***Supplemental Online Content***

**Predicting cardiovascular disease in patients with mental illness using machine learning**

Martin Bernstorff1,2,3, Lasse Hansen1,2,3, Kevin Kris Warnakula Olesen2,4,5,

Andreas Aalkjær Danielsen1,2, Søren Dinesen Østergaard1,2

1 Department of Affective Disorders, Aarhus University Hospital – Psychiatry, Aarhus, Denmark

2 Department of Clinical Medicine, Aarhus University, Aarhus, Denmark

3 Center for Humanities Computing, Aarhus University, Denmark

4 Department of Cardiology, Aarhus University Hospital, Aarhus, Denmark.

5 Department of Cardiology, Regional Hospital Gødstrup, Herning, Denmark.

Word count: X

**Corresponding author**

Martin Bernstorff, MD

Department of Affective Disorders

Aarhus University Hospital - Psychiatry

Palle Juul-Jensens Boulevard 175

8200 Aarhus N

Denmark

E-mail: [manber@rm.dk](mailto:manber@rm.dk)

Telephone: +45 4142 6636

**Table of Contents**

[eMethods 3](#_Toc168577853)

[**eTable 1.** Feature layers. 3](#_Toc168577854)

[Predictor flattening 6](#_Toc168577855)

[eFigures and eTables 7](#_Toc168577856)

[**eFigure 1.** Selection of eligible prediction times before model selection 7](#_Toc168577857)

[**eFigure 2.** Incidence of CVD over time, by type 8](#_Toc168577858)

[**eTable 2.** Hyperparameters for model selection. Parameters for the best performing model in bold. 9](#_Toc168577859)

[**eTable 3.** The 100 most important features by information gain (XGBoost with 5 year lookahead) 10](#_Toc168577860)

# Supplement Methods

## eTable 1. Interventions/procedures used for outcome definition.

|  |  |  |
| --- | --- | --- |
| **Procedure group** | **SKS-code** | **Definition** |
| Coronary | KFNG02, KFNG05, KFNG96 | Percutaneous coronary intervention (PCI) |
|  | KFNA-KFNE | Coronary artery bypass graft (CABG) |
| Intracranial | KAAL10-11 | Coronary endovascular thrombolysis |
|  | KAAL99 | Other intracranial endovascular surgery |
| Iliac artery | KPDE-KPDF\* | Thromb- and embolectomy |
|  | KPDH | Bypass |
|  | KPDN | Plastic |
|  | KPDP | Percutaneous plastic |
|  | KPDQ | Endoprotesis |
| Femoral artery | KPEE-KPEF | Thromb- and embolectomy |
|  | KPEH | Bypass |
|  | KPEN | Plastic |
|  | KPEP | Percutaneous plastic |
|  | KPEQ | Endoprotesis |
| Popliteal artery and distal | KPFE-KPFG\* | Thromb- and embolectomy |
|  | KPFH | Bypass |
|  | KPFN | Plastic |
|  | KPFP | Percutaneous plastic |
|  | KPFQ | Endoprotesis |
|  | KNFQ, KNGQ, KNHQ | Amputations |

## eTable 2. Feature layers.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Predictor** | **Layer** | **SCORE2** | **ICD10** | **ATC** | **NPU** |
| Sex | 1 | X |  |  |  |
| Age | 1 | X |  |  |  |
| LDL | 1 | X |  |  | 01568 |
| Systolic blood pressure | 1 | X |  |  |  |
| Smoking (pack years) | 2 |  |  |  |  |
| Smoking (categorical)1 | 2 | X |  |  |  |
| HbA1c | 3 |  |  |  | 27300 |
| Chronic lung disease | 3 |  | J40-44\* |  |  |
| Psychiatric diagnoses, by subchapter | 4 |  | F0-F9\* |  |  |
| Antipsychotics | 4 |  |  | N05AA01, N05AB06, N05AE03, N05AH02, N05AH03, N05AH04, N05AX08, N05AX11, N05AX13, N05AX14 |  |
| HDL | 4 | X |  |  | 01567 |
| Atrial fibrillation | 5 |  | I48\* |  |  |
| Antihypertensives | 5 |  |  | C02\* |  |
| T1D | 6 |  | E10\*, |  |  |
| T2D | 6 |  | E11\* |  |  |
| Weight | 6 |  |  |  |  |
| Height | 6 |  |  |  |  |
| BMI | 6 |  |  |  |  |
| Chronic kidney failure | 7 |  | N18\* |  |  |
| Angina pectoris | 7 |  | I20\* |  |  |
| Total cholesterol | 7 | X |  |  | 01566 |

1Smoking (categorical) was categorised as “daily”, “occasionally”, “prior smoker” and “never smoked”.

