**Experiments on NotebookLM for Study selection, study tagging and PROBAST**

**Study selection**

Excluded studies

**Study 1**: MyAirCoach: the use of home- monitoring and mHealth systems to predict deterioration in asthma control and the occurrence of asthma exacerbations; study protocol of an observational study

**Analysis of "MyAirCoach" Study's Adherence to Inclusion Criteria**

Based on the provided excerpts, **yes**, the "MyAirCoach" study fulfills all the inclusion criteria for the systematic review. Here's a breakdown:

* **Patients monitored in a remote setting:** The study explicitly states its focus on home-monitoring and mHealth systems. Participants are provided with devices to record various measurements at home, fulfilling this criterion.
* **Study focuses on disease outcomes related to chronic conditions:** The study concentrates on asthma, a chronic condition, and aims to predict outcomes like loss of control and exacerbations.
* **Relevant disease outcome monitored:** The primary endpoints are episodes of uncontrolled asthma and occurrences of moderate and severe asthma exacerbations, aligning with the review's scope of relevant outcomes.
* **ML applied and ML methodology sufficiently described:** While the specific ML algorithms are not detailed in the excerpt, the study mentions using techniques like cluster, spectral, and factor analyses, along with probabilistic methods and data mining. It also highlights the involvement of collaborators experienced in handling continuous data for predictive modeling. This suggests a clear application of ML with sufficient methodological description.
* **Prospective study design:** The study protocol describes a 1-year observational study with two phases, clearly indicating a prospective design.
* **Not related to surgery-related outcomes or congenital conditions:** The study exclusively focuses on asthma management and explicitly excludes patients with comorbidities causing overlapping symptoms. There's no mention of surgery-related outcomes or congenital conditions.
* **Includes data analysis and results:** Although the provided excerpts don't contain the final results, the study protocol outlines the planned data analysis methods and mentions the dissemination of results. This demonstrates the intention to analyze and report the findings.
* **Monitoring is not used as an intervention:** The study observes the association between collected data and asthma outcomes, not using the monitoring itself as an intervention to alter treatment. The study design emphasizes that participants continue their usual care and medication regimens.

Therefore, the "MyAirCoach" study appears to meet all the stipulated inclusion criteria based on the information available in the provided excerpts.

**Study 2**: Contact-Less Real-Time Monitoring of Cardiovascular Risk Using Video Imaging and Fuzzy Inference Rules

**Evaluation of Study in Source Against Inclusion Criteria**

Based on the information available in the sources, **no**, this study does not seem to fulfill all the inclusion criteria for the review.

* **Patients monitored in a remote setting:** The study uses a camera integrated into a smart mirror to measure vital signs. This method could potentially be used in a remote setting, however, the source specifically mentions that this study focuses on designing a smart device for home environments, not on using the device for remote patient monitoring.
* **Study focuses on disease outcomes related to chronic conditions:** The study aims to predict the risk of cardiovascular disease. Cardiovascular disease is a chronic condition, supporting this criterion.
* **Relevant disease outcome monitored:** The study uses fuzzy inference rules to predict a level of risk for cardiovascular disease, which is a relevant outcome.
* **ML applied and ML methodology sufficiently described:** The study uses fuzzy logic, a method often categorized as a type of machine learning. It describes the fuzzy inference system and the fuzzy rules used in detail, suggesting fulfillment of this criterion.
* **Prospective study design:** The source describes the architecture and methodology of the smart mirror system, and includes results from experiments that compare the system's measurements with a standard pulse oximeter. However, there is no mention of a prospective study design in which data is collected over time to predict future outcomes.
* **Not related to surgery-related outcomes or congenital conditions:** There is no mention of surgery-related outcomes or congenital conditions, supporting this criterion.
* **Includes data analysis and results:** The source presents data analysis and results from experiments comparing the performance of the system with a standard pulse oximeter. This aligns with the inclusion criteria.
* **Monitoring is not used as an intervention:** The source describes the smart mirror as a monitoring system, not an intervention, which supports this criterion.

While the study explores cardiovascular risk prediction, the lack of clarity regarding a prospective study design raises concerns about its fit with the review's inclusion criteria. The experiments conducted appear to be for validation and comparison purposes rather than a prospective study observing outcomes over time.

**Study 3**: A Novel Semi-Supervised Learning Model for Smartphone-Based Health Telemonitoring

**Inclusion Criteria Evaluation for Study in Source**

Based on the information available in the sources, **yes**, this study appears to fulfill all the inclusion criteria for the review.

