{1c}$ is associated with increased risk for long-term complications of diabetes [28].

Another key finding was that our model's sensitivity and PPV were higher when the predicted percent rise threshold was lower (eg, ≥0.3% vs ≥0.4%), whereas specificity and NPV were increased at higher predicted percent rise thresholds (eg, ≥0.4% vs ≥0.3%). We hypothesized that using a higher percent rise threshold would decrease the likelihood of false positives (ie, identifying a youth as someone who would experience a corresponding rise in HbA$_{1c}$ when they did not), and the data supported this conclusion. On the other hand, using a lower percent rise threshold reduced the likelihood of missing those who would experience a clinically significant rise in HbA$_{1c}$. If confirmed in future studies, these findings suggest that using the lowest clinically significant threshold may be useful for guiding clinical decision-making and subsequent initiation of interventions designed to mitigate rising glycemic levels.

We also evaluated our model's ability to augment clinical decision-making by using PPV and NPV to identify a percent-change cut point that optimized identification of youth who experienced a clinically significant rise in HbA$_{1c}$ at their subsequent diabetes clinic encounter. Although PPV and NPV are considered the metrics of choice for clinical decision-making at the level of an individual person, the selection of desirable PPV and NPV values in a particular use case depends on numerous factors. These factors include considerations about short- and long-term burdens and costs related to over- or undertreatment, associated psychological impacts on individuals receiving care, and short- and long-term costs imposed on the health care system

(eg, for increased staffing resources) [43]. Therefore, before implementing this model clinically, it would be important to allow clinicians to provide feedback about the most appropriate thresholds for defining clinically significant rise in HbA$_{1c}$, along with associated PPV and NPV values. For this work, we propose using the ≥0.3% cut-point to maximize capture of high-risk youth who are candidates for behavioral and care delivery interventions designed to reduce or prevent predicted rise in HbA$_{1c}$.

The top features impacting our model's predictions (ie, postal code, numerous metrics pertaining to HbA$_{1c}$, and history of low treatment engagement) have been shown in previous studies to be associated with elevated glycemic levels. Ample evidence suggests associations between geographic location and geographically linked measures of socioeconomic status (eg, area deprivation, social deprivation, and child opportunity indices) and T1D outcomes, including glycemic levels and diabetic ketoacidosis [44-47]. Previous HbA$_{1c}$ measurements have also been shown to significantly impact ML-based predictions of future HbA$_{1c}$, but previous investigations have only examined this in adults with type 2 diabetes (T2D) [48]. Finally, lower treatment engagement has been shown to have a substantial impact on HbA$_{1c}$ in youth with T1D [49,50]. This evidence collectively underscores the critical need for members of the diabetes care team to partner with affected youth and families to identify resources and tailored strategies for optimizing diabetes self-management behaviors.

Given the widespread use of EHRs in clinical care, as well as the growing volume and availability of these data, there exists tremendous potential for using EHR data to identify and personalize care pathways for improving health outcomes in T1D. Previous work has applied ML to EHR data, for example, to predict the onset of T1D in youth [51], as well as diabetic ketoacidosis in both youth and adults with T1D [19,20,52]. Recent research has focused on applying numerous ML classifiers to medical encounter data to predict HbA$_{1c}$ in individuals with T2D [48]. The area under the receiver operating curve for each of the top 5 best-performing classifiers in the aforementioned study was extremely high (>0.95). Of note, however, these model predictions were binary (ie, HbA$_{1c}$ <7% [<53 mmol/mol] vs ≥7% [≥53 mmol/mol]) rather than continuous and were evaluated in a primarily adult Chinese cohort diagnosed with T2D, limiting generalizability to other populations. Our approach is designed to predict unit change in HbA$_{1c}$ and to give clinicians a simple output (ie, HbA$_{1c}$ will or will not increase by ≥0.3%) for interpretation. This study is the first to

use EHR data to predict a clinically significant rise in HbA_{1c} in youth with T1D.