ICD: International classification of disease.

ATC: Anatomical Therapeutic Chemical

NPU: Nomenclature, Properties and Units

## eTable 3. Proportion of observations using fallback

|  |  |  |  |
| --- | --- | --- | --- |
| **Model** | **Predictor layer** | **Type** | **AUROC (95% CI)** |
| XGBoost | 1 | Default parameters, 730 days, mean | 0.80 (0.80-0.80) |
| XGBoost | 1 | Aggregations (min, mean, max) | 0.79 (0.79-0.79) |
| XGBoost | 1 | Lookbehinds (90, 365, 730 days) | 0.79 (0.79-0.80) |
| XGBoost | 1 | Hyperparameter tuned | 0.82 (0.82-0.82) |
| XGBoost | +2 | Hyperparameter tuned | 0.84 (0.84-0.84) |
| XGBoost | +3 | Hyperparameter tuned | 0.84 (0.83-0.84) |
| XGBoost | +4 | Hyperparameter tuned | 0.84 (0.84-0.84) |
| Logistic Regression | 1 | Hyperparameter tuned | 0.83 (0.83-0.83) |
| Logistic Regression | +2 | Hyperparameter tuned | 0.83 (0.83-0.84) |
| Logistic Regression | +3 | Hyperparameter tuned | 0.83 (0.83-0.83) |
| Logistic Regression | +4 | Hyperparameter tuned | 0.83 (0.83-0.83) |
| Logistic Regression | SCORE2 | See eTable 3 for predictors | 0.83 (0.83-0.83) |

## eTable 4. Performance comparison between experiments. Results from five-fold cross-validation.

|  |  |  |  |
| --- | --- | --- | --- |
| **Predictor** | **Layer** | **Lookbehind days** | **Proportion using fallback** |
| LDL | 1 | 90 | 0.73 |
| LDL | 1 | 365 | 0.38 |
| LDL | 1 | 730 | 0.26 |
| Systolic Blood Pressure | 1 | 90 | 0.65 |
| Systolic Blood Pressure | 1 | 365 | 0.34 |
| Systolic Blood Pressure | 1 | 730 | 0.23 |
| Smoking (pack-years) | 2 | 90 | 0.96 |
| Smoking (pack-years) | 2 | 365 | 0.88 |
| Smoking (pack-years) | 2 | 730 | 0.83 |
| Smoking (Categorical) | 2 | 90 | 0.74 |
| Smoking (Categorical) | 2 | 365 | 0.42 |
| Smoking (Categorical) | 2 | 730 | 0.28 |
| HbA1c | 3 | 90 | 0.69 |
| HbA1c | 3 | 365 | 0.32 |
| HbA1c | 3 | 730 | 0.20 |
| Chronic Lung Disease | 3 | 90 | 1.00 |
| Chronic Lung Disease | 3 | 365 | 1.00 |
| Chronic Lung Disease | 3 | 730 | 0.99 |
| HDL | 4 | 90 | 0.69 |
| HDL | 4 | 365 | 0.32 |
| HDL | 4 | 730 | 0.21 |

Note that, for layer 4, the description of psychiatric diagnoses by subchapter is available in Table 1A.

## Predictor flattening

All predictors are time-series. For example, each LDL measurement includes both the time it was measured as well as its value. Traditional machine learning models require that each prediction time, e.g., each psychiatric service contact, be represented as one row. However, each LDL measurement might be relevant to more than one prediction time. Furthermore, more than one measurement might be relevant to the same prediction time. This means we need a many-to-many join, filter by time, and aggregate within this time filter. Specifically, we looked back a given distance from the prediction time (lookbehind window), and aggregated the measurements within this window (see Figure 1F in the manuscript) using the “timeseriesflattener”Python package.32 The final dataset was quality checked manually by inspecting features, as well as with the Deepchecks suite v0.13.1. It passed all relevant checks, e.g. lack of patient overlap between splits and no predictor being too strongly correlated with the outcome as a sign of leakage.33

## Predictor addition by early stopping

Predictor selection is crucial in machine learning model development, as it balances model performance and resource efficiency. Adding predictors may improve discrimination but increases computational resources, complexity in external validation, and the risk of model failure during implementation.20 To balance this trade-off, we employed an early stopping-like predictor selection strategy.21 This iterative approach adds predictors, evaluates model performance, and repeats until no further improvement is observed, ensuring a smaller set of predictors while maintaining good performance.