* **Patients monitored in a remote setting:** The study explicitly focuses on **smartphone-based health telemonitoring**, where patient monitoring occurs remotely using electronic devices like smartphones. This aligns with the inclusion criteria of monitoring patients in a remote setting.
* **Study focuses on disease outcomes related to chronic conditions:** The study concentrates on Parkinson’s Disease (PD), a chronic neurodegenerative disorder, and aims to predict disease severity scores. This satisfies the requirement of focusing on disease outcomes related to a chronic condition.
* **Relevant disease outcome monitored:** The study aims to predict a disease severity score, specifically the MDS-UPDRS total score, using smartphone-collected activity data like tapping and speaking. Predicting disease severity is a relevant outcome listed in the inclusion criteria.
* **ML applied and ML methodology sufficiently described:** The study employs a novel semi-supervised learning model called S2SSL, incorporating feature and instance selection to predict disease severity scores from smartphone activity data. The source provides a detailed mathematical formulation of the S2SSL model, explaining the integration of feature and instance selection within the model framework. It further describes the algorithm for parameter estimation, encompassing instance selection, solving the predictive function, and feature selection using PSO. This detailed description satisfies the criterion of sufficient description of the ML methodology.
* **Prospective study design:** While the source doesn't explicitly mention the term "prospective," it states that the smartphone app can be used at the patient's chosen times and places, enabling frequent data collection for timely assessment. This suggests a prospective data collection approach where activity data is gathered over time to predict disease severity.
* **Not related to surgery-related outcomes or congenital conditions:** The study exclusively focuses on PD and its associated symptoms, with no mention of surgery-related outcomes or congenital conditions.
* **Includes data analysis and results:** The study involved human subjects and received ethical approval from the Western Institutional Review Board. It utilized data from 37 patients with PD, including 74 labelled instances with MDS-UPDRS scores and 563 unlabelled instances from smartphone-collected activity data. The study reports achieving a high accuracy of 0.828 correlation between the predicted and true disease severity scores on a validation dataset. This clearly demonstrates the inclusion of data analysis and the reporting of results.
* **Monitoring is not used as an intervention:** The study focuses on using smartphone-based telemonitoring to predict disease severity scores using activity data. It explicitly mentions that the app guides users through pre-designed activities to measure PD symptoms, highlighting the observational nature of the monitoring. The study does not involve any interventions or modifications to treatment based on the monitoring.

Thus, based on the available information, the study presented in source appears to meet all the stipulated inclusion criteria.

**Study 4**: How to reliably diagnose arterial hypertension: lessons from 24 h blood pressure monitoring

**Assessment of Study's Compliance with Inclusion Criteria**

Based on the information available in the sources, **no**, this study does not fulfill all the inclusion criteria.

While the study investigates arterial hypertension, a chronic condition, and uses a relevant outcome measure (blood pressure), the source does not state whether the study was conducted in a **remote setting**. The study analyses data from 24-hour ambulatory blood pressure monitoring (ABPM) devices, suggesting that patients were not continuously monitored in a remote setting, but rather wore a device for a period of time.

Additionally, the study appears to be **retrospective** in nature. Source states that the patient data were obtained from a local heart centre and ethical approval and informed consent were not required due to the retrospective study design. This contradicts the inclusion criterion of a **prospective study design**.

Here's a detailed analysis of how the study aligns with each inclusion criterion:

* **Patients monitored in a remote setting:** The sources do not explicitly state whether patient monitoring was conducted remotely. The study's reliance on 24-hour ABPM data suggests that monitoring might not have been fully remote.
* **Study focuses on disease outcomes related to chronic conditions:** The study clearly focuses on arterial hypertension, a prevalent chronic condition.
* **Relevant disease outcome monitored:** The study aims to determine the number and timing of blood pressure measurements required to approximate a patient's average blood pressure for accurate hypertension diagnosis. This aligns with the inclusion criteria of monitoring a relevant disease outcome.
* **ML applied and ML methodology sufficiently described:** The study does not seem to employ machine learning (ML). It primarily uses statistical methods like Kappa coefficients and canonical correlation.
* **Prospective study design:** The study is described as an observational study where patient data was obtained retrospectively from a local heart centre. This indicates a retrospective design, not a prospective one.
* **Not related to surgery-related outcomes or congenital conditions:** The study focuses solely on hypertension diagnosis and does not mention any surgery-related outcomes or congenital conditions.
* **Includes data analysis and results:** The sources provide detailed descriptions of the statistical analysis methods used and report the study results.
* **Monitoring is not used as an intervention:** The study uses 24-hour ABPM and home blood pressure measurements as monitoring tools to understand blood pressure patterns and inform diagnosis. There is no indication that the monitoring itself is used as an intervention to modify treatment.

