diabetes research and clinical practice 190 (2022) 109982

Contents lists available at ScienceDirect

Diabetes Research and Clinical Practice

journal homepage: www.journals.elsevier.com/diabetes-research-and-clinical-practice

Predicting poor glycemic control during Ramadan among non-fasting patients with diabetes using artificial intelligence based machine learning models

Imane Motaib [a,*], Faiçal Aitlahbib [b,c], Abdelhamid Fadil [b], Fatima Z.Rhmari Tlemcani [a], Saloua Elamari [a], Soukaina Laidi [a], Asma Chadli [a]

[a] Department of Endocrinology Diabetology Metabolic Disease and Nutrition, Cheikh Khalifa International University Hospital, Faculty of Medicine, Mohammed VI University of Health Sciences (UM6SS), Casablanca, Morocco
[b] Hassania School of Public Works, Casablanca, Morocco
[c] Office Chérifien des Phosphates (OCP), Casablanca, Morocco

A R T I C L E I N F O

Keywords:
Ramadan
Diabetes
Glycemic control
Machine learning
Artificial intelligence

A B S T R A C T

Aims: This study aims to predict poor glycemic control during Ramadan among non-fasting patients with diabetes using machine learning models.
Methods: First, we conducted three consultations, before, during, and after Ramadan to assess demographics, diabetes history, caloric intake, anthropometric and metabolic parameters. Second, machine learning techniques (Logistic Regression, Support Vector Machine, Naive Bayes, K-nearest neighbor, Decision Tree, Random Forest, Extra Trees Classifier and Catboost) were trained using the data to predict poor glycemic control among patients. Then, we conducted several simulations with the best performing machine learning model using variables that were found as main predictors of poor glycemic control.
Results: The prevalence of poor glycemic control among patients was 52.6%. Extra tree Classifier was the best performing model for glycemic deterioration (accuracy = 0.87, AUC = 0,87). Caloric intake evolution, gender, baseline caloric intake, baseline weight, BMI variation, waist circumference evolution and Total Cholesterol serum level after Ramadan were selected as the most significant for the prediction of poor glycemic control. We determined thresholds for each predicting factor among which this risk is present.
Conclusions: The clinical use of our findings may help to improve glycemic control during Ramadan among patients who do not fast by targeting risk factors of poor glycemic control.

1. Introduction

Ramadan is the ninth month of the lunar calendar. Muslims around the world fast during this month. They refrain from eating and drinking from sunrise to sunset. During Ramadan, the daily lifestyle routine of Muslims undergoes major changes, particularly in terms of meal timing and content, sleep schedule, and physical activity [1,2,3].

In patients with diabetes, these lifestyle modifications impact their glycemic control whether they fast or not during the month of Ramadan [4,5]. Indeed, fasting in patients with diabetes can cause complications such as hyperglycemia, ketoacidosis, hypoglycemia, dehydration, and thrombosis[5]. This implies that the management of diabetes during the month of Ramadan requires close follow-up and therapeutic adjustment.

To prevent acute complications related to fasting during Ramadan, the International Federation of Diabetes (IDF) has published in its latest recommendations a fasting risk calculator. The aim is to identify the patients with the highest risk of complications during fasting. The IDF also emphasized the usefulness of artificial intelligence and particularly machine learning algorithms in the future for risk stratification [6]. While diabetes management in patients who choose to fast during Ramadan is well codified [6], the risk of glycemic control deterioration in non-fasting patients with diabetes during this month is not well documented. Indeed, several studies have shown that the fasting period is followed by a feasting period. Large meals with excessive caloric and carbohydrate intake are consumed in patients with diabetes even if they do not fast which could cause poor glycemic control [7].

* Corresponding author at: Faculty of Medicine, Mohammed VI University of Health Sciences (UM6SS), 82403 Casablanca, Morocco.
E-mail address: imotaib@um6ss.ma (I. Motaib).

https://doi.org/10.1016/j.diabres.2022.109982
Received 10 March 2022; Received in revised form 17 June 2022; Accepted 4 July 2022
Available online 6 July 2022
0168-8227/

