**PROBAST**

Study:

Step 2: Type of prediction study

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

Hypoglycemia Detection Using Hand Tremors: Home Study of Patients with Type 1 Diabetes

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

Development only

**What is the model of interest?**

Ensemble model

**What is the outcome of interest?**

Hypoglycemia detection

Step 3: Assess risk of bias

**Domain 1: Participants**

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

The inclusion criterion was patients with T1DM who regularly used CGMs. To be consistent, only patients who were using a Dexcom CGM (G5 and G6; Dexcom Inc) were enrolled in the study.

**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**

Appropriate inclusion criteria

**Domain 2: Predictors**

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

We used a mobile app called TremorApp to record, archive, and transfer the accelerometer data. TremorApp is an app our team customized in the lab to run continuously in the background of the watch. It allows participants to make a single tap on the Apple watch whenever they feel they have low blood sugar, and it is logged automatically. In addition, the app is connected to participants’ iPhones, where they can track the number of hypoglycemic events they have reported, as well as their HR and acceleration. Every week, participants would upload their accelerometer datafile, subjective low blood sugar logs, HR data file, and CGM logs over their phones

**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?**

N

**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**

The accelerometer data and hypoglycemia logs rely on participants’ subjective perception of low blood sugar. Not all hypoglycemic events may be logged (e.g., if participants don’t notice symptoms or forget to log). Participants' knowledge of hypoglycemia status can impact when or if they log events, potentially skewing the dataset toward more noticeable events.

**Domain 3: Outcome**

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

Hypogylcemia detection via CGM.

**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?**

N

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

Y

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

High

**Rationale of bias rating**

Similar to above, outcome is dependent on patient subjective experience and influenced by predictor information.

**Domain 4: Analysis**

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

A total of 45 participants started the study, among whom 7 dropped out due to non conformance or technical issues with the phone, Apple Watch, or CGM. In addition, 5 patients’ devices did not correctly record accelerometer data. The data collected from 33 patients, including 21 (64%) females and 12 (36.4%) males, aged between18 and 56 (mean 25.35) years were included in this study.  
One of the challenges of training the algorithms to detect hypoglycemia was the imbalanced data set, with an average of 3.3% hypoglycemic windows per patient.

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

we used 3 machine learning models—random forest, SVM, and KNN—to classify hand tremors (hypoglycemic state) from nonhypoglycemic states in patients with hypoglycemia. The 3 machine learning models were trained on the acceleration features. We also used ensemble learning for the hypoglycemia classification. Ensemble methods are techniques that create multiple models and then merge them to improve classification performance [70]. Ensemble methods usually result in more accurate solutions than a single algorithm. We combined random forest, KNN, and SVM for the ensemble learning model.

**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)**

10-fold CV

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

AUC, SPE, PRE, REC, F1, ACC

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

A total of 45 participants started the study, among whom 7 dropped out due to non conformance or technical issues with the phone, Apple Watch, or CGM. In addition, 5 patients’ devices did not correctly record accelerometer data. The data collected from 33 patients, including 21 (64%) females and 12 (36.4%) males, aged between18 and 56 (mean 25.35) years were included in this study.

**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?**

PN

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

Y

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

N

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

PY

**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?**

Y

**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**

Small amount of outcomes due to imbalanced dataset. Unclear how many exactly. A considerable amount of patients dropped out. Handling of missing data not described.

**Overall Risk of bias**

High