Therefore, while the study investigates relevant aspects of hypertension management, its retrospective design and the lack of clarity regarding remote monitoring suggest that it does not meet all the inclusion criteria for the review.

**Study 5**: A Ranking-based Weakly Supervised Learning model for telemonitoring of Parkinson’s disease

**Inclusion Criteria Check for Study on Parkinson's Disease Telemonitoring**

Based on the information available in the sources, **yes**, this study appears to fulfill all the inclusion criteria for the review.

Here's a breakdown of the study's characteristics against each criterion:

* **Patients monitored in a remote setting:** The study uses **telemonitoring** to collect data from Parkinson’s disease (PD) patients using their mobile phones. Telemonitoring, by definition, involves monitoring patients remotely using electronic devices. This study aligns with the inclusion criterion of monitoring patients outside of hospitals or labs.
* **Study focuses on disease outcomes related to chronic conditions:** The study concentrates on PD, a chronic neurodegenerative condition, and aims to develop a model for translating data collected by a patient’s mobile device into a predicted score for disease severity assessment. This satisfies the requirement of focusing on disease outcomes related to a chronic condition.
* **Relevant disease outcome monitored:** The study aims to predict PD severity based on mobile-collected tapping activity data. The source lists disease detection and severity as relevant outcomes, and the inclusion criteria specify disease outcomes related to chronic conditions.
* **ML applied and ML methodology sufficiently described:** The study proposes a novel **Ranking-based Weakly Supervised Learning (RWSL)** model. The source provides a detailed description of the RWSL model, including its mathematical formulation, optimization algorithm, and criteria for ranking samples. The source also describes how the model integrates labelled and unlabelled data, leveraging domain knowledge to rank samples based on medication use. This level of detail suggests the fulfillment of the criterion requiring sufficient description of the ML methodology.
* **Prospective study design:** The study's objective is to create a model that can predict future disease severity scores based on mobile-collected data. This implies a prospective approach where data is collected over time. However, the source does not explicitly confirm whether data was collected prospectively.
* **Not related to surgery-related outcomes or congenital conditions:** The study focuses specifically on PD and excludes congenital diseases. There is no mention of surgery-related outcomes.
* **Includes data analysis and results:** The source describes the data analysis methods used to develop and validate the RWSL model. The source also presents a real-data application of the model using data from the mPower study, which collected tapping activity data and MDS-UPDRS scores from PD patients. The source reports the model's performance using metrics like MSE and predictive correlation. The inclusion of data analysis and results aligns with the criteria.
* **Monitoring is not used as an intervention:** The study aims to use mobile phone data to predict PD severity. The data collection process is described as patient self-administered, and the study emphasizes that the goal is to develop a model for translating data into a predicted score, not to directly intervene or modify treatment based on the monitoring. This supports the criterion that monitoring is not used as an intervention.

Although the source does not explicitly state that the study design was prospective, the objective of using mobile-collected data to predict future disease severity suggests a prospective approach. Given the information provided in the sources, the study appears to meet all other inclusion criteria.

Included studies

**Study 1**: Mining telemonitored physiological data and patient-reported outcomes of congestive heart failure patients

**Analysis of Study's Adherence to Inclusion Criteria**

Based on the information available in the sources, **yes**, this study fulfills all the inclusion criteria for the review.

