PeriMyo - Vorhersage von Perioperativen Myokordschäden mittels Machine Learning

MIRACUM-DIFUTURE Kolloquium 2024

EBERHARD KARLS UNIVERSITÄT TÜBINGEN

Universitätsklinikum Tübingen

Benjamin Sailer

17.09.24

Scientific background

- In Germany: 15 Mio. surgical interventions in 2020
- In 3% of cases, a postoperative increase in cardiac troponin (cTn) occurs during non-cardiac surgery
    - Increase in cardiac troponin: Marker for damage to the heart muscle tissue
- Perioperative Myocardial Injury (PMI):
    - Increase in cardiac troponin (cTn) above the 99th percentile
    - No signs of myocardial infarction
- Patients with PMI: increased mortality (8-15%)
- Goals:
    - Find risk factors of PMI
    - Use Machine Learning to develop a risk score for predicting PMI based on the risk factors identified

Inclusion/Exclusion criteria

- Inclusion criteria:
  - Presence of an operation (OPS 5.01-5.92, except 5.35-5.37)
  - Presence of a troponin measurement during the same inpatient stay as the operation
  - Age >= 18 years
- Exclusion criteria:
  - Cardiac surgery operations (OPS 5.35-5.37)
  - Diagnosis of STEMI (I21.1-I21.4) in the same inpatient stay as the operation
  - Further cardiological diagnostics (OPS code 1-274 to 1-277, 1-279, 1-283, 5-381, 8-837, 3-032, 3-607)

Data used

- Demographic data:
  - Age
  - Gender
- Patient history:
  - Diagnosis(s) (ICD-10 coded)
  - Type and date of surgery (OPS coded)
  - Laboratory values (code, measured value, unit and time of examination; LOINC coded)
- MII core data set modules:
  - Person
  - Diagnose
  - Prozedur
  - Laborbefund
  - Fall

Extraction from FHIR server

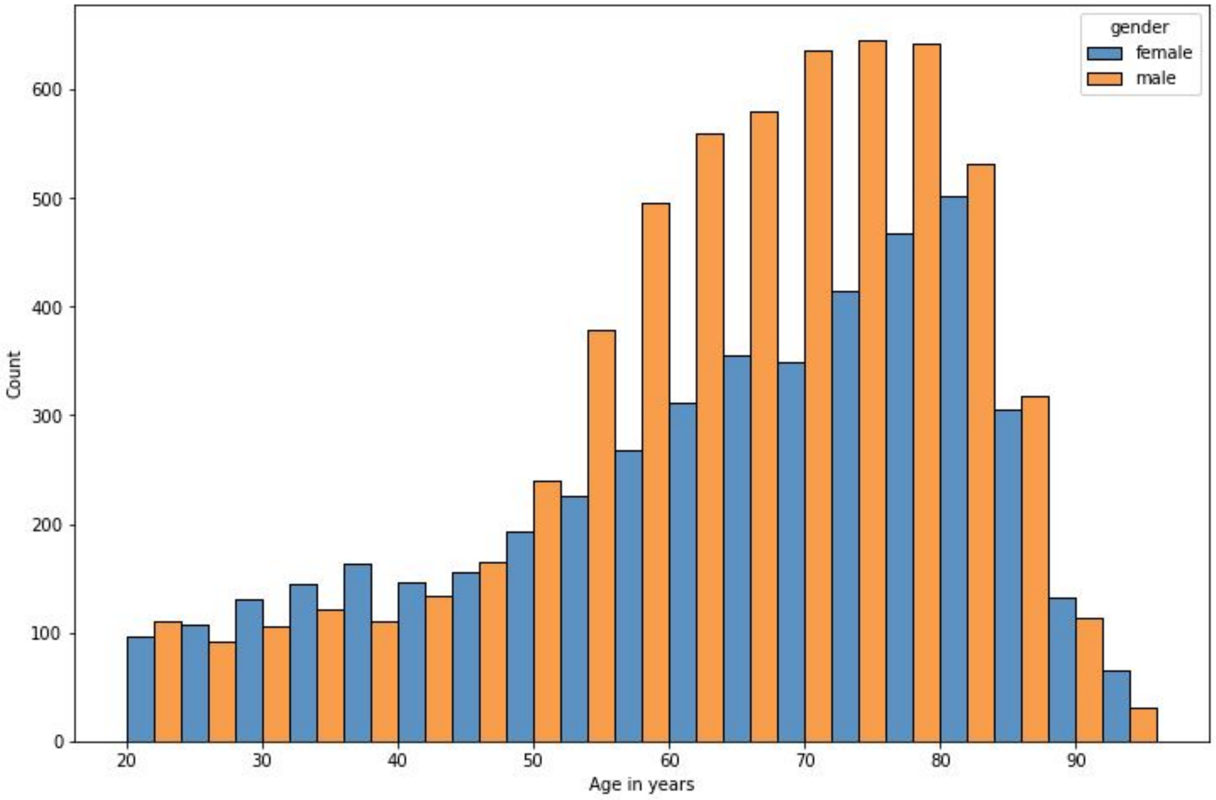
- Due to complexity not feasible with a single FHIR query
- Instead: results of multiple, sequential queries are joined together; use of fhircrackR
- Implementation:
  - Get all observations/patients with troponin measurements
  - Get procedures for those patients (OPS 5.01-5.92, except 5.35-5.37)
  - Match dates (+/- 7 days between observation and procedure)
  - Match multiple operations during the same time period (+/- 3 days)
  - Check exclusion criteria
  - Add other (lab) observations
  - Reduce ICD-10 and OPS granularity: ICD-10: E84.21 -> E84; OPS: 5-123.2 -> 5-12
  - Multiple lab observations -> use mean

