Journal of Neonatal Surgery
ISSN(Online): 2226-0439
Vol. 14, Issue 18s (2025)
https://www.jneonatalsurg.com

OPEN ACCESS

A machine learning based approach to factors predicting correlation between Hb1Ac and diabetes in patients of malnourished communities

Sweta Mishra[1], Dr. Ratnesh Litoriya[2]

1Research Scholar, Department of Computer Science and Engineering, Medi-Caps University, Indore, Madhya Pradesh, India, sweta.mishra@medicaps.ac.in

2Professor, Department of Computer Science and Engineering, Medi-Caps University, Indore, Madhya Pradesh, India ratnesh.litoriya@medicaps.ac.in

Cite this paper as: Sweta Mishra, Dr. Ratnesh Litoriya, (2025) A machine learning based approach to factors predicting correlation between Hb1Ac and diabetes in patients of malnourished communities. *Journal of Neonatal Surgery*, 14 (18s), 560-567.

ABSTRACT

Type 1 diabetes mellitus (T1DM) is associated with acute complications such as diabetic ketoacidosis (DKA) and long-term glycemic dysregulation. This study aimed to develop and validate machine learning models to predict DKA episodes and glycemic control, defined as HbA1c >7%, using a large multi-center, bi-national database from the Diabetes Data Network (DDN). Nine machine learning algorithms, including Deep Learning (DL), Support Vector Machine (SVM), Random Forest (RF), and Logistic Regression (LR), were trained and validated on clinical and demographic features collected longitudinally from 10,868 individuals aged 2–21 years between January 2012 and May 2022. The DL model demonstrated the highest predictive performance for DKA, achieving an area under the curve (AUC) of 0.887, while SVM was most effective in predicting HbA1c >7% with an AUC of 0.884. Key predictors for DKA included age at diagnosis, diabetes duration, prior DKA events, BMI z-score, HbA1c, CGM use, insulin regimen, and center. For HbA1c prediction, baseline HbA1c and BMI emerged as dominant features. The results suggest that integrating machine learning models into clinical care could enable early identification of high-risk individuals, facilitating timely interventions and potentially reducing hospitalization rates and healthcare costs associated with T1DM complications. These models underscore the importance of personalized management strategies and highlight the feasibility of real-world application in diabetes clinics.

Keywords: Type 1 Diabetes Mellitus, Diabetic Ketoacidosis (DKA), Machine Learning Models, Glycemic Control (HbA1c), Predictive Analytics in Healthcare

1. INTRODUCTION

Type 1 diabetes mellitus (DM) is becoming more common and is associated with a number of costly and difficult consequences. Type 1 diabetes is the most prevalent type among young people, according to reports, however many are diagnosed in adulthood. A serious and potentially fatal acute consequence of Type 1 diabetes is diabetic ketoacidosis (DKA) [1]. Although DKA can happen to anyone with Type 1 diabetes, it is the first symptom that 6–21% of individuals with the disease experience. It has been noted that among adults with a diabetes diagnosis, those with infections, other illnesses, psychological stress, or those who have neglected or used insulin therapy insufficiently are the ones most likely to get DKA [2, 3]. Tomic and colleagues discovered that among people with type 1 diabetes under the age of 20, more than 60% of hospitalizations were directly due to diabetes, with nearly half of them being for ketoacidosis.

The most accurate indicator for evaluating long-term diabetes control is HbA1c, which measures blood glucose levels over the previous two to three months. Higher HbA1c levels have been found to be independently associated with a number of factors over time, such as indigenous status, speaking a language other than English at home, using a multiple daily injection (MDI) regimen rather than continuous subcutaneous insulin infusion (CSII) or "insulin pump therapy," not using continuous glucose monitoring (CGM), having diabetes for a longer period of time, being older when diagnosed, having more annual visits, and having a lower body mass index (BMI) z-score. With severe DKA at diagnosis linked to increased HbA1c over a median follow-up period of 5.6 years, we also showed in that same study that there is a positive correlation between presentation at diagnosis with DKA and eventual HbA1c. A literature review on the use of artificial intelligence for diabetes management and decision support found evidence of acceleration in the development of artificial