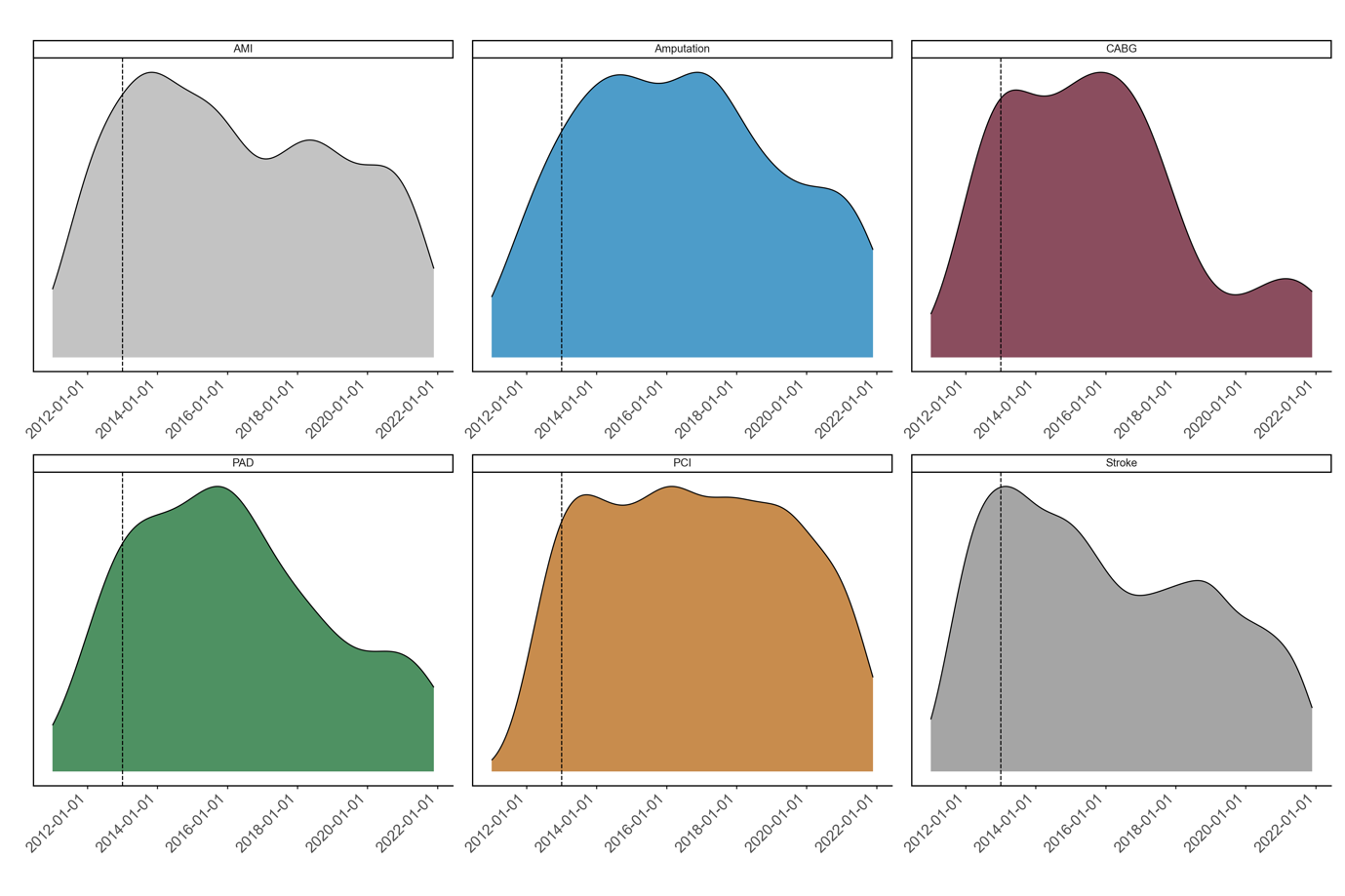
A limitation of this process is its inability to consider predictor interaction. When a model fails to learn from additional predictors due to suboptimal hyperparameters, it may falsely indicate that those predictors are uninformative. To address this, we performed hyperparameter tuning using the best layer and repeated the process with additional predictor layers. Similar performance suggests that the next set of predictors does not enhance prediction.