Here's a detailed examination of the study in source against each inclusion criterion:

* **Patients monitored in a remote setting:** The study involves collecting physiological and ambient parameters from congestive heart failure (CHF) patients using wearable and other devices. These devices transmitted data to a smartphone, enabling data collection outside of hospitals or labs. The patients wore wearable devices for specific durations during the day and took daily measurements using non-wearable devices. The study's use of wearable technology and smartphone-based data transmission suggests **remote monitoring**.
* **Study focuses on disease outcomes related to chronic conditions:** The study focuses on CHF, a chronic condition, and aims to investigate the relationship between objective parameters and patient-reported outcomes (PROs) describing the patients' feelings of health. This aligns with the inclusion criteria of focusing on disease outcomes related to chronic conditions.
* **Relevant disease outcome monitored:** The study uses PROs as a measure of patients' well-being. The inclusion criteria list "wellbeing" as a relevant outcome. The study also aims to understand how objective parameters relate to the feeling of health, further supporting its focus on a relevant disease outcome.
* **ML applied and ML methodology sufficiently described:** The study uses machine learning (ML) techniques to mine data for relations between objective parameters and PROs. The source details the feature construction process, including calculating averages, standard deviations, ratios, and changes in feature values. It also explains feature selection methods, including RelieF, Correlation-based Feature Subset Selection (CFS), and expert-driven selection. The source further explains imputation techniques used to handle missing data, such as k-nearest neighbours (kNN), Multiple Imputation by Chained Equations (MICE), and Singular Value Decomposition (SVD). Finally, the source describes the specific ML algorithms used, including Random forest (RF), Naïve Bayes (NB), Decision tree (DT), and sequential minimal optimization (SMO), along with their rationale for selection.
* **Prospective study design:** The Chiron study is described as a "clinical observational study" and the data was collected over 1,086 days of observation. Although the source does not explicitly use the term "prospective," the longitudinal nature of data collection over an extended period suggests a **prospective design**.
* **Not related to surgery-related outcomes or congenital conditions:** The study specifically excludes patients with implanted devices and those on a heart-transplant list. It also excludes congenital conditions as an area of focus.
* **Includes data analysis and results:** The source provides a detailed description of the data analysis process, including feature construction, selection, and imputation methods. It also presents the results, highlighting the accuracy achieved in predicting PROs based on objective parameters. The source also includes the ROC curve for the RF model, showing an AUC of 0.835. This detailed reporting of methods and results fulfills the criteria of including data analysis and results.
* **Monitoring is not used as an intervention:** The study aims to understand the relationship between objective parameters and PROs in CHF patients. The monitoring is observational in nature, used to collect data for analysis and model development. The sources do not describe any interventions or modifications to treatment based on the monitoring data.

**Study 2**: Automatic prediction of chronic obstructive pulmonary disease exacerbations through home telemonitoring of symptoms

**Fulfilling Inclusion Criteria: A Comprehensive Analysis**

Based on the information presented in the sources, **yes**, the study by Fernández-Granero *et al*. detailed in source fulfills all the inclusion criteria for the review.

Let's break down how the study aligns with each criterion:

* **Patients monitored in a remote setting:** The study used a telemonitoring system that allowed COPD patients to report their symptoms daily from home using a Dedicated mobile Device (DmD). This aligns with the inclusion criteria of remote monitoring conducted outside of hospitals or labs.
* **Study focuses on disease outcomes related to chronic conditions:** The study centers around COPD, a chronic lung disease, and its primary aim is the early detection of COPD exacerbations. The inclusion criteria specifically mention COPD as one of the ten chronic conditions of interest.
* **Relevant disease outcome monitored:** The study uses the Anthonisen criteria to define and identify COPD exacerbations. This aligns with the inclusion criteria, which list "disease detection" and "exacerbations" as relevant outcomes. Additionally, the study aims to predict these exacerbations several days before their clinical manifestation, further reinforcing its focus on a relevant disease outcome.
* **ML applied and ML methodology sufficiently described:** The study utilises a Probabilistic Neural Network (PNN) to predict COPD exacerbations. The sources provide a detailed explanation of the PNN's architecture, training process, and the specific parameters used in the model. The authors also describe the input features used for the classifier, data pre-processing steps like handling missing data, and the model's performance evaluation using metrics like sensitivity, specificity, and accuracy.
* **Prospective study design:** The study involved monitoring 16 COPD patients over a six-month period, collecting daily symptom data through the telemonitoring system. This suggests a prospective design, where data is gathered over time to observe outcomes. The authors clearly state that the study was a field trial with a defined duration, supporting a prospective approach.
* **Not related to surgery-related outcomes or congenital conditions:** The study explicitly focuses on COPD and includes specific exclusion criteria related to surgery and congenital conditions. Participants were recruited based on a history of COPD exacerbations requiring hospital admission or treatment with antibiotics/corticosteroids. The study does not investigate any outcomes related to surgery or congenital diseases.
* **Includes data analysis and results:** The sources present a comprehensive account of the data analysis process and the study's findings. They detail the characteristics of the patient cohort, the number of exacerbations observed, the data pre-processing steps undertaken, the model's performance metrics, and a discussion of the results. The inclusion of this information aligns with the criterion requiring data analysis and results.
* **Monitoring is not used as an intervention:** The study's objective is to develop a system for the early detection of COPD exacerbations using remote monitoring data. The telemonitoring system is used to collect symptom data, and the PNN model is developed to predict exacerbations based on this data. There is no indication that the monitoring itself was used to directly intervene in the patients' treatment or management. The authors highlight the potential of the system to improve patient self-management and clinical efficiency, suggesting that the monitoring is primarily observational and intended to inform clinical decisions rather than act as an intervention.