intelligence-powered tools for prediction and prevention of complications associated with diabetes [4]. Additionally, supervised machine learning techniques have been successfully used in the past to classify short and long term HbA1c response after treatment initiation with good AUC values of 0.80 and 0.81, respectively. It has been observed that supervised machine learning techniques outperform standard statistical regression models in big clinical datasets [5-9]. However, there has never been a comprehensive evaluation of several machine learning models to predict DKA and HbA1c>7% in a multi-center type 1 diabetes clinic environment. Using a large bi-national, multi-center database, our goal was to create and compare many machine learning approaches to predict which patients with type 1 diabetes would present with an undesirable result of DKA during their clinic visit. Predicting a HbA1c >7% at the most recent clinic visit was the secondary outcome [10, 11]. In order to enable the future integration of these models into clinical databases at the point of care, we sought to identify critical risk indicators (features) that were predictive of the outcomes for the best models [12-16].

2. SAMPLING METHODOLOGY

2.1. Study population

The Diabetes Data Network (DDN), which provided the data for this investigation. De-identified, prospectively gathered individual data from 33 participating DDN centers that treat patients with type 1 diabetes is included in the database. Data was moved from sites to the register twice a year using a common data dictionary. All participants aged 2 and older who were followed up with at the centers between January 2012 and May 2022 are included in the data extraction for this study, which was completed in May 2022. In our study, we only included people with type 1 diabetes. Patients who were older than 21 at the time of their clinic visit were not included in our analysis. The Monash University Human Research Ethics Committee granted ethics approval for this study (Project ID: 37978).

2.2. Feature variable prediction

The reported DKA at the most recent clinic visit (coded as Yes/No) served as our main outcome variable. Hyperglycemia, ketonuria or ketonemia, and a pH of less than 7.3 or a bicarbonate level of less than 15 mmol/L were all considered indicators of DKA. The glycaemic outcome of HbA1c>7%, coded Yes/No, was the secondary endpoint. It was assessed once again at the patient's most recent (final) visit. The following risk factors were included in the predictive features of the model: DKA at diagnosis (Yes/No/Missing), age at diagnosis (in years), DKA at previous visit, gender (Male/Female), insulin regimen at previous visit.
(MDI/CSII/Others/Missing), Index of Relative Socio-economic Index (IRSD) based on postcode of diagnosis, de-identified center ID, center type (Paediatric/Adult), duration of diabetes at previous visit, BMI percentile at previous visit, counts of comorbidity at previous visit, HbA1c% at previous visit, country of birth and use of CGM device before last visit (Yes/No). In order to establish a temporal relationship between the predictive features and outcomes, we attempted to use the individual's prior clinic measurements as predictors. To make it easier to investigate non-linear correlations, all continuous variables were divided into quartiles. Missing data was kept as a separate category within the variable to maintain the overall sample size. Since it was impossible to determine if the missing data was absent at random from our data, we decided against using multiple imputation. There was more missing visit data for patients who experienced DKA since a much larger percentage of them had only one visit documented in our database during the study period. For adults over the age of 18, height and weight were taken during each clinic visit using the standard procedures for each clinic. After that, measurements were used to calculate BMI. The height and weight data of children aged 2 to 18 years were used to calculate their BMI standard deviation or z-scores, which are measures of relative weight adjusted for the child's age and sex. The 2000 reference scale from the Centers for Disease Control and Prevention was used for the calculations. Both at the time of diagnosis and during follow-up clinic visits, HbA1c levels were measured using standardized techniques using either laboratory-performed HbA1c or desk-top analyzers. The IRSD is a proxy for socioeconomic status (SES) at the local level. A postcode with a low IRSD score has a higher percentage of people who are relatively disadvantaged.

