Cardiovascular risk prediction using automated machine learning: A prospective study of 4,23,604 UB Biobank participants

<https://journals.plos.org/plosone/article/file?id=10.1371/journal.pone.0213653&type=printable>

UK Biobank dataset

473 variables were included

Categorized all variables into 9:

> Health and medical history

> Lifestyle and environment

> Blood assays

> Physical activity

> Family history

> Physical measurements

> Psycho-social factors

> Dietary and nutritional information

> Socio-demographics

1. Framingham risk calculator: <https://www.mdcalc.com/framingham-risk-score-hard-coronary-heart-disease>

Miss-forest non-parametric data imputation algorithm

1. Cox proportional hazards model
2. Other standard machine learning models
   1. Linear support vector machines(SVM)
   2. Random forest
   3. Neural network
   4. AdaBoost
   5. Gradient boosting machines

Different risk calculators were used to bench-mark the AutoPrognosis model

AutoPrognosis has automatically selected and tuned best ML model and that no individually tuned ML model performed better than the AutoPrognosis model

AutoPrognosis source code: <https://github.com/ahmedmalaa/AutoPrognosis>