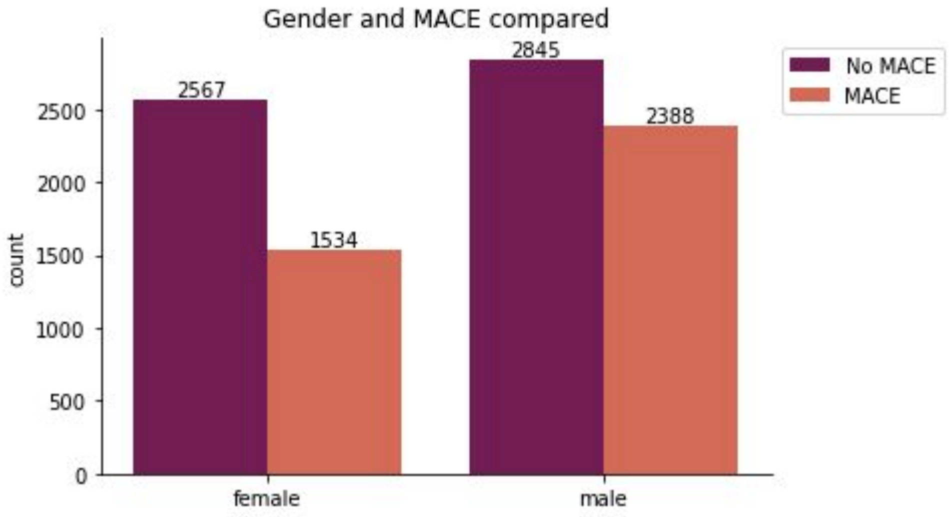
Data set (1)

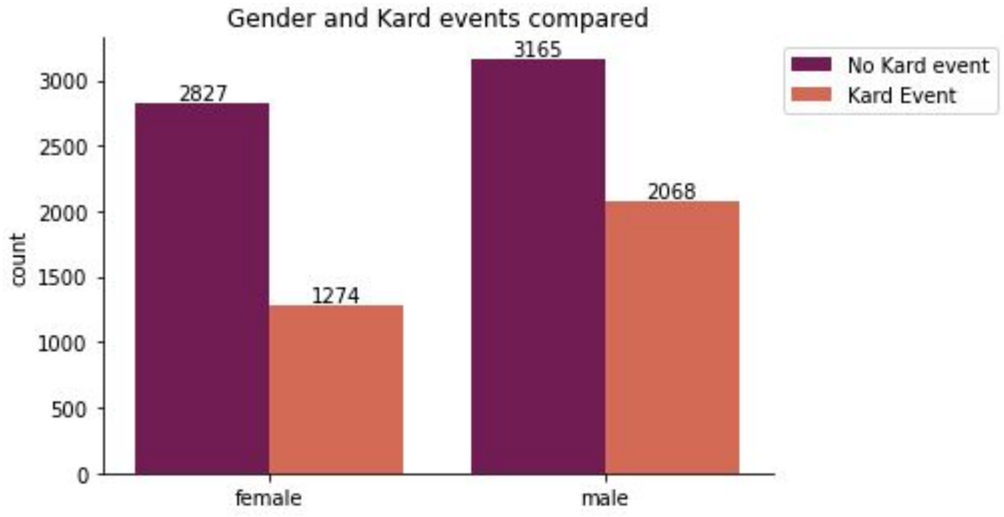
- 9334 patients at 10,617 operations with troponin measurements (3683 increased, 6934 not)
- Features (1314 in total):
  - Age at OP, Gender
  - ICD-10 + OPS codes before OP: True/False
  - Certain selected lab values:
    - 0-2 days before OP
    - 3-7 days before OP
  - Modified Charlson Comorbidity Index (without systems “Acute myocardial infarction" and "congestive heart failure)
  - Cardiac Events (presence of certain OPS codes): True/False
  - Major Adverse Cardiovascular Events (MACE, presence of certain ICD codes): True/False