3. MACHINE LEARNING MODELS

In our investigation, we used the following nine machine learning techniques: (1) Naïve Bayes (NB), a low variance, high bias estimator based on conditional probability and the Bayes Theorem, (2) In order to regularize parameters ranging between L1 (Lasso) and L2 (Ridge regression), the Generalized Linear Model (GLM), a statistical technique that is an extension of the standard linear regression model, is fitted by maximising the log-likelihood function with an elastic net penalty alpha parameter. (3) The Generalized Linear Model includes the Logistic Regression (LR) Model as a subset. (4) Deep Learning (DL): a technique based on a multi-layer artificial neural network that features numerous hidden layers made up of neurons and sophisticated features like adaptive learning rate and L1 or L2 regularization, and is trained using stochastic gradient descent using back-propagation, (5) Fast Large Margin (FLM), a model that optimizes the cost parameter, C, the error term's penalty parameter, is based on the linear support vector learning technique and is helpful for large sparse data with a large number of observations and features., The sixth machine learning technique is the Support Vector Machine (SVM), which creates a set of hyperplanes in a high-dimensional space for categorization. Effective

separation is achieved by determining the hyperplane that keeps the largest distance from each class's closest training data points, also known as the "functional margin." Decision Trees (DT), these represent a non-parametric binary classification method in which one tree is used to evaluate threshold points to find the one that yields the highest consistency within new subgroups while attempting to minimize impurity in the node. Generally speaking, a larger margin corresponds to a reduced generalization error for the classifier. When more divisions are unable to improve the current homogeneity, the process is over. (8) Random Forests (RF), which are made up of several DTs that were trained using the ensemble bagging approach. The final categorization is decided after a vote by all DTs. Data is continually divided into child nodes by DTs, resulting in binary trees. Trees that are not functioning well can be pruned. RF is a suggested classifier because it can manage noise, outliers, underfitting, and overfitting. Gradient Boosted Trees (GBT), XGBoost, and a well-known classification ensemble, through iterative improvements of a single tree model, this machine learning technique improves a model. It modifies example weights according to prior expectations at each iteration. A weighted combination of all the models produced is the final model. The gradient of the function generated by the errors is used to fine-tune the training parameters.

3.1. Optimal feature and model parameter selection

The 60% training dataset served as the basis for the optimal feature selection process (Fig. 1). To reduce classification error, a seven-step cross-validation procedure was used, evaluating several complicated functions and optimization heuristics on 90% training and 10% validation datasets. Following feature optimization, the features were applied to the entire training dataset, and using a 3-fold cross-validation stratified sampling technique with 90% training and 10% validation sub-sets, the parameters for each machine learning model were optimized using this 60% dataset.

3.2. Model evaluation

An assessment of the model's prediction accuracy and other performance metrics, which differ based on the particular classification task at hand, is given in this section. We use a 40% hold-out dataset, which hasn't been used in any of the previous model optimization procedures, to evaluate performance. We assess performance across seven different subsets using this hold-out set as the input for a multi-hold-out-set validation (Fig. 1). The following performance metrics were computed in order to quantitatively compare the efficacy of the different machine learning approaches: classification error (the total percentage of subjects incorrectly classified), sensitivity (the total percentage of true positives, such as those with DKA), specificity (the total percentage of true negatives), F-measure (the combination of sensitivity and precision), total time spent developing and validating the model, and lastly, the primary performance metric: AUC. The most dependable performance metric that combines sensitivity and specificity is the AUC. AUC is computed for the validation dataset, which allows for external validation, and values nearer 1 signify higher discriminatory qualities. Values over 0.8 are regarded as good performance. Rapidminer Studio V10.1(14) and Stata V18.0 (Stata Corp, College Station, TX, USA) were used for data analysis.