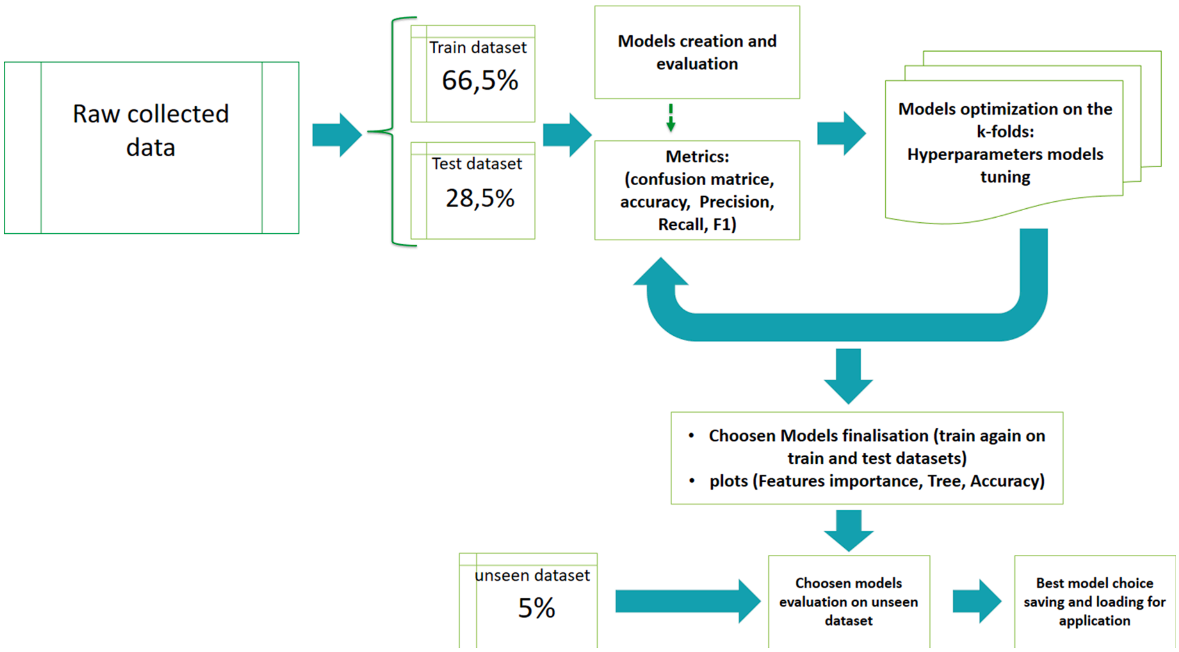

Fig. 1. Flowchart explaining methodology of using machine learning in the study.

Machine learning is a discipline of artificial intelligence that aims to develop learning by computers from their previous experience. This has enabled the development of algorithms that can predict events, with an amount of accuracy. In the medical field, especially in diabetes research, Machine learning (ML) has been widely introduced. Several studies have investigated the application of machine learning methods such as random forest, support vector machines (SVM), and naïve Bayes to predict the risk of diabetes and its complications [8,9,10,11].

Similarly to the machine learning models developed to predict hyperglycemic and hypoglycemic excursions in patients who fast during Ramadan[12], other models can be applied to predict glycemic imbalance in non-fasting patients with diabetes. Machine learning algorithms can identify the clinical parameters that cause poor glycemic control among those patients. Additionally, the deployment of these models into apps will help physicians to identify patients at risk of poor glycemic control and to propose an early and personalized management of those patients.

The aim of this study is to predict poor glycemic control during Ramadan among non-fasting patients with diabetes followed up in Sheikh Khalifa Ibn Zaid Hospital using machine learning models.

2. Subjects and methods

2.1. Study design and protocol

We carried out a prospective cohort that included all adult patients with diabetes who did not fast during three months of Ramadan (May-June 2018, May-June 2019, and April-May 2021). The study was conducted in the Department of Endocrinology and Diabetology in Sheikh Khalifa Ibn Zaid Hospital, Casablanca, Morocco. Pregnant women and patients who did not fully complete the follow-up during the study period were excluded. A week before the start of Ramadan, we conducted the first consultation during which we collected the following data: demographic characteristics (including age and gender), diabetes history (including type and duration of diabetes, diabetes medications, and degenerative complications), anthropometric parameters (weight, body mass index and waist circumference), biological characteristics including metabolic parameters (glycated haemoglobin, fasting blood glucose and lipid profile). An assessment of dietary intake before Ramadan was performed using the 24-hour dietary recall method and a food frequency questionnaire adapted for Moroccan adults [13]. Fasting risk stratification according to the IDF-DAR recommendations [14]was also conducted. Patients were stratified into 3 categories (Very high risk, High risk, Moderate/low risk) according to several criteria. Patients who did not fast were in the very high risk and the high-risk categories. These categories included patients with type 1 diabetes, or type 2 diabetes with persistent poor glycemic control, patients who presented acute complications within the three months before Ramadan (such as severe hypoglycemia, or ketoacidosis), a history of recurrent hypoglycemia or hypoglycemia unawareness, patients with chronic kidney disease, or macrovascular complications, patients performing intense physical labor, patients with treatment with drugs that may affect cognitive function and ill elderly patients.

A self-monitoring of blood glucose using a blood glucose logbook was performed by all patients during all the month of Ramadan, patients collected capillary blood glucose before and 2 h after each meal. A second consultation was performed the second week of Ramadan to assess physical activity level during Ramadan, using the Global Physical Activity Questionnaire (GPAQ2) [15]. Dietary intake and number of meals, as well as the occurrence of complications such as hyperglycemia (defined as capillary blood glucose rate higher than 16.5 mmol/l) and hypoglycaemia (defined as capillary blood glucose rate below 3.85 mmol/l), were also evaluated. To determine anthropometric and metabolic parameters evolution, we conducted a third consultationduring the first week after the end of Ramadan, to re-evaluate the parameters previously mentioned. The study complies with the Declaration of Helsinki, informed consent was obtained for all patients prior to inclusion. Data confidentiality and patient anonymity were maintained at all stages of the study. We deleted patient-identifying information before analysing the database.

2.2. Definition of poor glycemic control

- Poor glycemic control was defined as an increase of glycated hemoglobin level above 0.5% of its pre-Ramadan level with glycated hemoglobin above the glycemic target of patients according to ADA recommendations[16].