**Study 3**: Early Indication of Decompensated Heart Failure in Patients on Home-Telemonitoring: A Comparison of Prediction Algorithms Based on Daily Weight and Noninvasive Transthoracic Bio-impedance

**Evaluating the Study's Adherence to Inclusion Criteria**

Based on the provided excerpts from "Pasted text" and "ref\_16647.pdf", **yes**, the study by Cuba Gyllensten *et al*. meets all the inclusion criteria for the review.

Here's a detailed examination of how the study aligns with each criterion:

* **Patients monitored in a remote setting:** The study used a home telemonitoring system to collect daily weight and noninvasive transthoracic bio-impedance (NITTI) measurements from patients with chronic heart failure (CHF). Patients were instructed to perform measurements in the morning before breakfast using a weight scale and a wearable bio-impedance vest. This setup, where patients collected data themselves outside of clinical settings, clearly indicates remote monitoring.
* **Study focuses on disease outcomes related to chronic conditions:** The study specifically investigates the early detection of decompensated heart failure (HF) in patients with chronic HF. The review's inclusion criteria explicitly list heart disease as one of the ten chronic conditions of interest.
* **Relevant disease outcome monitored:** The primary objective of the study was to evaluate the predictive value of daily weight and NITTI measurements for detecting decompensation prior to hospitalization due to worsening heart failure. The authors aimed to identify changes in these parameters within a two-week period preceding hospitalization. The review's inclusion criteria include "disease detection," "exacerbations," "readmission," and "mortality" as relevant outcomes. As decompensation in HF often leads to hospitalization, the study's focus aligns with the review's criteria.
* **ML applied and ML methodology sufficiently described:** While the study doesn't explicitly mention "machine learning," it employs several algorithms to analyse weight and NITTI data for predicting HF decompensation. These algorithms include rule-of-thumb (RoT), moving average convergence divergence (MACD), and cumulative sum control chart (CUSUM) methods. The authors provide detailed descriptions of each algorithm, including their formulas, parameters, and how they were applied to the data. The study also describes the process of parameter optimization and cross-validation to ensure the models' generalizability and avoid overfitting.
* **Prospective study design:** The study used data collected as part of the MyHeart heart failure management observational study. Patients were recruited between October 2008 and July 2010, and the average monitoring time was 10 months. This longitudinal data collection over an extended period suggests a prospective study design.
* **Not related to surgery-related outcomes or congenital conditions:** The study explicitly excluded patients with severe chronic obstructive pulmonary disease (COPD), primary pulmonary hypertension, renal insufficiency requiring dialysis, and psychiatric or neurological disorders. Patients who had a prior acute myocardial infarction or coronary artery bypass grafting (CABG) in the previous three months were also excluded. These exclusion criteria ensure that the study does not focus on surgery-related outcomes or congenital conditions.
* **Includes data analysis and results:** The study presents comprehensive data analysis and results, including patient characteristics, the number of heart failure-related hospitalizations, and the predictive performance of different algorithms. The authors report sensitivity, specificity, positive predictive value, negative predictive value, and optimal parameters for each algorithm. They also discuss the statistical significance of findings and compare the performance of weight-based algorithms to NITTI-based algorithms.
* **Monitoring is not used as an intervention:** The study's objective is to evaluate the predictive capabilities of weight and NITTI measurements for detecting HF decompensation. The telemonitoring system serves as a tool for data collection, and the algorithms are developed to analyse this data for early detection purposes. There's no mention of interventions or changes in treatment based on the monitoring data. The authors suggest that the system's potential lies in improving patient self-management and clinical efficiency through early identification of decompensation.