Recent efforts have also explored the use of ML classifiers that use 2 weeks of CGM data to forecast 90-day HbA_{1c} in youth with T1D [53]. The first of these studies used a nested, ensemble learning approach to iteratively predict HbA_{1c} in stages: (1) HbA_{1c} ≤7.5% (58 mmol/mol) or >7.5% (stage 1), (2) HbA_{1c} ≤9% (75 mmol/mol) or >9% (stage 2, after stage 1 was complete), and (3) HbA_{1c} ≤12.5% (113 mmol/mol) or >12.5% (stage 3, after stage 2 was complete) [54]. A subsequent study used few-shot learning followed by K-nearest neighbors to classify transformed images of CGM time series data into multiclass HbA_{1c} intervals [55]. Generalizability of these HbA_{1c} prediction efforts is limited, however, by these methods' dependence on CGM data and by racial disparities in the relationship between CGM metrics and HbA_{1c} [56].

Currently, CGM systems are neither accessible to nor used by all individuals with T1D. Recent data from the T1D Exchange Quality Improvement Collaborative suggest that only 40%-50% of US youth with T1D currently use CGM systems [57,58]. Reasons for this are multifactorial and can include reluctance to use CGM technologies, financial constraints, lack of insurance coverage, device-related skin complications, CGM alarm fatigue, and sociodemographic and racial or ethnic disparities in access that adversely impact use of diabetes technologies [58-62]. At this time, CGM data also remain notably absent from most EHRs, are distributed across multiple proprietary commercial software, and are difficult for health systems to access. Although efforts to integrate CGM data into the EHR remain ongoing [63,64], large-scale implementation of these efforts will hinge on the development of CGM-related data standards and a data architecture that supports this integration [65].

In contrast, EHR data are routinely collected on every person receiving care from a given health care institution. These data thus provide a rich, longitudinal source of individual- and population-level health data that can be leveraged in near real-time for ML-driven clinical decision support [13,66]. Even so, the potential for integrating EHR-based ML-driven analytics in health care remains largely unrealized. A 2020 systematic review evaluating the number of clinical prediction models that have been embedded into EHRs noted that fewer than 45 such examples have been published [67]. Of note, only 36% (16/45) of model implementations occurred in outpatient settings, and none of the embedded models were specific to individuals affected by diabetes [67]. These findings highlight a critical gap, as well as opportunity, for leveraging real-world EHR data to facilitate real-time risk prediction and improve diabetes-related health outcomes.

Limitations and Strengths

A strength of this study is its use of longitudinal EHR data to predict 90-day unit-change in HbA_{1c} in a large cohort of youth with T1D. The scale and granularity of these data facilitated the creation of thousands of data features that we simultaneously analyzed as potential predictors

for suboptimal glycemic outcomes. Additional strengths of this study include its use of explainable ML methods for evaluating model predictions and our use of a clinician-led, postmodeling decision analysis to enhance clinicians' understanding and uptake of model predictions. The relevance of our model is underscored by its ability to forecast 90-day change in HbA_{1c} for all youth receiving care through our regional clinic network, and not only for those using CGM systems.

Several limitations also warrant consideration. The data used in this study originated from a regional network of diabetes clinics in the Midwest United States and may not generalize to other geographic locations or health care settings, to future cohorts using rapidly evolving diabetes treatment technologies, or to more racially or socioeconomically diverse cohorts. External validation of the geographic and demographic "transportability" of this and future iterations of our model will hinge on ensuring that data from different clinical settings are collected in similar ways and standardized according to a common data model. Examples of such data standards include the Observational Medical Outcomes Partnership Common Data Model [68] and the T1D Exchange Quality Improvement Collaborative data specification [69]. As well, EHR data are subject to data entry errors and missing data that inadvertently occur as a part of routine clinical care. EHRs are also characterized by data fragmentation and reflect biases in clinical data collection, documentation, and decision-making [13]. Therefore, results from this and all models constructed using EHR data must be interpreted carefully, given both known and unknown biases that impact model predictions.