2.3. Exploratory data analysis (EDA)

The dataset holds 154 records with 43 attributes such as age, gender, type of diabetes, baseline glycated hemoglobin, hypoglycemia during Ramadan, poor glycemic control. For data exploration, we used histograms for each attribute to evaluateits distribution in the dataset.

Supplemental Figure S1 showed that the majority of continuous data had a Gaussian distribution while some variables had a skewed distribution. We notice that the target parameter "poor glycemic control" is slightly unbalanced (56% no versus 44% yes).

To determine the strength of the linear relationship between the variables of the dataset, we carried out the correlation coefficient analysis using Pearson correlation. **Supplemental Figure S2** shows the scatterplot correlation coefficient of the attributes of the dataset. We notice that the target parameter "poor glycemic control" has no linear correlation with the other variables, which requires the use of nonlinear machine learning methods

2.4. Data preprocessing

The dataset has no missing value. To predict poor glycemic control during the month of Ramadan we created the variable "poor glycemic control" from the difference between glycated hemoglobin before and after Ramadan. This variable is a binary class corresponding to whether a patient has poor glycemic control during Ramadan or not.

Exploratory data analysis showed that the target parameter "poor glycemic control"is slightlyunbalanced, therefore we used the Synthetic Minority Oversampling Technique (SMOTE). This technique generates randomly new instances of the minority class to obtain balanced data to get better training for prediction.

In addition, to increase the performance of machine learning algorithms predictions, for the categorical variables that have more than two possible values (such as diabetes medications, degenerative complications), we applied a one-hot encoding technique which recorded them as multiple variables each having two possible values.

We also applied a normalization of the data set using the Z-score method to put different variables on a same scale. The Z-score is calculated according to the following formula $Z=(x-\mu)/\sigma$, where $\times$ is the original feature vector, μ is the mean and σ is the standard deviation of the feature vector.

2.5. Data splitting and metrics

We randomly divided the data into two parts as represented in Fig. 1. A first part representing 95% of the data has been divided into 2 subparts (respectively 70% and 30%), the first subpart represents the training set (66.5% of the data) and the secondrepresents the test-set (26.5%). The models of machine learning were trained on the train dataset and evaluated through the metrics on the test dataset. The remaining 5% of the database, is called the unseen data and was used for the last step of validation after the choice of the best model is made. The finalization of this selected model was done by training it on the whole first part of the data (including the train-set and test-set) and then evaluated on the unseen data. The optimization of the models was done by tuning the hyperparameters of the model using the k-fold cross-validation method (with $k = 10$ as the number of subsets to evaluate the model) [17]. The performance measurement of the different machine learning models was conducted using the following metrics:

- **The confusion matrix**: represented in **Supplemental Table S3**, is a table composed of the following parameters:
  - True positives (TP): cases we predicted positive and which are really positive.
  - True negatives (TN): cases we predicted negative and which are really negative.
  - False positives (FP): cases we predicted positive, but they are actually negative.
  - False negatives (FN): cases we predicted negative, but they are actually positive.

It is a performance measurement technique used to evaluate the performance of the model on the test-set for which the true values are already known. It also allows calculating other metrics such as Accuracy, Area under curve, recall, and precision.

- **Accuracy:** which is defined as a performance metric that calculates the ratio of true positives and true negatives to all positive and negative observations. Accuracy shows the probability that the machine learning model will correctly predict a classification out of the whole number of times it made predictions.
- **Area under curve (AUC):** defined from **Receiver Operating Characteristics** curve, evaluates how much the model is able to distinguish between classes. The Higher the AUC, the better the model is at predicting the target parameter.
- **Recall:** defined as the ability of a classification model to find all the relevant cases in a dataset which is calculated using the following formula: $Recall = TP/\ (TP + FN)$.
- **Precision:** is the ability of a classification model to identify only the relevant data points. It is calculated as following: $Precision = TP/\ (TP + FP)$
- **F1 score:** which is the harmonic mean of precision and recall. The formula for calculating this score is: $F1\ score = 2x(precision + recall)/\ (precision + recall)$

  A model is considered acceptable and good when the value of the metrics is greater than 0.7.

2.6. Supervised machine learning models and feature-selection techniques

To predict poor glycemic control, we used and compared the following supervised machine learning (ML) algorithms: Logistic Regression (LR), SupportVector Machine (SVM), Naive Bayes (NB), K-nearest neighbour (KNN), DecisionTree (DT), Random Forest, Extra Trees Classifier and Catboost. These ML models are widely used in the literature[8]. Brief definitions of each model are represented in **Supplemental Table S4.**

As a first step, we trained these ML classifiers with all the features of the dataset, as the target parameter "poor glycemic control" has no linear correlation with the other variables. Then as a second step, we selected the features that obtained the higher score using Recursive feature elimination (RFE) with the 3 best performing ML models. The higher the score, the more important is the feature toward poor glycemic control prediction. This process improves the accuracy of ML classifiers and reduces the risk of overfitting.