Data set (2)

```
male      5233
female    4101
```







Machine Learning

- Models used:
  - Logistic Regression: fits data to a logistic function
  - Explainable Boosting Machine: tree-based generalized additive model, support ‘shape functions’
  - Random Forest: ensemble of decision trees
  - XGBoost: ensemble of decision trees, sequential growing of trees
- Several (sub-)selections of features used:
  - All data
  - ICD data: ICD-10, age, gender, CCI, MACE
  - Lab data: Lab, age, gender
  - OPS data: OPS-Codes, age, gender, Cardiac Events
- Software:
  - ETL: R 4.1.1, fhircrackr, tidyverse
  - ML: Python 3.10.12, sklearn, interpret, xgboost, imblearn, pandas, numpy
  - Figures: R 4.4.0, tidyverse, patchwork, ggpubr

Machine Learning - Explainability

- SHAP (SHapley Additive exPlanations):
  - Can explain blackbox models after training
  - Shapley value quantifies its contribution to the difference between the actual model prediction and the expected prediction
  - positive SHAP value: pushes prediction for positive class -> potential risk factor for PMI
- Explainable Boosting Machines:
  - By design interpretable
  - Feature importance: average of the absolute predicted value of each feature for the training dataset
    → corresponds with which features have the largest impact on predictions in the training set

Machine Learning - Model training

- For each subselection of features: repeated 5 times with different seeds
  - Preprocessing:
    - Outlier removal (1.5 * IQR)
    - Features/patients with more than 50% missing values excluded
    - 80-20 train/test/validation split: 80% train/test, 20%: validation
  - Models:
    - Imputation of missing values
    - (if necessary) Scaling
    - Features with 80% pearson correlation are dropped
  - Training:
    - Hyperparametersearch
      - Random Search with 50 iterations / model (EBM: due to slower runtime only 10)
      - 10-fold crossvalidation, validation of models on validation data set (see above)
      - Best model: based on balanced accuracy

