**PROBAST**

Study:

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)

Step 2: Type of prediction study

**Is the study a diagnostic or a prognostic study?**

**Diagnostic**

**Is the study a development only, development and validation or validation only study?**

**Development only**

**What is the model of interest?**

XGBoost

**What is the outcome of interest?**

Predict hyper- and hypoglycemic events

Step 3: Assess risk of bias

**Domain 1: Participants**

**Describe the sources of data and criteria for participant selection**

PROFAST was a prospective observational cohort study under-taken at Hamad Medical Corporation, Doha, Qatar, during Ramadan of 1430 (May-June 2018). Inclusion criteria were adult patients aged 18–79 years with T2D on stable treatment of three or more anti-diabetic medications that include either a sulphonylurea (SU) or insulin; planning to fast during the month of Ramadan; HbA1c 13.0%; and estimated glomerular filtration rate (eGFR) > 30 ml/min. Exclusion criteria included a history of recurrent hypoglycaemia (more than 2episodes of symptomatic hypoglycaemia per week) or hypo-glycaemia unawareness; admission with more than 2 episodes of diabetic ketoacidosis (DKA) or with hyperosmolar non-ketotic coma in the preceding year or with an episode of DKA within the previous 3 months prior to the start of Ramadan; active coronary artery disease, congestive cardiac failure, or those with advanced co-morbidities and/or advanced diabetes microvascular complications; patients with newly diagnosed cancer or those undergoing cancer treatment.

**1.1 Were appropriate data sources used, e.g. cohort, RCT or nested case-control study data?**

**Y**

**1.2 Were all inclusions and exclusions of participants appropriate?**

Y

**Risk of bias introduced by selection of participants:**

Low

**Rationale of bias rating**

The exclusion criteria limit the generalizability of the findings to patients with less severe T2D. Yet this may be to avoid dropout from death or any confounding complications.

**Domain 2: Predictors**

**List and describe predictors included in the final model, e.g. definition and timing of assessment**

The best perform-ing model was XGBoost, which was trained with data on physical activity (mean and standard deviation of physical activity measured by number of steps performed in 1-hourintervals up to the last 5 h to predict subsequent blood glucose level). Additionally, the model included data such as hour of the day, day of the week, part of the day and a binary indicator for Ramadan vs non-Ramadan day, along with demographic and medication information from the EHR to enhance the predictive capability of the model.**2.1**

**Were predictors defined and assessed in a similar way for all participants?**

Y

**2.2 Were predictor assessments made without knowledge of outcome data?**

Y

**2.3 Are all predictors available at the time the model intended to be used?**

Y

**Risk of bias introduced by predictors or their assessment**

Low

**Rationale of bias rating**

Smartwatch data without much risk of bias.

**Domain 3: Outcome**

**Describe the outcome, how it was defined and determined, and the time interval between predictor assessment and outcome determination:**

Glucose prediction from continuous glucose data (Freestyle Libre sensor)

**3.1 Was the outcome determined appropriately?**

Y

**3.2 Was a pre-specified or standard outcome definition used?**

Y

**3.3 Were predictors excluded from the outcome definition?**

Y

**3.4 Was the outcome defined and determined in a similar way for all participants?**

Y

**3.5 Was the outcome determined without knowledge of predictor information?**

Y

**3.6 Was the time interval between predictor assessment and outcome determination appropriate?**

Y

**Risk of bias introduced by the outcome or its determination**

Low

**Rationale of bias rating**

The same CGM sensor was used in. all patients.

**Domain 4: Analysis**

**Describe number of participants, number of candidate predictors, outcome events and events per candidate predictor**

Complete data were available from 13 patients. This model accurately estimated normal glucose levels(2584/2715; 95.2%) and hyperglycaemic events (852/1031;82.6%), but fewer hypoglycaemic events (48/172; 27.9%).

**Describe how the model was developed, predictor selection and risk group definition**

Five separate machine learning techniques including Linear Regression, Random-Forest, SVM, XGBoost and Deep Learning were applied. The best performing model was XGBoost.

**Describe whether and how the model was validated, either internally (cross validation, random split sample) or externally (e.g. temporal validation, geographical validation, different setting, different type of participants)**

The data was divided randomly into 80% for training and the remaining 20% for test and all model hyper-parameters were optimized using 5-fold cross-validation on the training set.

**Describe the performance measures of the model, e.g. calibration, discrimination, classification, net benefit, and whether they were adjusted for optimism**

Accuracy

**Describe any participants who were excluded from the analysis**

None.

**Describe missing data on predictors and outcomes as well as methods used for missing data**

Not described

**4.1 Were there a reasonable number of participants with the outcome?**

N

**4.2 Were continuous and categorical predictors handled appropriately?**

Y

**4.3 Were all enrolled participants included in the analysis?**

U

**4.4 Were participants with missing data handled appropriately?**

U

**4.5 Was selection of predictors based on univariable analysis avoided?**

Y

**4.6 Were complexities in the data (e.g. censoring, competing risks, sampling of controls)**

**accounted for appropriately?**

N

**4.7 Were relevant model performance measures evaluated appropriately?**

N

**4.8 Were model overfitting and optimism in model performance accounted for?**

N

**4.9 Do predictors and their assigned weights in the final model correspond to the results**

**from multivariable analysis?**

Y

**Risk of bias introduced by the analysis**

High

**Rationale of bias rating**

Not described how many patients were included and how missing data were handled. Random train test split for time series data. Small amount of hyper and hypoglycemic events. Only accuracy mentioned.

**Overall Risk of bias**

High