**Study 4:** SugarMate

**Assessing Study Eligibility**

**No**, the study described in the sources does not fulfil all the inclusion criteria for this review.

While the study effectively utilises machine learning and demonstrates a prospective design focused on a chronic condition in a remote setting, its focus on glucose level *inference* rather than the prediction of specific disease outcomes presents a key discrepancy with the inclusion criteria.

Let's examine how the study aligns with each criterion:

* **Patients monitored in a remote setting:** The study utilises a smartphone-based system, SugarMate, to collect data on meals, drug and insulin intake, physical activity, and sleep quality. This information is gathered outside of hospitals or labs, indicating a remote monitoring approach.
* **Study focuses on disease outcomes related to chronic conditions:** The study aims to infer blood glucose levels, a crucial factor in managing diabetes, a chronic condition explicitly listed in the inclusion criteria.
* **Relevant disease outcome monitored:** This criterion poses a challenge. While the study meticulously infers blood glucose levels, it doesn't directly predict specific disease outcomes like disease detection, severity, exacerbations, readmission, mortality, or medication status as specified in the inclusion criteria. The study's objective is primarily focused on providing users with insights into their blood glucose fluctuations, potentially aiding in self-management but not directly predicting clinically relevant events.
* **ML applied and ML methodology sufficiently described:** The study employs a novel machine learning paradigm, Md3RNN (multi-division deep dynamic recurrent neural network), to infer blood glucose levels. The sources provide a comprehensive explanation of Md3RNN's architecture, training process, and parameters.
* **Prospective study design:** The study involved monitoring 112 participants, including individuals with and without diabetes, over varying periods ranging from 6 to 30 days. This data collection over time to observe blood glucose fluctuations points towards a prospective design.
* **Not related to surgery-related outcomes or congenital conditions:** The study focuses specifically on diabetes management and includes individuals with Type I and Type II diabetes as well as non-diabetic participants. There is no indication of investigating outcomes related to surgery or congenital diseases.
* **Includes data analysis and results:** The sources present detailed data analysis and results, encompassing participant characteristics, blood glucose measurements, model performance metrics (precision, recall, accuracy), and comparisons with other machine learning models.
* **Monitoring is not used as an intervention:** The study aims to develop a system for inferring blood glucose levels using data collected through smartphones. The primary function of the monitoring is to provide users with information about their blood glucose levels, potentially enabling better self-management. There is no indication that the monitoring itself is intended as a direct intervention in treatment.

Therefore, although the study excels in various aspects, its emphasis on inferring blood glucose levels rather than predicting specific disease outcomes ultimately makes it ineligible for inclusion in the review based on the provided criteria.

**Study 5**: Proactive Identification of Patients with Diabetes at Risk of Uncontrolled Outcomes during a Diabetes Management Program: Conceptualization and Development Study Using Machine Learning

**Study Inclusion Criteria Evaluation**

Based on the provided excerpts, the study appears to fulfil all the inclusion criteria for the review.

**Therefore, the answer is Yes.**

Here's a breakdown of how the study aligns with each criterion:

* **Patients monitored in a remote setting:** The study uses data from Teladoc Health’s Livongo for Diabetes RDMP, where patient monitoring is conducted remotely using a cellular-enabled device and mobile apps.
* **Study focuses on disease outcomes related to chronic conditions:** The study explicitly focuses on predicting uncontrolled diabetes outcomes, a significant concern in diabetes management.
* **Relevant disease outcome monitored:** The study uses HbA1c levels, a clinically relevant measure of long-term blood sugar control, to define and predict uncontrolled diabetes.
* **ML applied and ML methodology sufficiently described:** The study utilises machine learning, specifically a Light Gradient Boosting Machine (LightGBM) model, and provides extensive details about the model development, feature engineering, and hyperparameter tuning processes.
* **Prospective study design:** The study uses registry data from participants enrolled in the RDMP between 2019 and 2022 to predict future diabetes outcomes, indicating a prospective approach.
* **Not related to surgery-related outcomes or congenital conditions:** The study focuses solely on diabetes management and does not involve surgery-related outcomes or congenital conditions.
* **Includes data analysis and results:** The study presents detailed data analysis, including descriptive statistics, statistical analysis using Pearson r correlation, and performance evaluation of the ML models using metrics like recall, precision, AUC, and F1-score.
* **Monitoring is not used as an intervention:** The study uses monitoring data for predictive modelling and does not involve using the monitoring itself as an intervention.