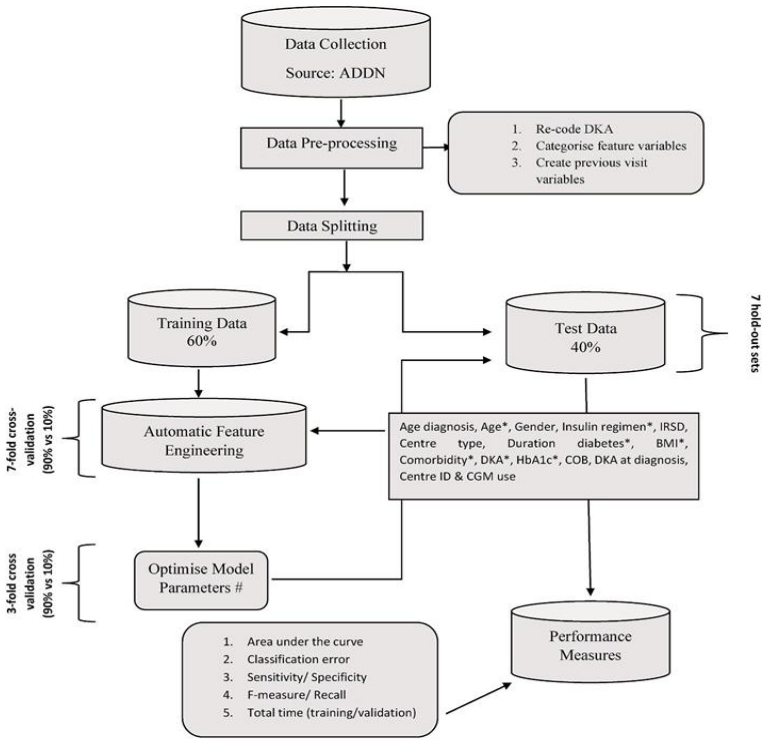

Fig. 1. Conceptual framework of data preparation, model development and validation

Table 1 Clinical characteristics of cohort by DKA status at last visit

DKA at last visit	Age at diagnosis	Age at previous visit	Gender	Insulin regimen	Duration of diabetes	HbA1c previous	DKA at diagnosis
0	12	18	Male	Others	7.5	8.41	Missing
0	18	14	Male	Missing	8.9	7.46	No
0	9	18	Male	CSII	4.4	7.55	Yes
0	5	13	Male	CSII	6.3	8.26	Yes
0	7	19	Female	BD	1.8	6.77	Missing
0	9	14	Female	BD	5.4	6.85	Yes
0	4	18	Male	MDI	8.1	7.94	Missing
0	17	17	Male	Missing	11.1	9.02	Yes
0	4	12	Male	Others	0.5	8.48	No

4. RESULTS AND DISCUSSION

Our study included 10,968 individuals with type 1 diabetes who had available DKA information at the time of their visit to DDN centers between January 2012 and May 2023 (Fig. 2). 367 (2.9%) of the 10,868 participants in our research reported having at least one DKA episode during their most recent visit (Table 1). The cohort was followed up in the trial for a median of 3.4 years (IQR 1.3–5.7 years), with the majority (50.1%) being between the ages of 15 and 21 at the last visit. The clinical and demographic characteristics of those who reported DKA and those who did not differed significantly. Those who were under 15 years old at the last visit, had been diagnosed and had had diabetes for less than 5 years, had lower missing values for BMI z-score and HbA1c at the last visit. Comorbidities were present in a substantially greater percentage of DKA patients at the prior visit. Based on the validation dataset, Table 2 presents an overview of the performance of the different machine learning techniques. With an AUC of 0.887, DL outperformed GLM and LR in terms of the major performance metric, AUC. AUC < 0.5 indicates that DT did not perform properly. Fig. 3 displays the AUC for comparing the different machine learning models for DKA.

A low classification error rate of 0.9%, high sensitivity of 100%, specificity of 50%, and F-measure of 99.6% were also provided by the DL. It took about eleven minutes to compute, which is a respectable amount of time. At the price of AUC (0.853), the next best-performing RF model offered a little greater specificity of 53%. Two input blocks with 50 layers each that were activated by the rectifier function were the ideal model parameters for the DL model. The following variables were measured at prior visits: DL used age, duration of diabetes, Center, BMI z-score, HbA1c, CGM usage, insulin regimen, and DKA. Out of five distinct machine learning models, DKA at the prior visit was the most frequently chosen attribute. Center, which was chosen by four machine learning models, was the next most popular feature.