Model generalizability could be enhanced by using standardized geographic-based features (eg, an area deprivation index or the Child Opportunity Index [45,47]) rather than zip code, as well as by creating a final prediction model that includes only a limited number of the "top-N features identified via cross-validation. Using additional data preprocessing methods (eg, one-hot encoding) when transforming categorical demographic features (eg, race and ethnicity) for ML would facilitate interpretability of model results pertaining to those features. Model performance may improve with additional hyperparameter tuning. This model's predictive utility could also be compared with that of models constructed using other ML methods, including other explainable AI methods and deep learning models. Finally, for youth who adopt diabetes technologies, such as CGM and automated insulin delivery systems, the inclusion of diabetes device data would likely significantly augment our model's predictions.

We acknowledge that translation of this work into clinical practice will be accompanied by various logistical and practical challenges. This study was designed as an "initial step" to evaluate the feasibility of using EHR data to predict change in HbA_{1c}. As previously described, additional research is needed to address issues related to model refinement, validation (using data from external organizations, as well as future EHR data collected from our network of diabetes centers), and deployment in clinical settings.

Future work can, for example, evaluate whether a limited set of standardized features may be useful for developing a more parsimonious model that can be readily disseminated to other institutions. Once deployed, ongoing monitoring of model performance will also be needed.

Furthermore, we acknowledge that refining and successfully incorporating this approach into clinical and decision workflows will hinge on the collection of additional evidence from future studies with even larger and more diverse patient cohorts, as well as buy-in and trust from both clinicians and patients. Although, in this iteration, our modeling approach yielded a nontrivial number of false positives, we note as well that our model's performance represents a substantial improvement over existing capabilities. Compared, for example, with initiating interventions randomly or initiating interventions at every diabetes clinic visit (to address youths' rising glycemic levels, which occurred 40.7% of the time in our cohort), our modeling efforts facilitated pre-emptive identification of rising glucose levels three-fifths of the time. The 1.5-fold risk enrichment demonstrated in this work represents a meaningfully improved opportunity for more targeted initiation and delivery of interventions designed to lower youths' glucose levels.

Conclusions

Using EHR data to develop an ML-based prediction model to identify youth who will experience a clinically significant rise in HbA_{1c} between diabetes clinic visits is both timely and feasible. Future research should aim to further optimize model performance, as well as evaluate model performance in racially or ethnically, socioeconomically, and geographically diverse cohorts. Future work is also needed to evaluate whether model results vary by duration of diabetes, use of technology (eg, CGM system users vs nonusers), and insulin delivery modality. Findings from this study may help to inform risk stratification and resource allocation efforts and serve as a catalyst for future quality improvement efforts focused on developing and evaluating personalized strategies and supports for optimizing diabetes self-management behaviors.

Data Availability

The datasets generated and analyzed for this study are not publicly available due to sensitive information contained in patient medical records. Interested parties should contact the corresponding author to inquire about access. Source code for this study is available from the corresponding author upon reasonable request.

Authors Contributions

EMT and CS participated in interpreting data and drafting the manuscript. DDW, CM, and MAC participated in study conceptualization and design, analysis, interpreting data, and drafting or revising the manuscript. BL, DF, CAV, MSB, ACS, ARK, SRP, SM, RM, ML, and LD participated in editing and revising the manuscript. All authors approved this manuscript for submission.

Conflicts of Interest

CM and LD are employees of Blue Circle Health. RM is a consultant for Sanofi. ML has received research grants from Eli Lilly and Novo Nordisk and has been a consultant or has received honoraria from Astra Zeneca, Boehringer Ingelheim, Eli Lilly, Nordicinfu Care, Novo Nordisk, and Rubin Medical, all outside the submitted work. MAC is a consultant for Glooko, Inc. and receives research support from Dexcom and Abbott Diabetes Care. All other authors are responsible for the reported research and stated that they have no affiliation, financial agreement, or involvement with any company or other organization with a financial interest in the subject matter of the submitted manuscript.

Multimedia Appendix 1

Electronic health records–based identification of a cohort of youth with type 1 diabetes.
[DOCX File (Microsoft Word File), 27 KB-Multimedia Appendix 1]

Multimedia Appendix 2

Top 30 most important features for predicting 90-day percent change in glycated hemoglobin in youth with type 1 diabetes.
[DOCX File (Microsoft Word File), 30 KB-Multimedia Appendix 2]

Multimedia Appendix 3

Sensitivity, specificity, positive predictive value, and negative predictive value of predicted versus true percent change in glycated hemoglobin for each cross-validation K-fold.