Results - all data - prediction of PMI

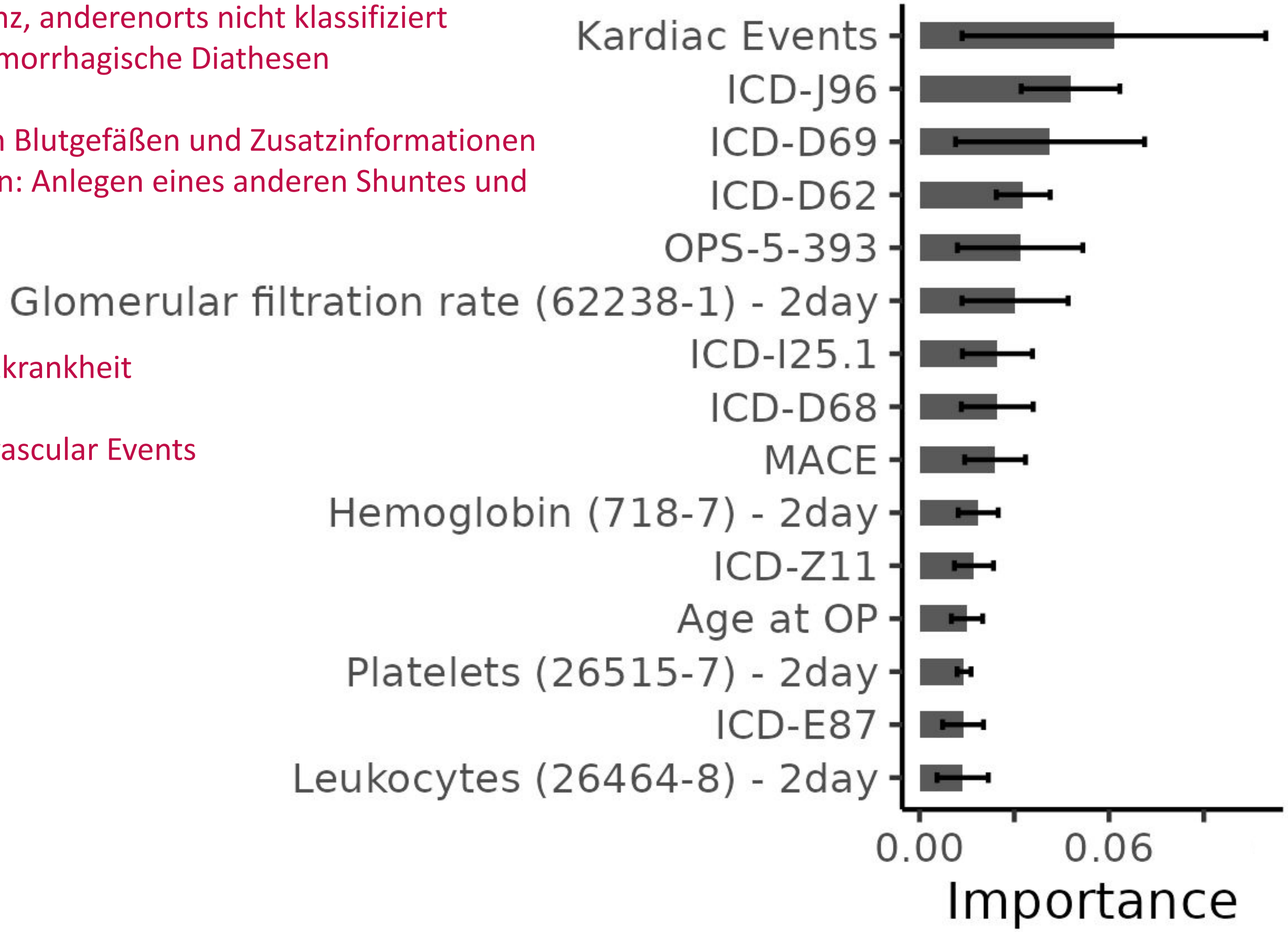
- Note: results are still preliminary (!)
- Average performance metrics per model across 5 seeds

Model	Balanced accuracy	ROC-AUC	Precision	Recall
Explainable boosting machine	0.676 (+/-0.007)	0.826 (+/-0.002)	**0.704 (+/-0.014)**	0.42 (+/-0.014)
Logistic regression	0.697 (+/-0.011)	0.821 (+/-0.008)	0.672 (+/-0.021)	0.487 (+/-0.024)
Random forest classifier	0.692 (+/-0.013)	0.824 (+/-0.008)	0.693 (+/-0.023)	0.464 (+/-0.031)
XGBoost	**0.701 (+/-0.015)**	**0.833 (+/-0.008)**	0.675 (+/-0.029)	**0.496 (+/-0.029)**