An overview of how well the different machine learning techniques performed in predicting HbA1c >7% is provided. With an AUC of 0.884, SVM outperformed the others based on the AUC metric. Several other models with somewhat lower AUCs came right after this. Fig. S1 displays the AUC for comparing the different machine learning models for glycemic levels greater than 7%. Additionally, the SVM yielded the highest specificity of 95.6%, the lowest classification error rate of 11.6%, and the F-measure of 58.9%. At 39 minutes, it was the second slowest in terms of computing time. Although the LR model was the fastest, its performance was somewhat compromised (AUC = 0.876). Gamma = 0.005 and C = 10 were the ideal model parameters for SVM. Regarding the features chosen, SVM made use of the following variables: center, number of comorbidities, BMI z-score at the prior visit, and HbA1c at the prior visit. The HbA1c from the prior visit was chosen by all of the machine learning models, and it was evident that this was a powerful predictor of the upcoming HbA1c. With only one feature, HbA1c, NB and DT seemed to be the most economical models. They also had similar AUCs, making them potential substitutes for practical use.

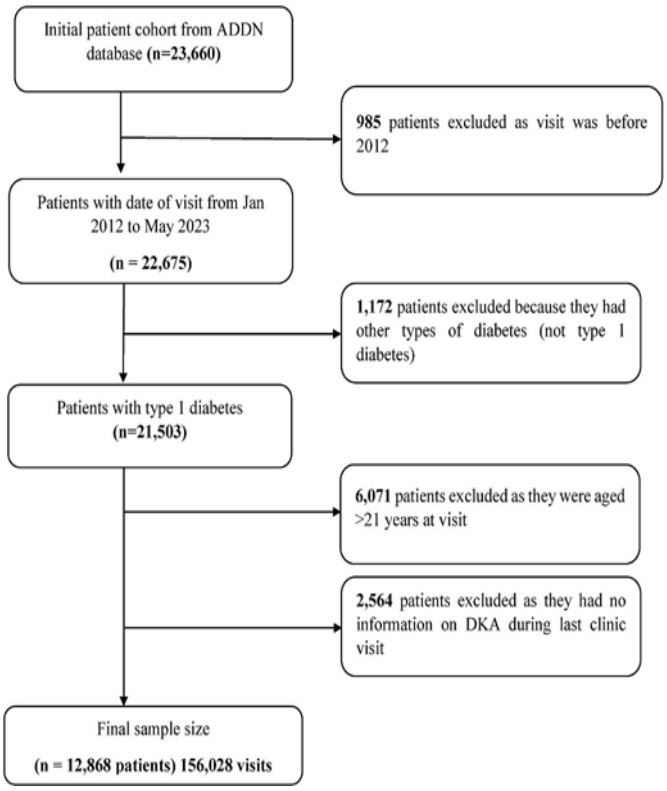

Fig. 2. Flowchart of patient inclusion/exclusion criteria and final cohort for stud

Based on the AUC determined on a validation dataset, our work has determined that it is possible to apply cutting-edge machine learning models to categorize people according to whether they report DKA during their clinic visits with good discriminatory qualities. In particular, we discovered that the DL model had the highest AUC of 0.887 with good sensitivity when it came to the following factors: age, duration of diabetes, BMI z-score, HbA1c, CGM use, insulin regimen, DKA during a prior visit, and center. Our discovery that the use of CGM was predictive of DKA is in line with previous research, which found that hospitalizations for hypoglycemia or ketoacidosis were linked to "time in range" data from CGM and that longitudinal measurements of CGM data were also highly predictive of identifying patients at risk of DKA. Our study has discovered a center effect in addition to what has already been noted in the literature. Clinical practice, healthcare resources, and other unmeasured variations in patient characteristics are examples of unmeasured or "latent" center level effects that can be quantified using the Center, an objective metric. In terms of differentiating individuals for this outcome, we discovered that the SVM model offered a decent match for predicting HbA1c >7% (AUC = 0.884). The HbA1c, comorbidities, BMI z-score at the prior visit, and center were among the important features that the model used in the prediction.

