**PROBAST**

Study:

Deep transfer learning and data augmentation improve glucose levels prediction in type 2 diabetes patients

Step 2: Type of prediction study

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

Prognostic

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

Development only

**What is the model of interest?**

Convolutional neural network

**What is the outcome of interest?**

Detect hypo- and hyperglycemic events

Step 3: Assess risk of bias

**Domain 1: Participants**

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

We analyzed data obtained from 40 outpatients with diabetes (19 males; age 65 ± 8 years; BMI at 30 ± 5; with a mean HbA1clevel at 7.33%), who contributed a mean of 130.6 mg/dL blood glucose level through CGM (BG ranging from 40 to 400 mg/dL). Individuals were eligible for inclusion if they were adults with a diagnosis of T2D patients using CGM.

**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**

No further eligibility criteria

**Domain 2: Predictors**

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

The primary outcome of interest in this study is the BG values in the future, e.g., 5 min to 1 hr later. We take the BG measured in 30 min (7 BG values) as one input data segment and predict the future BG level after a prediction horizon, a time period from the most recent CGM measurement in the input BG values

**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**

Glucose forecasting. Predictors independent of outcome

**Domain 3: Outcome**

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

The primary outcome of interest in this study is the BG values in the future, e.g., 5 min to 1 hr later. We take the BG measured in 30 min (7 BG values) as one input data segment and predict the future BG level after a prediction horizon, a time period from the most recent CGM measurement in the input BG values.

We identified all level 1 hypoglycemic (BG level <80 mg/dL) and hyperglycemic (BG level >180 mg/dL) episodes from the CGM recordings.

**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**

Only past and present values used for forecasting

**Domain 4: Analysis**

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

40 patients. Amount of hypo- and hyperglycemic events unclear. 83 hours of training data on average per patient.

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

We developed new deep-learning methods for patient-specific blood glucose level prediction. We considered three different neural network architectures, including recurrent neural networks (RNNs)44,69, gated convolutional neural networks (CNNs)45, and self-attention networks (SAN)46, as well as three different transfer-learning strategies. We also implemented Gaussian process regression (GP), fully connected feed-forward neural networks (FNNs), and support vector machine (SVM) as the baseline models.

**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)**

LOOCV

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

ACC, AUC, SEN, SPE, 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**

We discard any training sequences with one or more missing data points.

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

U

**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?**

Y

**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**

Unclear

**Rationale of bias rating**

The amount of hypo- and hyperglycemic events are not shown. Also unclear how missing data were handled.

**Overall Risk of bias**

Unclear