2.7. Clinical application

As a last step, and to better assess the clinical significance of our results, we conducted several simulations with the best performing machine learning model for poor glycemic control prediction using features that were found as main predictors for the model. Simulations were conducted using mean first quartile, median third quartile, and maximum for each continuous variable, and for categorical variables classes have been kept. Finally, we deployed our machine learning calculator model in a localhost using a Streamlite powered framework (www.streamlit.io), which is an open-source app framework for machine learning to predict poor glycemic control.

2.8. Statistical analysis

All statistical analysis including dataset preprocessing, splitting,

Table 1
Comparison of training K-fold results of machine learning models to predict poor glycemic control during Ramadan among non-fasting patients with diabetes using train data-set and evaluation of the top-3 models using test-set.

Machine learning Model	Data	Trainset					Test-set				
	Metrics	**Accuracy**	**AUC**	**Recall**	**Prec.**	**F1**	**Accuracy**	**AUC**	**Recall**	**Prec.**	**F1**
SVM - Linear Kernel		0.8945	0.9010	0.9200	0.8871	0.8944	0.7955	0.7723	0.6875	0.7333	0.7097
Naive Bayes		0.7736	0.7603	0.8350	0.7421	0.7796	–	–	–	–	–
Random Forest Classifier		0.8745	0.9342	0.8800	0.8781	0.8716	–	–	–	–	–
Extra Trees Classifier		0.8718	0.9540	0.8750	0.8800	0.8669	–	–	–	–	–
LogisticRegression		0.9018	0.9610	0.8950	0.9083	0.8938	0.7955	0.8504	0.6875	0.7333	0.7097
CatBoost Classifier		0.8936	0.9613	0.9000	0.8931	0.8916	0.7273	0.8795	0.7500	0.6000	0.6667
DecisionTree Classifier		0.8045	0.8189	0.7950	0.8117	0.7926	–	–	–	–	–
K Neighbors Classifier		0.8636	0.9410	0.9400	0.8252	0.8704	–	–	–	–	–

statistical learning methods, and hyperparameters finetuning were performed using Pycaret Auto ML library version 2.0 on Python version 3.6.

3. Results

3.1. Clinical and biochemical characteristics

The study cohort consists of 154 patients. The median (interquartile range) age of patients was 64 (57.25–71) years, 60.38% of patients were male and 92.8% of patients had type 2 diabetes. The medians (IQR) of glycated hemoglobin before and after Ramadan were respectively 8.45% (69 mmol/mol) (IQR: 7.7–10% (61–86 mmol/mol)) and 8.55%(70 mmol/mol)(IQR: 7.8–12.5% (62–108 mmol/mol)). The medians (IQR) of baseline BMI and waist circumference were respectively 28 kg/m^2 (IQR: 24.9–31.9) and 99 cm (IQR: 91–105). The median of Caloric intake was 1641.12Kcal (IQR: 1454–1915.16). The prevalence of poor glycemic control among non-fasting patients with diabetes during Ramadan was 52.6%. All Other Characteristics of patients are represented in **Supplemental TableS5.**

3.2. Comparison of supervised Machine learning models

In this study, outcomes were achieved by applying eight classification algorithms (logistic regression, support vector machine (SVM), naive Bayes, K-nearest neighbor (KNN), decision tree, random forest, Extra Trees classifier, and Catboost)) to display maximize accuracy in poor glycemic control prediction.

As a first step, we used all the features of the dataset. Models' comparison with k-fold showed that the best performing model using the train-set was Logistic regression with an accuracy rate of 0.90 and an AUC rate of 0.96, followed by SVM (accuracy = 0.894) and Catboost (accuracy = 0.893, AUC = 0.96)**.** Table 1 shows the results of the performance evaluation of the models with all metrics using the train set. As represented in **Table 3,** the evaluation of these three ML models, using the test-set showed that the best performing test was logistic regression (Accuracy = 0.7955, AUC = 0.8504). The last validation of logistic regression model with the unseen data showed that the accuracy rate is 0.875, AUC rate is 0.875, recall rate is 0.75, precision rate is 1 and F1 rate is 0.875.

3.3. Variable importance

The analysis of variables importance based on RFE showed that caloric intake evolution is the top-ranked variable by the top three of models conducted with the train-set (LR, SVM, and Catboost). The top ten of variables' importance according to each model are represented in **Supplemental Table S6**. Among these variables, seven were selected as the most significant variables for the prediction of poor glycemic control during Ramadan. These variables were caloric intake evolution, gender, baseline caloric intake, baseline weight, BMI variation, waist circumference evolution, Total Cholesterol level after Ramadan.

3.4. Supervised machine learning models comparison after variables selection

Supervised machine learning models were performed once again with the seven selected variables. As represented in Table 2, the training results showed that Extra Trees Classifier model was the best performing model regarding the metrics (Accuracy of 0.90, AUC = 0.95, Recall = 0.94, Precision = 0.87and F1 = 0.90). The evaluation using the test-set, of the different models, confirmed thatthe extra trees classifier was the best performing model for prediction of poor glycemic control (accuracy = 0.81, AUC = 0.86, Recall = 0.81, precision = 0.72 and F1 = 0.76) Table 2. The Extra tree classifier confusion matrix using the test-set is represented in **Supplemental FigureS7.**

The final validation using the unseen data showed an accuracy of 0.87, an AUC of 0.87, a recall of 0.75, a precision of 1, and an F1 rate of 0.85.