[DOCX File (Microsoft Word File), 30 KB-Multimedia Appendix 3]

References

1. DiMeglio LA, Evans-Molina C, Oram RA. Type 1 diabetes. Lancet. Jun 16, 2018;391(10138):2449-2462. [doi: 10.1016/S0140-6736(18)31320-5] [Medline: 29916386]

2. Fang M, Wang D, Selvin E. Prevalence of type 1 diabetes among US children and adults by age, sex, race, and ethnicity. JAMA. Apr 23, 2024;331(16):1411-1413. [doi: 10.1001/jama.2024.2103] [Medline: 38573653]

3. American Diabetes Association Professional Practice Committee. 6. Glycemic goals and hypoglycemia: standards of care in diabetes—2024. Diabetes Care. Jan 1, 2024;47(Supplement_1):S111-S125. [doi: 10.2337/dc24-S006]

4. American Diabetes Association Professional Practice Committee. 14. Children and adolescents: standards of care in diabetes—2024. Diabetes Care. Jan 1, 2024;47(Supplement_1):S258-S281. [doi: 10.2337/dc24-S014]

5. Patiño-Fernández AM, Eidson M, Sanchez J, Delamater AM. What do youth with type 1 diabetes know about the HbA1c test? Child Health Care. Apr 1, 2010;38(2):157-167. [doi: 10.1080/02739610902813328] [Medline: 20563233]

6. Foster NC, Beck RW, Miller KM, et al. State of type 1 diabetes management and outcomes from the T1D Exchange in 2016-2018. Diabetes Technol Ther. Feb 2019;21(2):66-72. [doi: 10.1089/dia.2018.0384] [Medline: 30657336]

7. American Diabetes Association Professional Practice Committee. 7. Diabetes technology: standards of care in diabetes-2024. Diabetes Care. Jan 1, 2024;47(Suppl 1):S126-S144. [doi: 10.2337/dc24-S007] [Medline: 38078575]

8. Redondo MJ, Libman I, Maahs DM, et al. The evolution of hemoglobin A_{1c} targets for youth with type 1 diabetes: rationale and supporting evidence. Diabetes Care. Feb 2021;44(2):301-312. [doi: 10.2337/dc20-1978] [Medline: 33431422]

9. Hermann JM, Miller KM, Hofer SE, et al. The Transatlantic HbA_{1c} gap: differences in glycaemic control across the lifespan between people included in the US T1D Exchange Registry and those included in the German/Austrian DPV registry. Diabet Med. May 2020;37(5):848-855. [doi: 10.1111/dme.14148] [Medline: 31557351]

10. Clements MA, Schwandt A, Donaghue KC, et al. Five heterogeneous HbA1c trajectories from childhood to adulthood in youth with type 1 diabetes from three different continents: a group-based modeling approach. Pediatr Diabetes. Nov 2019;20(7):920-931. [doi: 10.1111/pedi.12907] [Medline: 31418521]

11. Lachin JM, Nathan DM, DCCT/EDIC Research Group. Understanding metabolic memory: the prolonged influence of glycemia during the Diabetes Control and Complications Trial (DCCT) on future risks of complications during the study of the Epidemiology of Diabetes Interventions and Complications (EDIC). Diabetes Care. Sep 21, 2021;44(10):2216-2224. [doi: 10.2337/dc20-3097] [Medline: 34548284]

12. Writing Team for the Diabetes Control and Complications Trial/Epidemiology of Diabetes Interventions and Complications Research Group. Sustained effect of intensive treatment of type 1 diabetes mellitus on development and progression of diabetic nephropathy: the Epidemiology of Diabetes Interventions and Complications (EDIC) study. JAMA. Oct 22, 2003;290(16):2159-2167. [doi: 10.1001/jama.290.16.2159] [Medline: 14570951]

13. Tang AS, Woldemariam SR, Miramontes S, Norgeot B, Oskotsky TT, Sirota M. Harnessing EHR data for health research. Nat Med. Jul 2024;30(7):1847-1855. [doi: 10.1038/s41591-024-03074-8] [Medline: 38965433]