Previous studies have demonstrated baseline HbA1c, age, BMI, and duration of diabetes to be major drivers of subsequent HbA1c, thus our conclusion that clinical factors like HbA1c and BMI evaluated at a prior visit are a strong regulator of glycaemic outcome is not surprising. These findings also demonstrate that it is possible (computationally) to apply these models in practice with a fair degree of accuracy to help clinicians identify patients who may be at risk for developing DKA or who have a HbA1c of greater than 7% and who may benefit from targeted education, especially in the area of DKA prevention. AUC, which measures the discriminatory capacity of the model as a trade-off between sensitivity and specificity, served as the primary performance metric in our investigation. This would account for the model's high specificity and low sensitivity for HbA1c >7% and its high sensitivity and low specificity for DKA. Depending on the goal of the clinical implementation, models can be chosen using a variety of criteria (such as maximizing sensitivity or true positives).

Our models outperformed other relevant studies, such as a case-control analysis of a large analysis that compared the effectiveness of multiple machine learning techniques utilizing electronic medical information [20]. With AUCs ranging from 0.72 (partial out of sample validation cohort) to 0.85 (full out of sample validation cohort), our AUCs were also higher than those of a large electronic medical record analysis that predicted admissions linked to DKA within 180 days. The type of machine learning models examined and the study strategy (case-control versus our prospective cohort) may be the causes of the discrepancies. Elastic net regularization was one of the strategies examined in earlier related studies. For our investigation, we were able to take advantage of the longitudinal character of a sizable, real-world multi-center clinical dataset spanning two nations. Our analysis found the center effect in addition to what was previously noted in the literature, suggesting that these models must be tailored to certain geographic contexts.

Table 2 Comparison of accuracy, error rate and total time for different machine learning methods

Model	AUC	SD	Classification Error	Sensitivity	Specificity	F-Measure
Naive Bayes (NB)	0.811	0.019	1.2 %	100.0 %	24.0 %	99.4 %
Generalized Linear Model (GLM)	0.853	0.035	0.8 %	100.0 %	51.7 %	99.6 %
Logistic Regression (LR)	0.837	0.017	0.9 %	99.9 %	53.4 %	99.6 %
Fast Large Margin (FLM)	0.662	0.023	1.2 %	100.0 %	22.5 %	99.4 %
Deep Learning (DL)	0.887	0.033	0.9 %	99.9 %	50.3 %	99.6 %
Decision Tree (DT)	0.5	0.001	1.6 %	100.0 %	0.0 %	99.2 %
Random Forest (RF)	0.615	0.029	1.6 %	100.0 %	0.0 %	99.2 %
Gradient Boosted Trees (GBT)	0.636	0.037	1.6 %	100.0 %	0.0 %	99.2 %
Support Vector Machine (SVM)	0.536	0.049	1.4 %	99.7 %	24.4 %	99.3 %

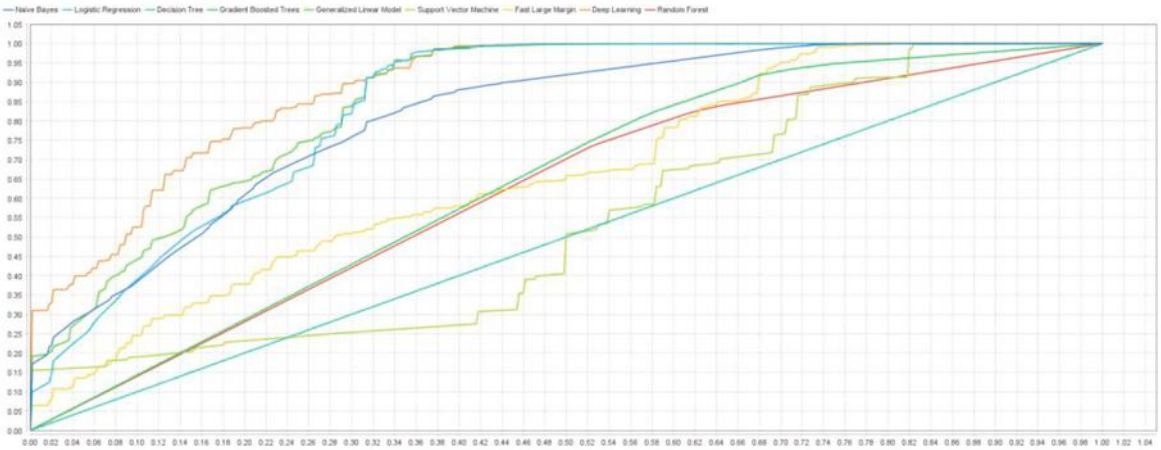

Fig. 3. Area under the curve comparing the various machine learning models for predicting DKA