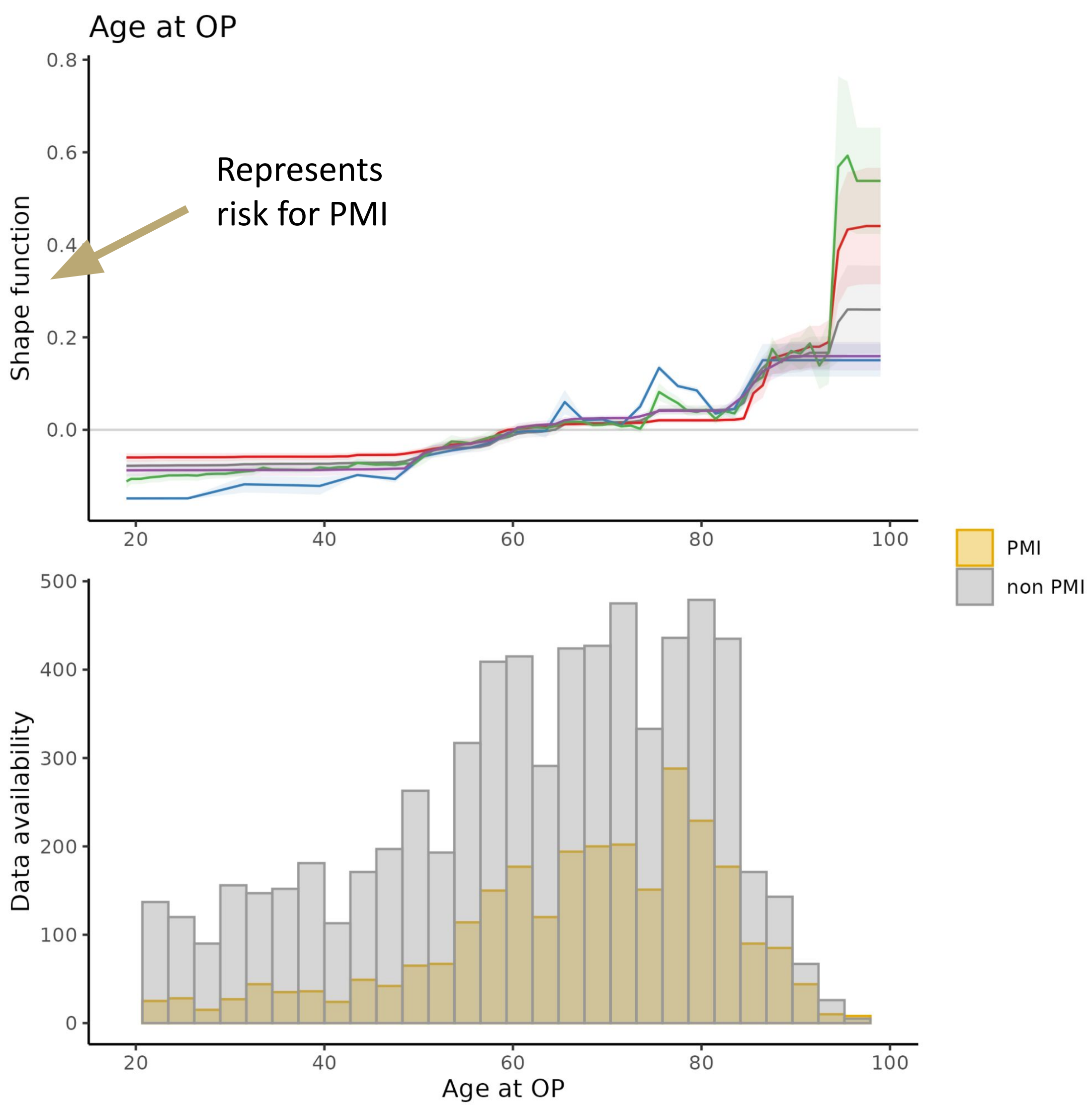
Results - all data - feature importance across all models