14. Sauer CM, Chen LC, Hyland SL, Girbes A, Elbers P, Celi LA. Leveraging electronic health records for data science: common pitfalls and how to avoid them. Lancet Digit Health. Dec 2022;4(12):e893-e898. [doi: 10.1016/S2589-7500(22)00154-6] [Medline: 36154811]

15. Goldstein BA, Navar AM, Pencina MJ, Ioannidis JPA. Opportunities and challenges in developing risk prediction models with electronic health records data: a systematic review. J Am Med Inform Assoc. Jan 2017;24(1):198-208. [doi: 10.1093/jamia/ocw042] [Medline: 27189013]

16. Zrubka Z, Kertész G, Gulácsi L, et al. The reporting quality of machine learning studies on pediatric diabetes mellitus: systematic review. J Med Internet Res. Jan 19, 2024;26:e47430. [doi: 10.2196/47430] [Medline: 38241075]

17. Liu K, Li L, Ma Y, et al. Machine learning models for blood glucose level prediction in patients with diabetes mellitus: systematic review and network meta-analysis. JMIR Med Inform. Nov 20, 2023;11:e47833. [doi: 10.2196/47833] [Medline: 37983072]

18. Fregoso-Aparicio L, Noguez J, Montesinos L, García-García JA. Machine learning and deep learning predictive models for type 2 diabetes: a systematic review. Diabetol Metab Syndr. Dec 20, 2021;13(1):148. [doi: 10.1186/s13098-021-00767-9] [Medline: 34930452]

19. Williams DD, Ferro D, Mullaney C, et al. An "All-Data-on-Hand" deep learning model to predict hospitalization for diabetic ketoacidosis in youth with type 1 diabetes: development and validation study. JMIR Diabetes. Jul 18, 2023;8:e47592. [doi: 10.2196/47592] [Medline: 37224506]

20.   Subramanian D, Sonabend R, Singh I. A machine learning model for risk stratification of postdiagnosis diabetic ketoacidosis hospitalization in pediatric type 1 diabetes: retrospective study. JMIR Diabetes. Aug 7, 2024;9:e53338. [doi: 10.2196/53338] [Medline: 39110490]

21.   Tallon EM, Ebekozien O, Sanchez J, et al. Impact of diabetes status and related factors on COVID-19-associated hospitalization: a nationwide retrospective cohort study of 116,370 adults with SARS-CoV-2 infection. Diabetes Res Clin Pract. Dec 2022;194:110156. [doi: 10.1016/j.diabres.2022.110156] [Medline: 36400172]

22.   What is random forest? IBM. URL: https://www.ibm.com/topics/random-forest [Accessed 2024-12-23]

23.   Breiman L. Random Forests. Mach Learn. Oct 2001;45(1):5-32. [doi: 10.1023/A:1010933404324]

24.   Oracle. Oracle Health EHR. URL: https://www.oracle.com/health/clinical-suite/electronic-health-record/ [Accessed 2024-12-23]

25.   Prahalad P, Yang J, Scheinker D, Desai M, Hood K, Maahs DM. Hemoglobin A1c trajectory in pediatric patients with newly diagnosed type 1 diabetes. Diabetes Technol Ther. Aug 2019;21(8):456-461. [doi: 10.1089/dia.2019.0065] [Medline: 31180244]

26.   Ibfelt EH, Wibaek R, Vistisen D, et al. Trajectory and predictors of HbA1c in children and adolescents with type 1 diabetes-a Danish nationwide cohort study. Pediatr Diabetes. Sep 2022;23(6):721-728. [doi: 10.1111/pedi.13337] [Medline: 35366046]

27.   Miller KM, Foster NC, Beck RW, et al. Current state of type 1 diabetes treatment in the U.S.: updated data from the T1D Exchange clinic registry. Diabetes Care. Jun 2015;38(6):971-978. [doi: 10.2337/dc15-0078] [Medline: 25998289]