**Study tagging**

Our study tag results

Note that we sometimes use more generalized tags for some study characteristics

|  |  |
| --- | --- |
| Condition | Diabetes |
| Subcondition | Different subtypes |
| Outcomes | Disease exacerbation |
| Dataset size | 112 |
| Algorithm | Deep neural network |
| Specific model | Recurrent neural network |
| Remote data sources | Wearable, Questionnaire |
| Monitored parameters | Food intake information, drug intake information, accelerometer data, calory consumption, sleep parameters |
| Monitoring duration | One week to four weeks |
| Labeling | Parallel measurement via continuous glucose measurement device |
| Evaluation metrics | Accuracy, Precision, Recall |
| Model interpreted | Yes |
| Code / data available | Not available / not mentioned |
| Monitoring frequency | More often than once every hour |
| Data processing | Feature engineering |

NotebookLM tag results

Here's a tagging of the study described in the sources, based on the characteristics you provided:

* **Condition:** Diabetes
* **Subcondition:**
  + Type 1 Diabetes
  + Type 2 Diabetes
  + Non-Diabetic
* **Outcomes:**
  + **Disease severity (chronic disease status)**
  + **Disease risk**
  + **disease exacerbation (acute disease related patient state that is potentially dangerous)**
* **Dataset size:** 112 participants
* **Algorithm:** Multi-division deep dynamic recurrent neural network (Md3RNN)
* **Remote data sources:**
  + **Smartphone sensors**: accelerometer, microphone and light sensor
  + **Questionnaire**: Food intake, drug intake, insulin injection
* **Monitored Parameters:**
  + Blood glucose levels
  + Physical activity
  + Sleep quality
  + Food intake
  + Drug intake
  + Insulin intake
* **Monitoring duration:** **Longer than a month** (up to 30 days, with re-calibration every three weeks)
* **Labeling:**
  + **Parallel measurement**: Blood glucose levels collected by CGM
  + **Forecasting**
  + Expert labeling
  + Validated test: Blood glucose levels measured by finger-pricking
* **Evaluation Metrics:**
  + Precision
  + Recall
  + Accuracy
  + Clarke Error Grid Analysis (CEGA)
* **AUC:** Yes, average accuracy is 82.14%
* **SEN:** Yes, the recalls for levels 1 and 4 are 83.13% and 85.23% respectively
* **SPE:** Yes
* **ACC:** Yes
* **Prediction Model Interpreted:** Yes, through feature importance analysis and layer contribution analysis.
* **Code / data availability:** Not specified in the sources.
* **Monitoring frequency:** **More often than once an hour** (every 3 minutes)
* **Data processing methods:** Feature extraction from physiological and temporal views, missing imputation, normalization using 10-cross validations.

**PROBAST assessment**

Our PROBAST assessment

**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), is proposed. To include 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 external factors 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

NotebookLM PROBAST assessment

**PROBAST Risk of Bias Assessment for SugarMate Study**

This response assesses the risk of bias of the SugarMate study according to the PROBAST tool. Applicability is not assessed. The assessment is done for the development of the Md3RNN model, which predicts blood glucose levels.

**Domain 1: Participants**

* **Signalling Question 1.1: Were appropriate data sources used, e.g., cohort, RCT or nested case-control study data?**
  + **Probably Yes**. The study used data from 112 participants with varying diabetes status (Type 1, Type 2, non-diabetic). While the study doesn't explicitly state the study design, the data collection methodology suggests a prospective cohort study.
* **Signalling Question 1.2: Were all inclusions and exclusions of participants appropriate?**
  + **Unclear**. The study doesn't specify inclusion and exclusion criteria.
* **Risk of Bias introduced by selection of participants:** **Unclear**. Lack of clarity regarding participant selection criteria makes it difficult to assess the risk of bias.

**Domain 2: Predictors**

* **Signalling Question 2.1: Were predictors defined and assessed in a similar way for all participants?**
  + **Yes**. The study clearly defines and assesses predictors like food intake, drug intake, insulin injection, physical activity, and sleep quality in the same manner for all participants.
* **Signalling Question 2.2: Were predictor assessments made without knowledge of outcome data?**
  + **Probably Yes**. The study uses CGM data as the outcome (blood glucose levels). While the study doesn't explicitly mention blinding, the manual input of predictors like food intake, drug intake, and insulin injection suggests that these were likely recorded without knowledge of the CGM readings. The automatic collection of physical activity and sleep quality data further supports this assumption.
* **Signalling Question 2.3: Are all predictors available at the time the model is intended to be used?**
  + **Yes**. The study uses real-time data from smartphone sensors and user inputs for predictions, ensuring predictor availability.
* **Risk of bias introduced by predictors or their assessment:** **Low**. Predictors are well-defined and likely assessed without knowledge of the outcome data, minimising bias risk.