5. SUGGESTIONS

To offer reliable and accurate estimations, a model must be developed from a training dataset and externally validated on fresh data before being used in clinical practice. We conducted an analysis of a sizable multi-center dataset with this goal in mind, enabling accuracy and generalizability in our findings. Implementing such models into standard clinical care data collection systems in diabetes clinics is the last phase. A collaborative shared decision-making process with clinicians and healthcare workers to take preventive measures and interventions to avoid this serious complication from occurring, as well as to avoid the healthcare costs and other psycho-social impact on individuals associated with the potential hospitalizations, can be made possible by machine learning models' ability to accurately predict which individuals may present with DKA (or HbA1c >7%) during the next clinic visit. Physicians should think about screening programs to prevent diseases including infections and mental health issues, which can lead to the development of DKA. In order to prevent DKA presentation, targeted case management and education about glycemic management, including ketone testing, extra insulin administration, and self-monitoring of blood glucose (SMBG), could be given priority for the most vulnerable groups. With prompt education and guidance, having access to phone assistance throughout illness can avoid admittance. The correct estimation of correction insulin boluses and the matching of insulin needs with carbohydrate intake can both be facilitated by education. In order to prevent DKA and improve glycaemic levels as determined by HbA1c, it is beneficial to encourage and involve young people in avoiding missed bolus doses or in catching up on missed bolus doses, particularly to avoid missing basal insulin.

Early detection of DKA risk and prevention of DKA treatment and related hospitalization expenses have financial advantages. Individuals can also benefit from things like less time away from work or school and less negative psycho-social effects. Our models may also be helpful in clinical practice when it comes to giving patients access to more recent targeted medications. Based on our model's predictions, it might be feasible to recruit participants for randomized controlled trials that are integrated into the clinical database according to their risk of getting DKA. A sizable multi-center, multi-country study with a sizable sample size served as the foundation for our investigation. The goal of this strategy is to lessen inequalities in healthcare accessibility and standards. About 40% of children and young adults in both nations have type 1 diabetes, and the DDN is essential to their care. Furthermore, healthcare facilities reach out to both urban and rural people through established outreach initiatives. As a result, our data offers a representative sample that captures the two countries' disparate health care systems and demographics. The use of numerous machine learning models and their

assessment using a robust design framework in terms of feature selection and model parameter selection via cross-validation constitute another strength of our research.
We admit that the list might not be all-inclusive in terms of the predictive factors, and we omitted laboratory measurements. Although the predictive performance of the models may be enhanced by these factors, systematic missingness, especially among non-DKAs in terms of the clinicians who prescribe the tests, may lead to bias in the data' analysis and interpretation. In a similar vein, 2564 people were not included in our study because they did not have DKA at clinic visits. In contrast to those with available DKA data, those with missing DKA values are more likely to be significantly older, have had diabetes for a longer period of time, be on an MDI regimen, live in moderately disadvantaged areas, be from adult centers, and have a higher proportion of people born abroad, according to a comparison of the characteristics of those with missing DKA at their most recent visit. Our study's machine learning models may perform worse than expected, especially when it comes to underestimating the more severe category. We also admit that there was a substantial amount of missing data for certain variables, such as insulin regimen, BMI, DKA at diagnosis, and IRSD. In the case of data that might not be missing at random, multiple imputation may be a viable remedy and is a topic for future study. Since this is a computationally costly operation, not every parameter in each model was optimized. As a result, our research's findings are conservative, and there may be room for additional model performance improvement.