28.   Lind M, Polonsky W, Hirsch IB, et al. Continuous glucose monitoring vs conventional therapy for glycemic control in adults with type 1 diabetes treated with multiple daily insulin injections: the GOLD randomized clinical trial. JAMA. Jan 24, 2017;317(4):379-387. [doi: 10.1001/jama.2016.19976] [Medline: 28118454]

29.   Multum MediSource Lexicon (MMSL) Source Information. National Library of Medicine. Mar 31, 2020. URL: https://www.nlm.nih.gov/research/umls/rxnorm/sourcereleasedocs/mmsl.html

30.   Kavlakoglu E, Winland V. What is k-means clustering? IBM. 2024. URL: https://www.ibm.com/topics/k-means-clustering [Accessed 2024-12-23]

31.   StringIndexer. Apache Software Foundation. URL: https://spark.apache.org/docs/latest/api/python/reference/api/pyspark.ml.feature.StringIndexer.html [Accessed 2024-12-23]

32.   Agency for Healthcare Research and Quality. Healthcare Cost & Utilization Project user support: Clinical Classifications Software Refined (CCSR) for ICD-10-CM diagnoses. 2024. URL: https://hcup-us.ahrq.gov/toolssoftware/ccsr/dxccsr.jsp [Accessed 2024-12-23]

33.   Manning CD, Raghavan P, Schütze H. An Introduction to Information Retrieval. Cambridge, England: Cambridge University Press; 2008. URL: https://nlp.stanford.edu/IR-book/pdf/irbookonlinereading.pdf [Accessed 2025-09-19] ISBN: 0521865719

34.   What is bagging? IBM. URL: https://www.ibm.com/topics/bagging [Accessed 2024-12-23]

35.   Liaw A, Wiener M. Classification and regression by randomforest. R News. 2002:18-22. URL: https://journal.r-project.org/articles/RN-2002-022/RN-2002-022.pdf [Accessed 2025-09-19]

36.   Breiman L. Bagging predictors. Mach Learn. Aug 1996;24(2):123-140. [doi: 10.1007/BF00058655]

37.   Jackson EK, Roberts W, Nelsen B, Williams GP, Nelson EJ, Ames DP. Introductory overview: error metrics for hydrologic modelling – a review of common practices and an open source library to facilitate use and adoption. Environ Model Softw. Sep 2019;119:32-48. [doi: 10.1016/j.envsoft.2019.05.001]

38.   Apache Software Foundation. RandomForestRegressor. URL: https://spark.apache.org/docs/3.5.2/api/scala/org/apache/spark/ml/regression/RandomForestRegressor.html [Accessed 2024-12-23]

39.   Ishwaran H. The effect of splitting on random forests. Mach Learn. Apr 2015;99(1):75-118. [doi: 10.1007/s10994-014-5451-2] [Medline: 28919667]

40.   Nembrini S, König IR, Wright MN. The revival of the Gini importance? Bioinformatics. Nov 1, 2018;34(21):3711-3718. [doi: 10.1093/bioinformatics/bty373] [Medline: 29757357]

41.   Machine Learning Library (MLlib) Guide. Apache Software Foundation. URL: https://spark.apache.org/docs/latest/ml-guide [Accessed 2024-12-23]

42.   StataCorp. Stata Statistical Software. 18th ed. College Station, Texas: StataCorp LLC; 2023.

43.   Trevethan R. Sensitivity, specificity, and predictive values: foundations, pliabilities, and pitfalls in research and practice. Front Public Health. 2017;5:307. [doi: 10.3389/fpubh.2017.00307] [Medline: 29209603]

44.   Holm TF, Jensen MH, Hejlesen OK, Hagstrøm S, Madsen M, Hangaard S. Prediction of poor glycemic control in children with type 1 diabetes. Stud Health Technol Inform. Aug 22, 2024;316:1759-1760. [doi: 10.3233/SHTI240770] [Medline: 39176556]

45. Everett E, Mathioudakis N. Association of area deprivation and diabetic ketoacidosis readmissions: comparative risk analysis of adults vs children with type 1 diabetes. J Clin Endocrinol Metab. Aug 1, 2019;104(8):3473-3480. [doi: 10.1210/jc.2018-02232] [Medline: 31220288]