**Domain 3: Outcome**

* **Signalling Question 3.1: Was the outcome determined appropriately?**
  + **Yes**. The study uses CGM readings as the outcome, which is an appropriate measure of blood glucose levels.
* **Signalling Question 3.2: Was a pre-specified or standard outcome definition used?**
  + **Yes**. The study uses standard blood glucose level categories based on clinical definitions (Table 1).
* **Signalling Question 3.3: Were predictors excluded from the outcome definition?**
  + **Yes**. The outcome (blood glucose levels) is distinct from the predictors used in the model.
* **Signalling Question 3.4: Was the outcome defined and determined in a similar way for all participants?**
  + **Yes**. The study used the same CGM device and blood glucose level categories for all participants.
* **Signalling Question 3.5: Was the outcome determined without knowledge of predictor information?**
  + **Probably Yes**. The CGM device records blood glucose levels independently of the predictor data collected through the smartphone app.
* **Signalling Question 3.6: Was the time interval between predictor assessment and outcome determination appropriate?**
  + **Yes**. The study uses a 3-minute time resolution for both predictor and outcome data, aligning them temporally.
* **Risk of bias introduced by the outcome or its determination:** **Low**. The outcome is well-defined, measured appropriately, and likely determined without knowledge of predictor information.

**Domain 4: Analysis**

* **Signalling Question 4.1: Were there a reasonable number of participants with the outcome?**
  + **Probably Yes**. The study included 112 participants and collected a large dataset of over 760,000 blood glucose measurements, indicating a reasonable sample size.
* **Signalling Question 4.2: Were continuous and categorical predictors handled appropriately?**
  + **Yes**. The study employs feature engineering to transform raw data into meaningful predictors, appropriately handling both continuous and categorical variables.
* **Signalling Question 4.3: Were all enrolled participants included in the analysis?**
  + **Unclear**. The study doesn't explicitly state if all enrolled participants were included in the final analysis.
* **Signalling Question 4.4: Were participants with missing data handled appropriately?**
  + **Yes**. The study mentions using missing data imputation and normalisation techniques.
* **Signalling Question 4.5: Was selection of predictors based on univariable analysis avoided?**
  + **Unclear**. The study describes feature engineering but doesn't explicitly mention whether predictor selection was based on univariable or multivariable analysis.
* **Signalling Question 4.6: Were complexities in the data (e.g., censoring, competing risks, sampling of controls) accounted for appropriately?**
  + **Probably Yes**. While the study doesn't explicitly address censoring or competing risks, the use of a deep dynamic RNN model inherently accounts for temporal dependencies and potential complexities in the sequential data.
* **Signalling Question 4.7: Were relevant model performance measures evaluated appropriately?**
  + **Yes**. The study uses precision, recall, accuracy, and Clarke Error Grid Analysis (CEGA) to assess model performance, providing a comprehensive evaluation.
* **Signalling Question 4.8: Were model overfitting and optimism in model performance accounted for?**
  + **Yes**. The study employs regularisation techniques during model training to prevent overfitting. Additionally, it uses 10-fold cross-validation and a separate test set to evaluate the model's generalisability.
* **Signalling Question 4.9: Do predictors and their assigned weights in the final model correspond to the results from multivariable analysis?**
  + **Unclear**. The study doesn't provide details about the final model weights or their correspondence to multivariable analysis results.
* **Risk of bias introduced by the analysis:** **Unclear**. While the study uses appropriate statistical techniques and addresses some potential sources of bias, the lack of information regarding participant inclusion and predictor selection leaves the overall risk of bias in the analysis unclear.

**Overall Risk of Bias:** **Unclear**. The lack of details regarding participant selection criteria and predictor selection process, combined with the unclear risk of bias in the analysis domain, makes it difficult to make a definitive judgment about the overall risk of bias. However, the study demonstrates a robust approach to data collection, outcome definition, and model development, mitigating several potential sources of bias.