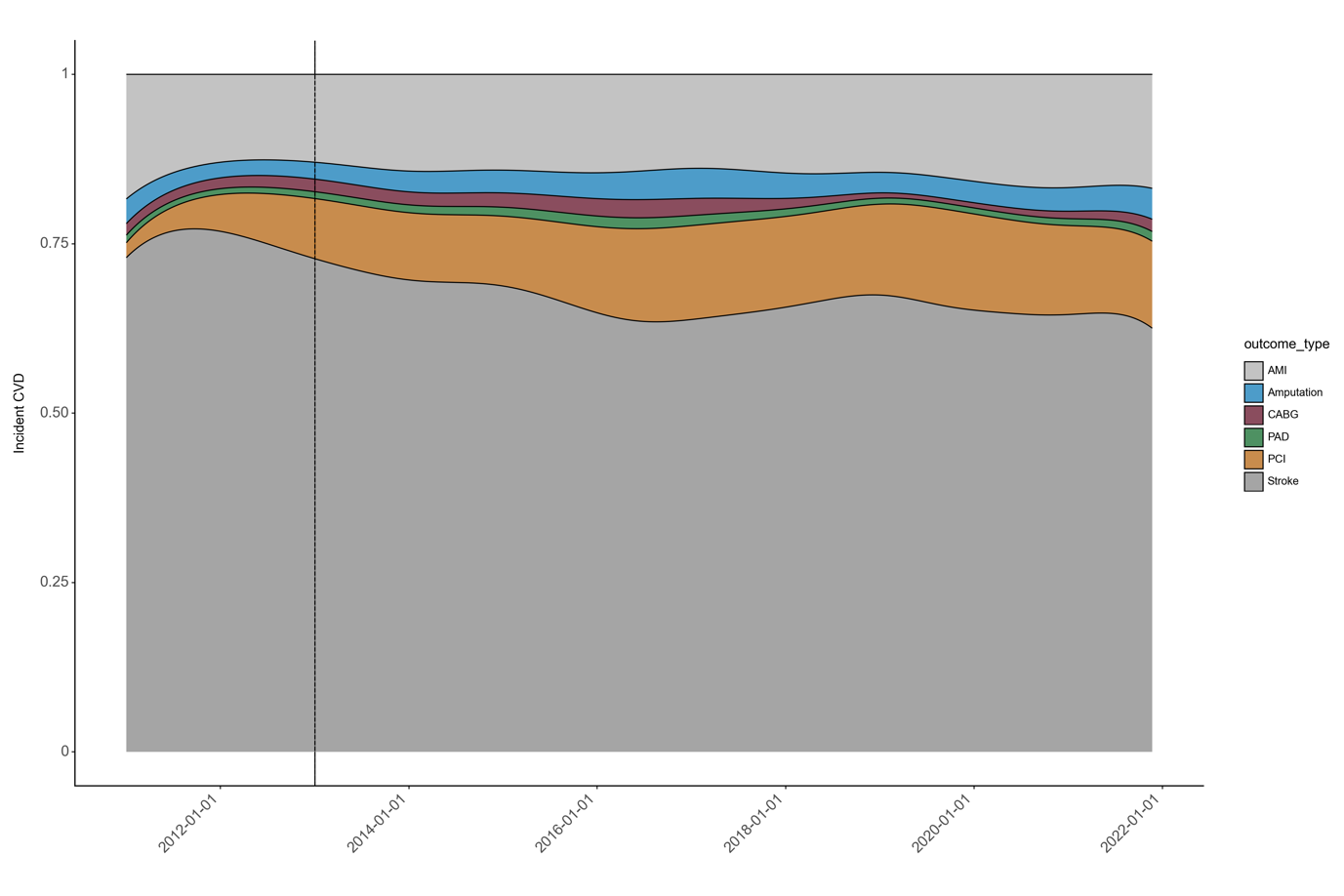
## eFigures and eTables

## eFigure 1. Selection of eligible prediction times before model selection

##### 

## eFigure 2. Incidence of CVD over time, stratified by type of CVD





The dashed line represents the end of the wash-in period (January 1st 2013).

## eTable 2. Hyperparameters for model selection. Parameters for the best performing model in bold.

|  |  |
| --- | --- |
| Model hyperparameters | |
| **XGBoost** | |
| N estimators | [100; 1200], **238** |
| Alpha | [10-8; 0.1], **1.44·10-4** |
| Lambda | [10-8; 1.0], **3.5-5** |
| Max depth | [3; 8], 6 |
| Learning rate | [10-8; 1], **0.014** |
| Gamma | [10-8; 10-3], **1.47·10-4** |

## eTable 4. Feature importance for the best performing model (XGBoost)

|  |  |  |
| --- | --- | --- |
|  | Feature Name | Gain |
| 1 | Age | 0.274978 |
| 2 | Smoking categorical 730-day mean | 0.175136 |
| 3 | Female | 0.168614 |
| 4 | Systolic blood pressure 730-day mean | 0.141200 |
| 5 | Smoking continuous 730-day mean | 0.123549 |
| 6 | LDL 730-day mean | 0.116524 |