To better appreciate the contribution and the ranking importance of the seven selected variables mentioned above in the onset of poor glycemic control among non-fasting patients during Ramadan, results obtained with the Extra tree classifier model were represented in a Shapley Additive Explanations (SHAP) value plot. As shown in Fig. 2**,** gender was considered the most important parameter for predicting the risk of poor glycemic control. Indeed, females were at greater risk of poor glycemic control than men. The second important predicting parameter was waist circumference evolution, patients with a higher waist circumference

Table 2
Comparison of performance of the machine-learning algorithms on train-set and test-set using the selected features.

	Data	Trainset					Test-set				
	Metrics	**Accuracy**	**AUC**	**Recall**	**Prec.**	**F1**	**Accuracy**	**AUC**	**Recall**	**Prec.**	**F1**
SVM - Linear Kernel		0.8636	0.913	0.9400	0.8162	0.8679	0.7727	0.7812	0.8125	0.6500	0.7222
Naive Bayes		0.8927	0.9260	0.9000	0.8900	0.8903	0.7500	0.8817	0.7500	0.6316	0.6857
Random Forest Classifier		0.8545	0.9440	0.8800	0.8381	0.8538	0.7727	0.8460	0.7500	0.6667	0.7059
Extra Trees Classifier		0.9036	0.9587	0.9400	0.8781	0.9029	0.8182	0.8638	0.8125	0.7222	0.7647
LogisticRegression		0.8436	0.9000	0.8400	0.8581	0.8262	0.7955	0.8817	0.6875	0.7333	0.7097
CatBoost Classifier		0.8645	0.9200	0.9000	0.8481	0.8686	0.7727	0.8348	0.7500	0.6667	0.7059
DecisionTree Classifier		0.8636	0.9040	0.8800	0.8564	0.8603	0.7273	0.8158	0.7500	0.6000	0.6667
K Neighbors Classifier		0.8755	0.9173	0.9000	0.8567	0.8741	0.7727	0.8627	0.7500	0.6667	0.7059

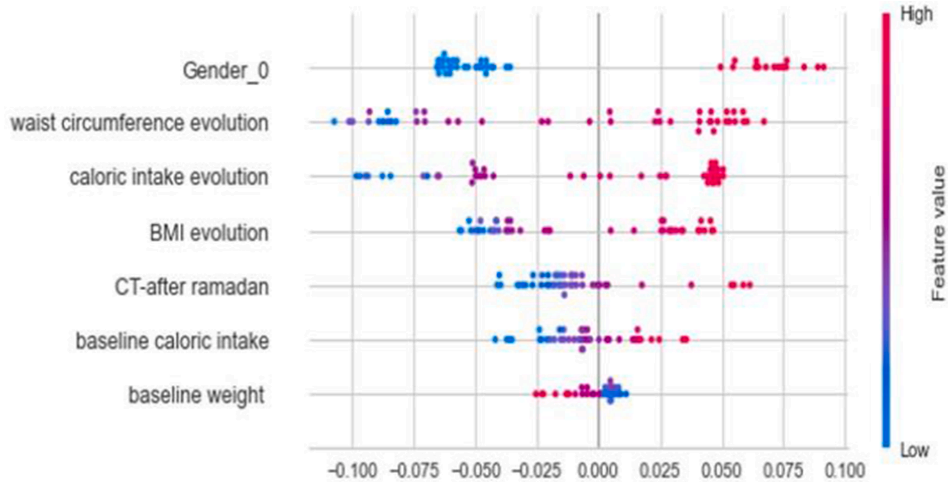

Fig. 2. SHAP value plot using Extra tree classifiers model. (Blue = low risk of poor glycemic control, red = high risk of poor glycemic control; Gender_0 = women, BMI = body mass index, CT: Serum level of total cholesterol**)**. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

evolution during Ramadan were more prone to poor glycemic control. The third parameter was caloric intake evolution, patients with higher caloric intake evolution tended more to poor glycemic control during Ramadan than patients with lower caloric intake evolution. The fourth parameter was BMI variation, patients with higher BMI variation were at higher risk of poor glycemic control than patients with lower BMI variation. The fifth parameter was serum level of total cholesterol after Ramadan, patients with Higher serum levels of total cholesterol after Ramadan were more prone to poor glycemic control than those with lower serum levels of total cholesterol. The sixth parameter was baseline caloric intake, patients with higher baseline caloric intake were at a higher risk of poor glycemic control than patients with a lower baseline caloric intake. The least important feature among the seven chosen parameters is baseline weight, patients with a lower baseline weight had an increased risk for poor glycemic control during Ramadan than those with a higher baseline weight.

3.5. Clinical use

Using thebest performing machine learning model (extra trees classifier), we conducted 31,250 simulations [(5 values of the selected features)$^{6 \text{ continuous parameters}}$ x2]. A focus has been made on classification score probabilities higher than 70%, the following risk factors and thresholds have been noted, allowing us to conclude if the patient is at risk of poor glycemic control during Ramadan:

- Women are more likely to have poor glycemic control during Ramadan.

- Patients with baseline caloric intake above 1641.16 Kilocalories (Kcal).