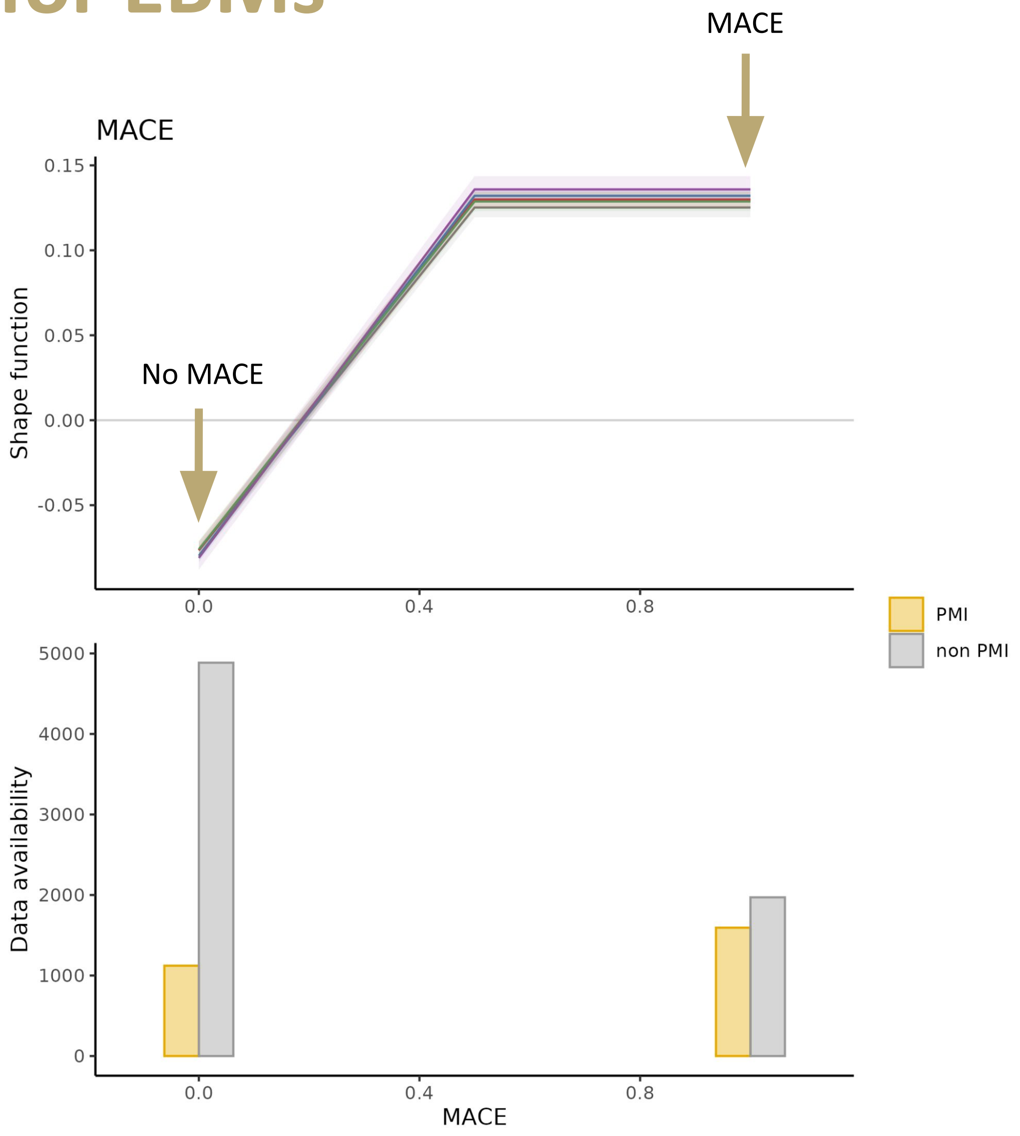
J96: Respiratorische Insuffizienz, anderenorts nicht klassifiziert
D69: Purpura und sonstige hämorrhagische Diathesen
D62: Akute Blutungsanämie
5-393: Andere Operationen an Blutgefäßen und Zusatzinformationen zu Operationen an Blutgefäßen: Anlegen eines anderen Shuntes und Bypasses an Blutgefäßen

I25.1: Atherosklerotische Herzkrankheit
D68: Sonstige Koagulopathien
MACE: Major Adverse Cardiovascular Events



