**PROBAST**

Study:

SugarMate

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 and validation

**What is the model of interest?**

Recurrent neural network

**What is the outcome of interest?**

Hypoglycemia prediction

Step 3: Assess risk of bias

**Domain 1: Participants**

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

We validate SugarMate on a dataset of 112 participants (35 non-diabetes, 38 type I diabetic patients and 39 type II diabetic patients) collected from July 2016 to January 2017.

**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**

Mix of type 1 and type 2 patients with no further eligibility criteria

**Domain 2: Predictors**

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

Food Intake. SugarMate provides a food menu for users to keep track of their meals. Based on the carbohydrate food list [40], meals are categorized into five types, including grains, vegetables, milk and egg, fruits, and meat. Users are asked to enter the food types and amounts of their meals, based on which SugarMate calculates UC , the carbohydrate of a meal.

Drug Intake. Oral diabetic drugs enhance the secretion of insulin into the blood and are usually used by Type IIdiabetic patients. In SugarMate, a menu of 11 common oral medicines is presented for users based on [5]. Usersare required to select the drug name and record the dosage.

Insulin Injection. Inulin injection is widely used for blood glucose control for Type I and Type II patients. SugarMate provides an insulin type list based on [4] for users to record the usage and dosage of their insulin injection. SugarMate automatically transforms drug intake and insulin injection into the amount of acting insulin UI via bolus and basal rate information [44].

Physical Activity. Daily activities e.g., exercises, consume the carbohydrate and affect blood glucose levels. In SugarMate, we adopt an efficient activity recognition scheme [29], which leverages the accelerometer to automatically record six common physical activities (walking, running, going upstairs, going downstairs, sitting and standing) along with the corresponding durations. SugarMate then calculates the caloric expenditure using a calorie calculator [36].

Sleep Quality. Sleep quality has a long-term influence on the blood glucose level [24]. SugarMate automatically measures sleep quality using [21], which invokes the accelerometer, microphone and light sensor for sleep quality inference. The output sleep quality score US is used for physiological feature extraction.

**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**

All predictors are applicable and independent of outcome and measures similarly for all patients.

**Domain 3: Outcome**

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

Blood glucose value at current time categorized.

**3.1 Was the outcome determined appropriately?**

PY

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

PY

**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**

Binning the BG levels into four categories is a bit uncommon, however the thresholds for hypo- and hyperglycemic events are reference standards. It may also have been useful to make it a regression task.

**Domain 4: Analysis**

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

112 patients. In total we obtain 762639 samples of blood glucose concentration and the corresponding external factors covering around 38132 hours.

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

a novel machine learning paradigm, namely Multi-division deep-dynamic RNN (Md3RNN), isproposed. To include the the aforementioned information sources in an unified framework, we develop two-key ideas that extend the classical RNN. Firstly, the single hidden layer in RNN is replaced with several deep stacked layers.

**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)**

Since all participants collected both measurements of CGM and externalfactors for at least 6 days, we use measurements during the former 5 days for training and the rest for testing.

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

ACC, PRE, REC

**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?**

Y

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

Y

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

Y

**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?**

Y

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

Y

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

PY

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

**from multivariable analysis?**

U

**Risk of bias introduced by the analysis**

Low

**Rationale of bias rating**

Validation may not be optimal: taking only the last day for testing and the rest for training. However, due to the great amount of datapoints this may still be overall low risk of bias.

**Overall Risk of bias**

Low