- Patients with baseline weight higher than 69 kg (Kg).

- An increase in caloric intake during Ramadan above 341.34 Kcal.

- An increase in BMI after Ramadan higher than 1.49 kg/m^2.

- An increase in waist circumference after Ramadan higher than 0.75 cm.

- Poor glycemic control is more prevalent in patients with values of total cholesterol after Ramadan higher than 4.11 mmol/l.

For convenient use of our model inclinical settings, we have deployed our prediction model in a calculator in which the inputs are the values of the 7 selected variables and the output is the poor glycemic control. An example of the calculator run is represented in **Supplemental Figure S8**.

4. Discussion

Many studies have investigated the glycemic control among fasting patients with diabetes during the month of Ramadan but to the best of our knowledge, there are no studies that have evaluated the glycemic control of non-fasting patients during this month. To predict poor glycemic control among these patients, we have developed supervised machine learning models. Indeed, many studies have been carried out using supervised machine learning to build predictive models for risk of diabetes[18], diabetes detection[19], diabetes complications [920], and management of diabetes[21,22]. In terms of Ramadan fasting, to our knowledge, the only study that has used a ML approach was developed by Elhadd et al. In this study, authorspredict glucose variability and hypoglycemia risk in patients with type 2 diabetes on a multiple drug regimen who fast during Ramadan [12].

Our study showed that the Extra Trees Classifier model was the best performing model to predict poor glycemic control with high rates of accuracy and AUC (0.87 for both metrics).

We also identified, using Recursive feature elimination, the variables that were most significantly related to poor glycemic control among our patients. The increase in caloric intake during the month of Ramadan was presented as a top-ranked variable by all models. In addition, baseline caloric intake was ranked as an important factor of poor glycemic control by most models. Simulations based on the Extra Trees Classifier model showed that the probability of poor glycemic control is higher than 70% when baseline caloric intake is above 1641.16 Kilocalories, and the increase of caloric intake during Ramadan is above 341.34 Kcal. Few studies have investigated food intake modifications among non-fasting patients with diabetes during Ramadan. Indeed, Sebbani et al have shown that the daily caloric intake in non-fasting diabetics was significantly higher in non-fasting patients with diabetes than those who fasted [6]. Excessive caloric intake leads to an increase in weight and maintains insulin resistance. In addition, excessive carbohydrate intake causes postprandial hyperglycemia [23]. However, no study has investigated the causal relationship between increased caloric intake and glycemic control among those patients.

Total cholesterol blood level after Ramadan was also identified as an important factor in the predictive factors. Indeed, our simulations showed that poor glycemic control is more prevalent in patients with values of serum level of total cholesterol after Ramadan higher than 4.11 mmol/l. This could be explained by the change in eating habits during the month of Ramadan, and the festive nature of meals during this month. Sebbani et al, showed that total fat intake consumption was

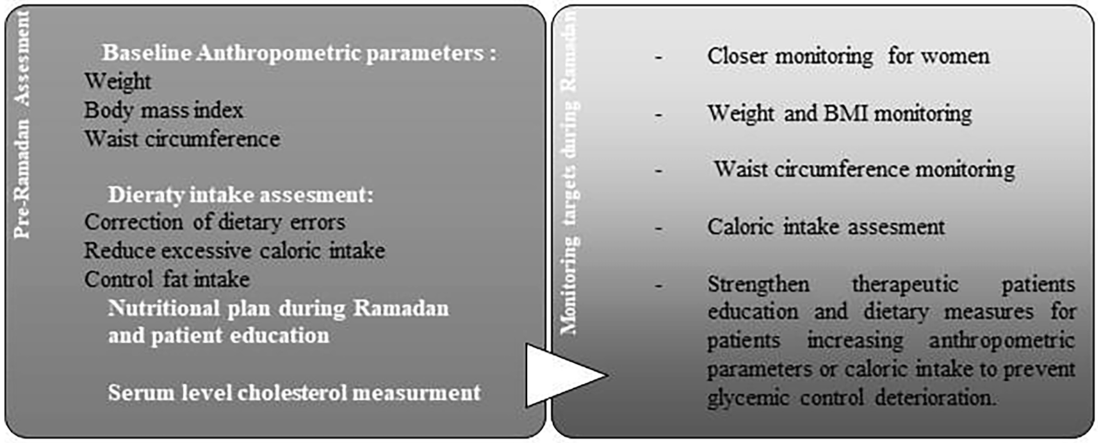

Fig. 3. Suggested Flowchart for the monitoring of non-fasting patients with diabetes during the month of Ramadan based on Machine learning models results.

significantly higher in non-fasting patients with diabetes compared to fasting patients (total fat intake (g) 57 g versus 41.3 g $p = 0.016$), cholesterol consumption was also higher among non-fasting patients but this result was not statistically significant [6].

Additionally, anthropometric parameters, such as baseline weight and evolutions of BMI and waist circumference repeatedly were appeared at the top of the list of top-ranked variables in our study. The simulations have allowed us to set thresholds beyond which the probability of poor glycemic control is greater than 70%. These thresholds are a baseline weight greater than 69 kg, an evolution of the BMI after Ramadan greater than 1.49 kg/m^2, and evolution of waist circumference greater than 3.75 cm. Indeed, previous studies showed that increased weight has been shown to worsen glycemic control[24], evolution of waist circumference has also been associated with glycemic control deterioration in patients with type 2 diabetes[25].