46. Carter PJ, Cutfield WS, Hofman PL, et al. Ethnicity and social deprivation independently influence metabolic control in children with type 1 diabetes. Diabetologia. Oct 2008;51(10):1835-1842. [doi: 10.1007/s00125-008-1106-9] [Medline: 18679654]

47. Hoyek K, Libman I, Mkparu N, Hong YH, Arslanian S, Vajravelu ME. Child Opportunity Index and clinical characteristics at diabetes diagnosis in youth: type 1 diabetes versus type 2 diabetes. BMJ Open Diabetes Res Care. Apr 17, 2024;12(2):e003968. [doi: 10.1136/bmjdrc-2023-003968] [Medline: 38631820]

48. Tao X, Jiang M, Liu Y, et al. Predicting three-month fasting blood glucose and glycated hemoglobin changes in patients with type 2 diabetes mellitus based on multiple machine learning algorithms. Sci Rep. Sep 30, 2023;13(1):16437. [doi: 10.1038/s41598-023-43240-5] [Medline: 37777593]

49. Bombaci B, Torre A, Longo A, et al. Psychological and clinical challenges in the management of type 1 diabetes during adolescence: a narrative review. Children (Basel). Sep 4, 2024;11(9):1085. [doi: 10.3390/children11091085] [Medline: 39334618]

50. Lee JM, Rusnak A, Garrity A, et al. Feasibility of electronic health record assessment of 6 pediatric type 1 diabetes self-management habits and their association with glycemic outcomes. JAMA Netw Open. Oct 1, 2021;4(10):e2131278. [doi: 10.1001/jamanetworkopen.2021.31278] [Medline: 34709387]

51. Daniel R, Jones H, Gregory JW, et al. Predicting type 1 diabetes in children using electronic health records in primary care in the UK: development and validation of a machine-learning algorithm. Lancet Digit Health. Jun 2024;6(6):e386-e395. [doi: 10.1016/S2589-7500(24)00050-5] [Medline: 38789139]

52. Li L, Lee CC, Zhou FL, et al. Performance assessment of different machine learning approaches in predicting diabetic ketoacidosis in adults with type 1 diabetes using electronic health records data. Pharmacoepidemiol Drug Saf. May 2021;30(5):610-618. [doi: 10.1002/pds.5199] [Medline: 33480091]

53. Bergenstal RM, Beck RW, Close KL, et al. Glucose Management Indicator (GMI): a new term for estimating A1C from continuous glucose monitoring. Diabetes Care. Nov 2018;41(11):2275-2280. [doi: 10.2337/dc18-1581] [Medline: 30224348]

54. Islam MS, Qaraqe MK, Belhaouari S, Petrovski G. Long term HbA1c prediction using multi-stage CGM data analysis. IEEE Sensors J. 2021;21(13):15237-15247. [doi: 10.1109/JSEN.2021.3073974]

55. Qaraqe M, Elzein A, Belhaouari S, Ilam MS, Petrovski G. A novel few shot learning derived architecture for long-term HbA1c prediction. Sci Rep. Jan 4, 2024;14(1):482. [doi: 10.1038/s41598-023-50348-1] [Medline: 38177624]

56. Bergenstal RM, Gal RL, Connor CG, et al. Racial differences in the relationship of glucose concentrations and hemoglobin A1c levels. Ann Intern Med. Jul 18, 2017;167(2):95-102. [doi: 10.7326/M16-2596] [Medline: 28605777]

57. DeSalvo DJ, Lanzinger S, Noor N, et al. Transatlantic comparison of pediatric continuous glucose monitoring use in the diabetes-patienten-verlaufsdokumentation initiative and type 1 diabetes exchange quality improvement collaborative. Diabetes Technol Ther. Dec 2022;24(12):920-924. [doi: 10.1089/dia.2022.0248] [Medline: 35947079]

58. DeSalvo DJ, Noor N, Xie C, et al. Patient demographics and clinical outcomes among type 1 diabetes patients using continuous glucose monitors: data from T1D exchange real-world observational study. J Diabetes Sci Technol. Mar 2023;17(2):322-328. [doi: 10.1177/19322968211049783] [Medline: 34632823]