Results - all data - shape functions for EBMs



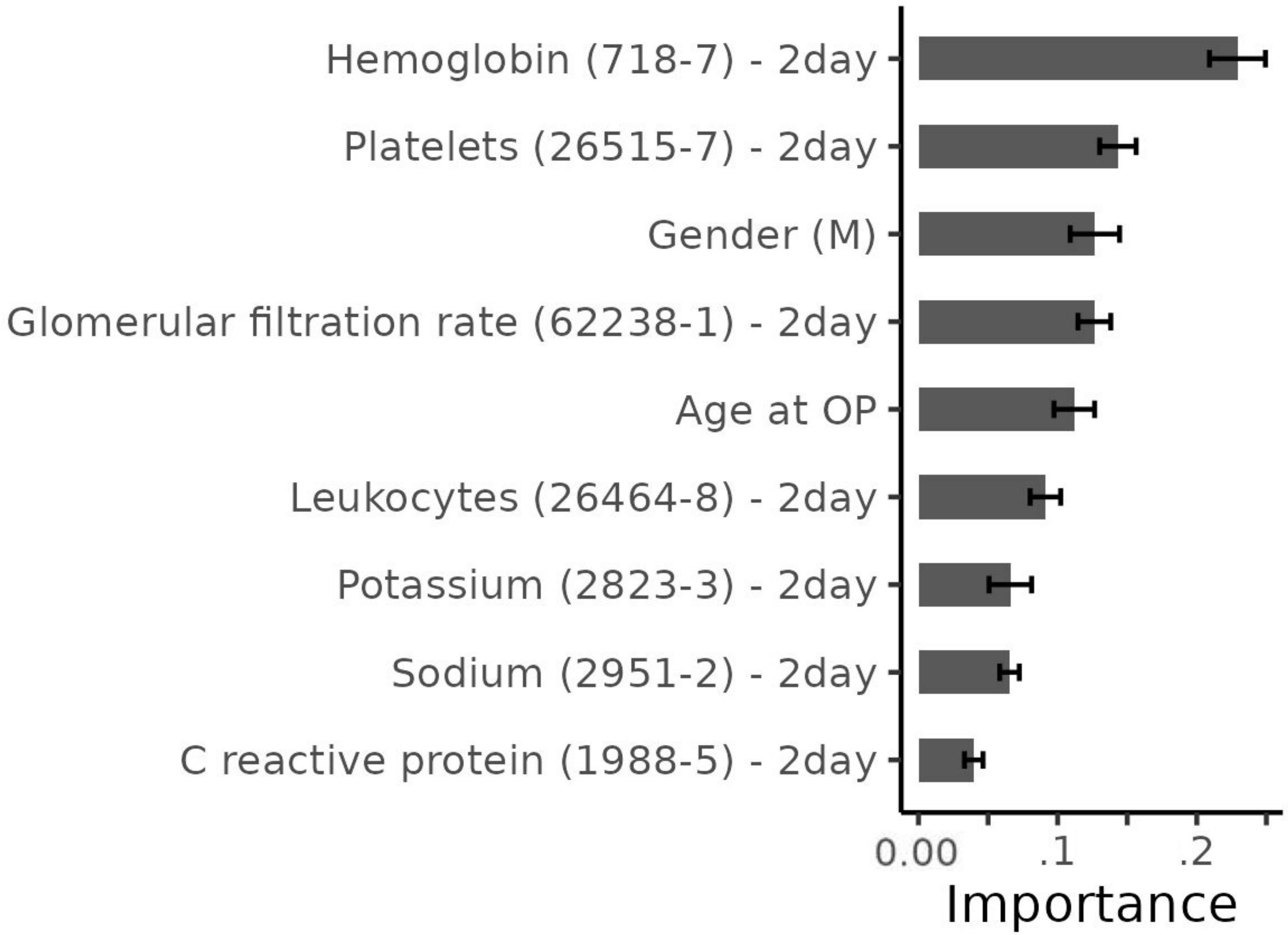


Results - lab data - prediction of PMI

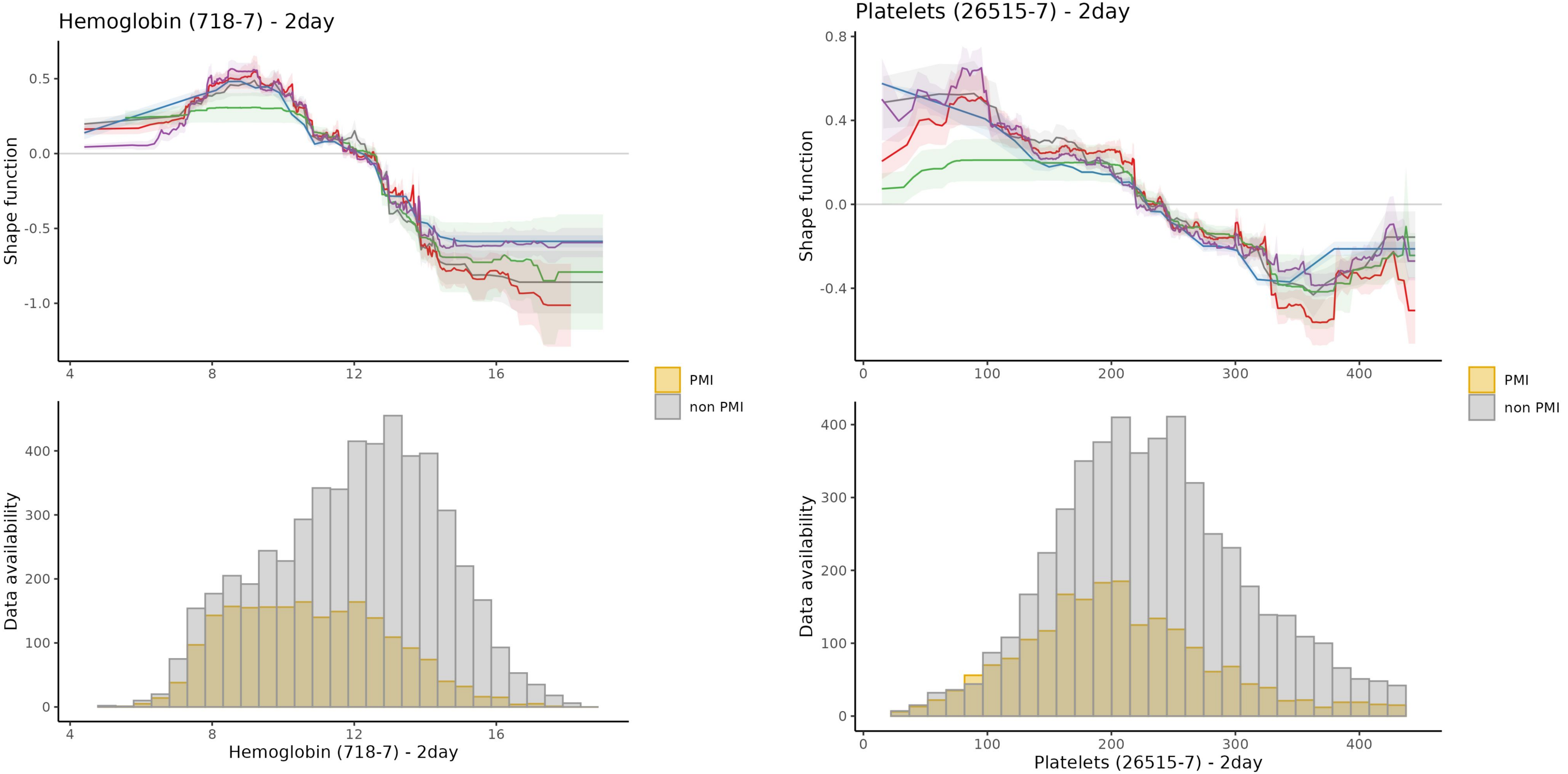
- Average performance metrics per model across 5 seeds

Model	Balanced accuracy	ROC-AUC	Precision	Recall
Explainable boosting machine	0.54 (+/-0.014)	0.686 (+/-0.009)	0.591 (+/-0.042)	0.11 (+/-0.036)
Logistic regression	0.557 (+/-0.007)	0.687 (+/-0.018)	0.556 (+/-0.037)	0.166 (+/-0.015)
Random forest classifier	0.549 (+/-0.008)	0.655 (+/-0.01)	0.453 (+/-0.029)	0.187 (+/-0.024)
XGBoost	0.561 (+/-0.007)	0.645 (+/-0.025)	0.438 (+/-0.041)	0.248 (+/-0.02)

Results - lab data - feature importance across EBM models



Results - lab data - shape functions for EBMs



Results - overall

- Prediction can work quite well:
  - All data: 0.826 (+/-0.005) ROC-AUC
  - ICD data: 0.813 (+/-0.01)
  - OPS data: 0.774 (+/-0.006)
  - Lab data: 0.668 (+/-0.022)
- Results very stable across multiple models/seeds
- Certain diagnosis and procedures seem to be very related with PMI; prediction works good using this data only
- low hemoglobin and platelet counts indicate higher risk of PMI
  -> interpretation TBD

Next steps

- Dig into details of results
- Extend prediction to patients without troponin measurement -> ethics proposal
- Validate results on other data -> FDPG proposal submitted:
  - Decentral analysis:
    - descriptive statistics at sites
    - validation of already trained model
    - training new models at sites
    - (optional): validation of new models at other sites

People involved

- Lars-Christian Achauer (DIZ Tübingen)
- Stephanie Biergans (DIZ Tübingen)
- Michaela Hardt (DIZ Tübingen)
- Michael Köppen (Klinik für Anästhesiologie und Intensivmedizin Tübingen)
- Benjamin Sailer (DIZ Tübingen)
- Sibel Sari-Yavuz (Klinik für Anästhesiologie und Intensivmedizin Tübingen)
- Raphael Verbücheln (DIZ Tübingen)

Questions?

Benjamin Sailer

17.09.24

EBERHARD KARLS UNIVERSITÄT TÜBINGEN

Universitätsklinikum Tübingen