6. CONCLUSION

In this study, we successfully developed and externally validated machine learning models capable of predicting DKA episodes and elevated HbA1c levels (>7%) in individuals with type 1 diabetes using a comprehensive, real-world multi-center dataset. The Deep Learning model achieved superior performance in predicting DKA with an AUC of 0.887, while the Support Vector Machine model excelled in predicting glycemic dysregulation with an AUC of 0.884. Significant predictors included clinical features such as prior DKA history, BMI z-score, CGM use, HbA1c levels, and diabetes duration, reflecting the multifactorial nature of risk assessment in T1DM management. These findings affirm the potential of machine learning models to enhance risk stratification and clinical decision-making in routine diabetes care. By facilitating the early identification of individuals at high risk for DKA or poor glycemic control, these models offer a pathway toward targeted educational interventions and resource allocation. Future implementation into clinical databases could further support proactive care models, aiming to reduce complications, improve patient outcomes, and optimize healthcare delivery systems across diverse demographic settings.

REFEENCES

[1] American Diabetes Association Professional Practice Committee, “14. Children and adolescents: standards of care in diabetes—2024,” Diabetes Care, vol. 47, pp. S258–S281, 2024.

[2] G. E. Umpierrez et al., “Hyperglycemic crises in adults with diabetes: a consensus report,” Diabetes Care, vol. 47, pp. 1257–1275, 2024.

[3] D. Tomic, M. E. Craig, D. J. Magliano, and J. E. Shaw, “Reasons for hospitalisation in youth with type 1 diabetes, 2010–2019,” Diabetic Medicine, vol. 40, no. 1, p. e15218, 2023.

[4] I. Contreras and J. Vehi, “Artificial intelligence for diabetes management and decision support: literature review,” Journal of Medical Internet Research, vol. 20, p. e10775, 2018.

[5] S. B. Nagaraj, G. Sidorenkov, J. F. M. van Boven, and P. Denig, “Predicting short- and long-term glycated haemoglobin response after insulin initiation in patients with type 2 diabetes mellitus using machine-learning algorithms,” Diabetes, Obesity and Metabolism, vol. 21, no. 12, pp. 2704–2711, 2019.

[6] Y. Fan et al., “Development of machine learning models for mortality risk prediction after cardiac surgery,” Cardiovascular Diagnosis and Therapy, vol. 12, no. 1, pp. 12–23, 2022.

[7] J. I. Wolfsdorf et al., “ISPAD clinical practice consensus guidelines 2018: diabetic ketoacidosis and the hyperglycemic hyperosmolar state,” Pediatric Diabetes, vol. 19, Suppl. 27, pp. 155–177, 2018.

[8] C. L. Ogden et al., “Centers for disease control and prevention 2000 growth charts for the United States: improvements to the 1977 national center for health statistics version,” Pediatrics, vol. 109, no. 1, pp. 45–60, 2002.

[9] I. Mierswa, Data Science, Machine Learning, Predictive Analytics: RapidMiner Studio, Version 10.1, 2019. [Online]. Available: https://rapidminer.com

[10] R. E. Fan et al., “LIBLINEAR: a library for large linear classification,” Journal of Machine Learning Research, vol. 9, pp. 1871–1874, 2008.

[11] A. Earnest, G. A. Tesema, and R. G. Stirling, “Machine learning techniques to predict timeliness of care among lung cancer patients,” Healthcare (Basel), vol. 11, no. 2, 2023.

[12] A. Garavand et al., “Efficient model for coronary artery disease diagnosis: a comparative study of several machine learning algorithms,” Journal of Healthcare Engineering, vol. 2022, Article ID 5359540, 2022.

[13] E. M. El Malahi et al., “Relationship between time in range, glycemic variability, HbA1c and complications in adults with type 1 diabetes mellitus,” Journal of Clinical Endocrinology and Metabolism, vol. 107, no. 2, 2021.

[14] S. L. C. Lebech and C. Bender, “Development of machine learning models for the identification of elevated ketone bodies during hyperglycemia in patients with type 1 diabetes,” Diabetes Technology & Therapeutics, vol. 26, no. 6, pp. 403–410, 2024.

[15] L. Li et al., “Performance assessment of different machine learning approaches in predicting diabetic ketoacidosis in adults with type 1 diabetes using electronic health records data,” Pharmacoepidemiology and Drug Safety, vol. 30, no. 5, pp. 610–618, 2021.

[16] D. D. Williams 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, vol. 8, p. e47592, 2023.