**PROBAST**

Study:

Adaptive Smart eHealth Framework for Personalized Asthma Attack Prediction and Safe Route Recommendation

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?**

Asthma attack prediction

Step 3: Assess risk of bias

**Domain 1: Participants**

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

The researchers collected the dataset from twenty-one volunteers aged twelve years or more, ranging in asthma disease severity. All study participants were recruited from Makkah health centers.

**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**

Representative population of asthma patients

**Domain 2: Predictors**

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

The system operates using two primary datasets: the asthma dataset and the route dataset. The asthma dataset comprises three primary components, namely the patient’s historical medical record (HMR), daily bio-signals symptoms, and environmental conditions.

Daily bio-signals are an essential part of managing asthma, as they include symptoms such as cough, shortness of breath, and decreased work capacity. One vital feature that is recorded daily is the Peak Expiratory Flow Rate (PEFR). This measures the maximum airflow produced during a forceful exhale after complete lung inflation. The PEFR is determined by the patient’s voluntary effort and muscular strength and provides valuable information on airway flow [11].

Asthma patients are significantly impacted by their environmental conditions, and sudden changes in these conditions can trigger an attack. The environmental data used to monitor these conditions include temperature, wind, humidity, and air quality index (AQI).

**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 can be applied and are 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:**

Unclear how asthma attack is defined exactly. Most likely it is a direct PRO or there is a rule to infer it from patient reported symptoms.

**3.1 Was the outcome determined appropriately?**

U

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

U

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

U

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

U

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

U

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

U

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

U

**Rationale of bias rating**

Outcome definition unclear.

**Domain 4: Analysis**

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

The proposed system was tested and evaluated by ten volunteers, all of whom had asthma, to determine its effectiveness and measure the accuracy of both prediction models. According to confusion matrix only 3 false negatives and 2 false positives.

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

Two machine learning models were developed in this layer: one for personalized asthma attack prediction, and another for predicting route risk levels. XGBoost was used as the machine-learning algorithm for both 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)**

XGBoost algorithm is used in both proposed models, asthma attack and risk level prediction. Each utilized dataset was divided into an 80:20 ratio for training and testing the model.

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

Accuracy, recall

**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 adopted two different methods to fill in the missing values of the dataset. For patient variables, we used the mean value, whereas, for environmental variables, we used the interpolated value.

**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?**

Y

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

Y

**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?**

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**

Small amount of patients, no proper validation.

**Overall Risk of bias**

High