59. Ebekozien O, Mungmode A, Sanchez J, et al. Longitudinal trends in glycemic outcomes and technology use for over 48,000 people with type 1 diabetes (2016-2022) from the T1D exchange quality improvement collaborative. Diabetes Technol Ther. Nov 2023;25(11):765-773. [doi: 10.1089/dia.2023.0320] [Medline: 37768677]

60. Rigo RS, Levin LE, Belsito DV, Garzon MC, Gandica R, Williams KM. Cutaneous reactions to continuous glucose monitoring and continuous subcutaneous insulin infusion devices in type 1 diabetes mellitus. J Diabetes Sci Technol. Jul 2021;15(4):786-791. [doi: 10.1177/1932296820918894] [Medline: 32389062]

61. Tilden DR, French B, Datye KA, Jaser SS. Disparities in continuous glucose monitor use between children with type 1 diabetes living in urban and rural areas. Diabetes Care. Mar 1, 2024;47(3):346-352. [doi: 10.2337/dc23-1564] [Medline: 37906202]

62. Barnard-Kelly KD, Martínez-Brocca MA, Glatzer T, Oliver N. Identifying the deficiencies of currently available CGM to improve uptake and benefit. Diabet Med. Aug 2024;41(8):e15338. [doi: 10.1111/dme.15338] [Medline: 38736324]

63. Espinoza J, Shah P, Raymond J. Integrating continuous glucose monitor data directly into the electronic health record: proof of concept. Diabetes Technol Ther. Aug 2020;22(8):570-576. [doi: 10.1089/dia.2019.0377] [Medline: 31904260]

64. Okuno T, Macwan SA, Miller D, Norman GJ, Reaven P, Zhou JJ. Assessing patterns of continuous glucose monitoring use and metrics of glycemic control in type 1 diabetes and type 2 diabetes patients in the veterans health care system: integrating continuous glucose monitoring device data with electronic health records data. Diabetes Technol Ther. Nov 2024;26(11):806-813. [doi: 10.1089/dia.2024.0083] [Medline: 38768417]

65. Espinoza J, Xu NY, Nguyen KT, Klonoff DC. The need for data standards and implementation policies to integrate CGM data into the electronic health record. J Diabetes Sci Technol. Mar 2023;17(2):495-502. [doi: 10.1177/19322968211058148] [Medline: 34802286]
66. Kamel Rahimi A, Canfell OJ, Chan W, et al. Machine learning models for diabetes management in acute care using electronic medical records: a systematic review. Int J Med Inform. Jun 2022;162:104758. [doi: 10.1016/j.ijmedinf.2022.104758] [Medline: 35398812]
67. Lee TC, Shah NU, Haack A, Baxter SL. Clinical implementation of predictive models embedded within electronic health record systems: a systematic review. Informatics (MDPI). Sep 2020;7(3):25. [doi: 10.3390/informatics7030025] [Medline: 33274178]
68. Standardized data: the OMOP common data model. Observational Health Data Sciences and Informatics. URL: https://www.ohdsi.org/data-standardization/ [Accessed 2025-06-19]
69. Mungmode A, Noor N, Weinstock RS, et al. Making diabetes electronic medical record data actionable: promoting benchmarking and population health improvement using the T1D exchange quality improvement portal. Clin Diabetes. 2022;41(1):45-55. [doi: 10.2337/cd22-0072] [Medline: 36714251]

Abbreviations

CCSR: Clinical Classification Software Revised
CGM: continuous glucose monitoring
CPT: Current Procedural Terminology
EHR: electronic health record
HbA$_{1c}$: glycated hemoglobin
ICD-10: *International Statistical Classification of Diseases, Tenth Revision*
ICD-9: *International Classification of Diseases, Ninth Revision*
MAE: mean absolute error
ML: machine learning
NPV: negative predictive value
PPV: positive predictive value
RF: random forest
RMSE: root-mean-square error
SNOMED CT: Systematized Nomenclature of Medicine Clinical Terms
T1D: type 1 diabetes
T2D: type 2 diabetes
TF-IDF: term frequency-inverse document frequency