Interestingly, our predictive model showed that women with diabetes who do not fast are more likely at risk of poor glycemic control. This finding is in accordance with the results of the Qatari study PRO-FAST that investigates biophysical and biochemical changes among patients with type-2 diabetes during Ramadan fasting, and find that that there is a significant difference in terms of glycemic control and lipids profile between the two sexes [26].

Finally, to improve the management of non-fasting patients with diabetes during the month of Ramadan, we suggest in Fig. 3, a flowchart including all the findings obtained using machine learning models. We also deployed our prediction model in a friendly user calculator to facilitate its clinical use. This calculator can be used in future Ramadan for early screen of patients at risk of poor glycemic control in order to monitor their diabetes during this month.

4.1. Limitations and strength

This is the first study that uses Artificial intelligence and Machine learning models to predict glycemic control in patients who do not fast during Ramadan. Besides that, very few studies have investigated the impact of the month of Ramadan on this group of patients. The main limitations of our study include the limited sample size, this is due to the fact that a large majority of patients with diabetes fast during the month of Ramadan. However, we used a consistent methodology to make the machine learning models more accurate. The second limitation is that all the data were collected from a single center. Further multicentric studies further multicentric studies would allow us to include a larger sample and to collect more data and to consolidate the findings of our study.

In the present study, we highlighted the contribution of machine learning in the healthcare field. To the best of our knowledge, this is the first study that used machine learning models to predict poor glycemic control among patients with diabetes who do not fast during Ramadan. Indeed, our model identified the risk factors for poor glycemic control among the 43 initial features. Additionally, running the model allowed us to determine thresholds for each risk factor of poor glycemic control. Moreover, to help physicians to predict poor glycemic control among non-fasting patients, we suggested a flowchart to guide the management of diabetes among non-fasting patients during Ramadan. We also deployed the model using a streamlite framework that predict poor glycemic control among those patients. However, a larger study should be conducted to improve our model and generate more accurate predictors.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

Funding/financial support: The authors declare to have no finding for this article.

Authors contributions

I.M. and F.A prepared the first draft. I.M., F.A., F.R.T., S.E. and S.L, acquired and analyzed data. All authors assisted in interpretation of results. I.M, F.A, A.F and A.C. critically revised the manuscript and approved the final version of the article. I.M, F.A, F.R.T. and A.C conceptualized and designed the study and supervised the analysis. I.M and F.A. are the guarantors of this work and, as such, had full access to all the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.diabres.2022.109982.

References

[1] Bahammam A. Does Ramadan fasting affect sleep? Int J Clin Pract 2006;60:1631–7. https://doi.org/10.1111/j.1742-1241.2005.00811.x.
[2] Azizi F. Islamic Fasting and Health. Ann Nutr Metab 2010;56(4):273–82. https://doi.org/10.1159/000295848.
[3] Alghamdi AS, Alghamdi KA, Jenkins RO, Alghamdi MN, Haris PI. Impact of Ramadan on Physical Activity and Sleeping Patterns in Individuals with Type 2 Diabetes: The First Study Using Fitbit Device. Diabetes Ther Res Treat Educ

Diabetes Relat Disord 2020;11(6):1331–46. https://doi.org/10.1007/s13300-020-00825-x.
[4] CREED study: Hypoglycaemia during Ramadan in individuals with Type 2 diabetes mellitus from three continents- ClinicalKey n.d. https://www.clinicalkey.fr/#!/content/playContent/1-s2.0-S0168822717303492?returnurl=https:%2F%2Flinkinghub.elsevier.com%2Fretrieve%2Fpii%2FS0168822717303492%3Fshowall%3Dtrue&referrer= (accessed April 8, 2021).
[5] The characteristics and pattern of care for the type 2 diabetes mellitus population in the MENA region during Ramadan: An international prospective study (DAR-MENA T2DM)- ClinicalKey n.d. https://www.clinicalkey.fr/#!/content/playContent/1-s2.0-S0168822718316619?returnurl=https:%2F%2Flinkinghub.elsevier.com%2Fretrieve%2Fpii%2FS0168822718316619%3Fshowall%3Dtrue&referrer= (accessed April 8, 2021).
[6] International Diabetes Federation and DAR International Alliance. Diabetes and Ramadan: Practical Guidelines, Brussels, Belgium: International Diabetes Federation, 2021. n.d.
[7] Sebbani M, El Ansari N, El Mghari G, Amine M. Food intake during the month of Ramadan in Moroccan patients with type 2 diabetes. East Mediterr Health J 2013; 19:276–81. https://doi.org/10.26719/2013.19.3.276.
[8] Chaki J, Thillai Ganesh S, Cidham SK, Ananda Theertan S. Machine learning and artificial intelligence based Diabetes Mellitus detection and self-management: A systematic review. J King Saud Univ - Comput Inf Sci 2022;34(6):3204–25. https://doi.org/10.1016/j.jksuci.2020.06.013.
[9] Ljubic B, Hai AA, Stanojevic M, Diaz W, Polimac D, Pavlovski M, et al. Predicting complications of diabetes mellitus using advanced machine learning algorithms. J Am Med Inform Assoc 2020;27(9):1343–51. https://doi.org/10.1093/jamia/ocaa120.
[10] Ruan Y, Bellot A, Moysova Z, Tan GD, Lumb A, Davies J, et al. Predicting the Risk of Inpatient Hypoglycemia With Machine Learning Using Electronic Health Records. Diabetes Care 2020;43(7):1504–11. https://doi.org/10.2337/dc19-1743.
[11] Segar MW, Vaduganathan M, Patel KV, McGuire DK, Butler J, Fonarow GC, et al. Machine Learning to Predict the Risk of Incident Heart Failure Hospitalization Among Patients With Diabetes: The WATCH-DM Risk Score. Diabetes Care 2019;42 (12):2298–306. https://doi.org/10.2337/dc19-0587.
[12] Elhadd T, Mall R, Bashir M, Palotti J, Fernandez-Luque L, Farooq F, et al. Artificial Intelligence (AI) based machine learning models predict glucose variability and hypoglycaemia risk in patients with type 2 diabetes on a multiple drug regimen who fast during ramadan (The PROFAST – IT Ramadan study). Diabetes Res Clin Pract 2020;169:108388. https://doi.org/10.1016/j.diabres.2020.108388.
[13] El Kinany K, Garcia-Larsen V, Khalis M, Deoula MMS, Benslimane A, Ibrahim A, et al. Adaptation and validation of a food frequency questionnaire (FFQ) to assess dietary intake in Moroccan adults. Nutr J 2018;17(1). https://doi.org/10.1186/s12937-018-0368-4.
[14] IDF-DAR-Practical-Guidelines_15-April-2016_low_SA.pdf n.d.
[15] GPAQ_Analysis_Guide.pdf n.d.
[16] Association AD. 2. Classification and Diagnosis of Diabetes: Standards of Medical Care in Diabetes—2020. Diabetes Care 2020;43:S14–31. https://doi.org/10.2337/dc20-S002.
[17] Rodríguez JD, Pérez A, Lozano JA. Sensitivity analysis of kappa-fold cross validation in prediction error estimation. IEEE Trans Pattern Anal Mach Intell 2010;32:569–75. https://doi.org/10.1109/TPAMI.2009.187.
[18] Zhang L, Wang Y, Niu M, Wang C, Wang Z. Machine learning for characterizing risk of type 2 diabetes mellitus in a rural Chinese population: the Henan Rural Cohort Study. Sci Rep 2020;10:4406. https://doi.org/10.1038/s41598-020-61123-x.
[19] Zou Q, Qu K, Luo Y, Yin D, Ju Y, Tang H. Predicting Diabetes Mellitus With Machine Learning Techniques. Front Genet 2018;9:515. https://doi.org/10.3389/fgene.2018.00515.
[20] Reddy R, Resalat N, Wilson LM, Castle JR, El Youssef J, Jacobs PG. Prediction of Hypoglycemia During Aerobic Exercise in Adults With Type 1 Diabetes. J Diabetes Sci Technol 2019;13(5):919–27. https://doi.org/10.1177/1932296818823792.
[21] Baum A, Scarpa J, Bruzelius E, Tamler R, Basu S, Faghmous J. Targeting weight loss interventions to reduce cardiovascular complications of type 2 diabetes: a machine learning-based post-hoc analysis of heterogeneous treatment effects in the Look AHEAD trial. Lancet Diabetes Endocrinol 2017;5(10):808–15.
[22] Alfonsi JE, Choi EEY, Arshad T, Sammott S-AS, Pais V, Nguyen C, et al. Carbohydrate Counting App Using Image Recognition for Youth With Type 1 Diabetes: Pilot Randomized Control Trial. JMIR MHealth UHealth 2020;8:e22074. https://doi.org/10.2196/22074.
[23] Mohd Yusof B-N, Yahya NF, Hasbullah FY, Wan Zukiman WZHH, Azlan A, Yi RLX, et al. Ramadan-focused nutrition therapy for people with diabetes: A narrative review. Diabetes Res Clin Pract 2021;172:108530. https://doi.org/10.1016/j.diabres.2020.108530.
[24] Ross SA, Dzida G, Vora J, Khunti K, Kaiser M, Ligthelm RJ. Impact of weight gain on outcomes in type 2 diabetes. Curr Med Res Opin 2011;27(7):1431–8. https://doi.org/10.1185/03007995.2011.585396.
[25] Blaha MJ, Gebretsadik T, Shintani A, Elasy TA. Waist Circumference, Not the Metabolic Syndrome, Predicts Glucose Deterioration in Type 2 Diabetes. Obes Silver Spring Md 2008;16:869–74. https://doi.org/10.1038/oby.2008.12.
[26] Dabbous Z, Bashir M, Elzouki A-N, Ahmed MS, Farouk S, Hassanien M, et al. Differential effects of gender and patient background diversity on the changes in metabolic and biophysical profiles in people with type-2 diabetes from different ethnicities who fast during Ramadan (H1439); a prospective study from Qatar. Diabetes Res Clin Pract 2019;152:171–6. https://doi.org/10.1016/j.diabres.2019.03.032.