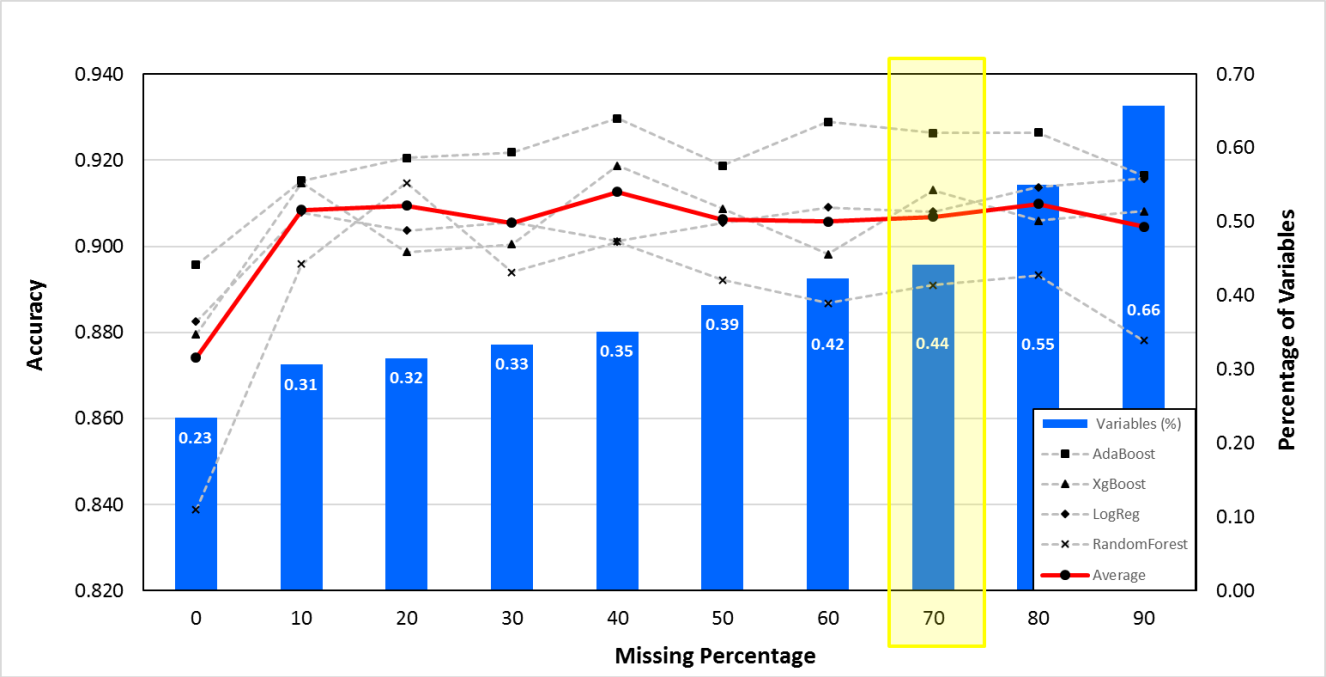
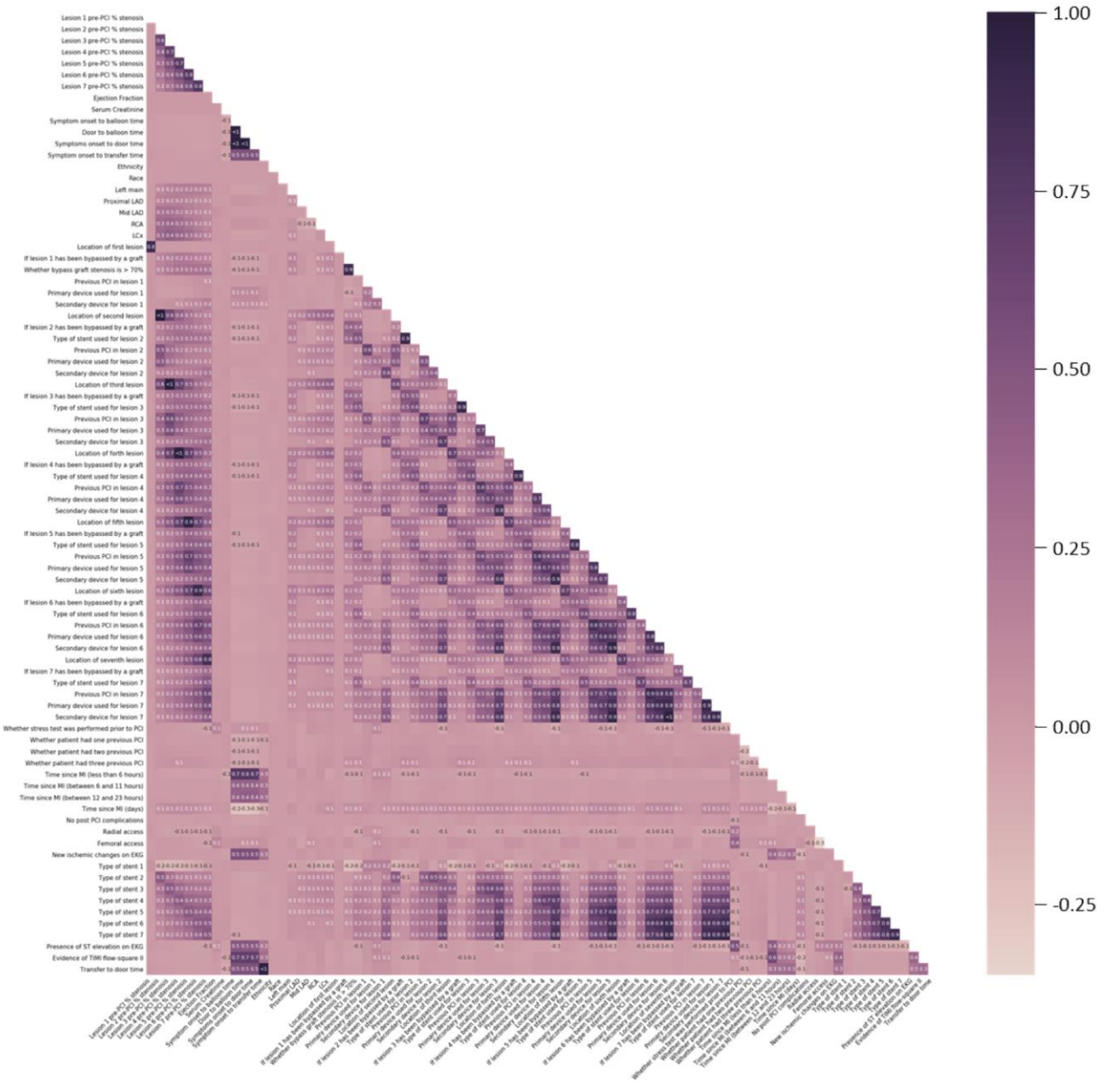
# Supplemental Material

**Figure S1. Comparison between model accuracy and variable missingness.**



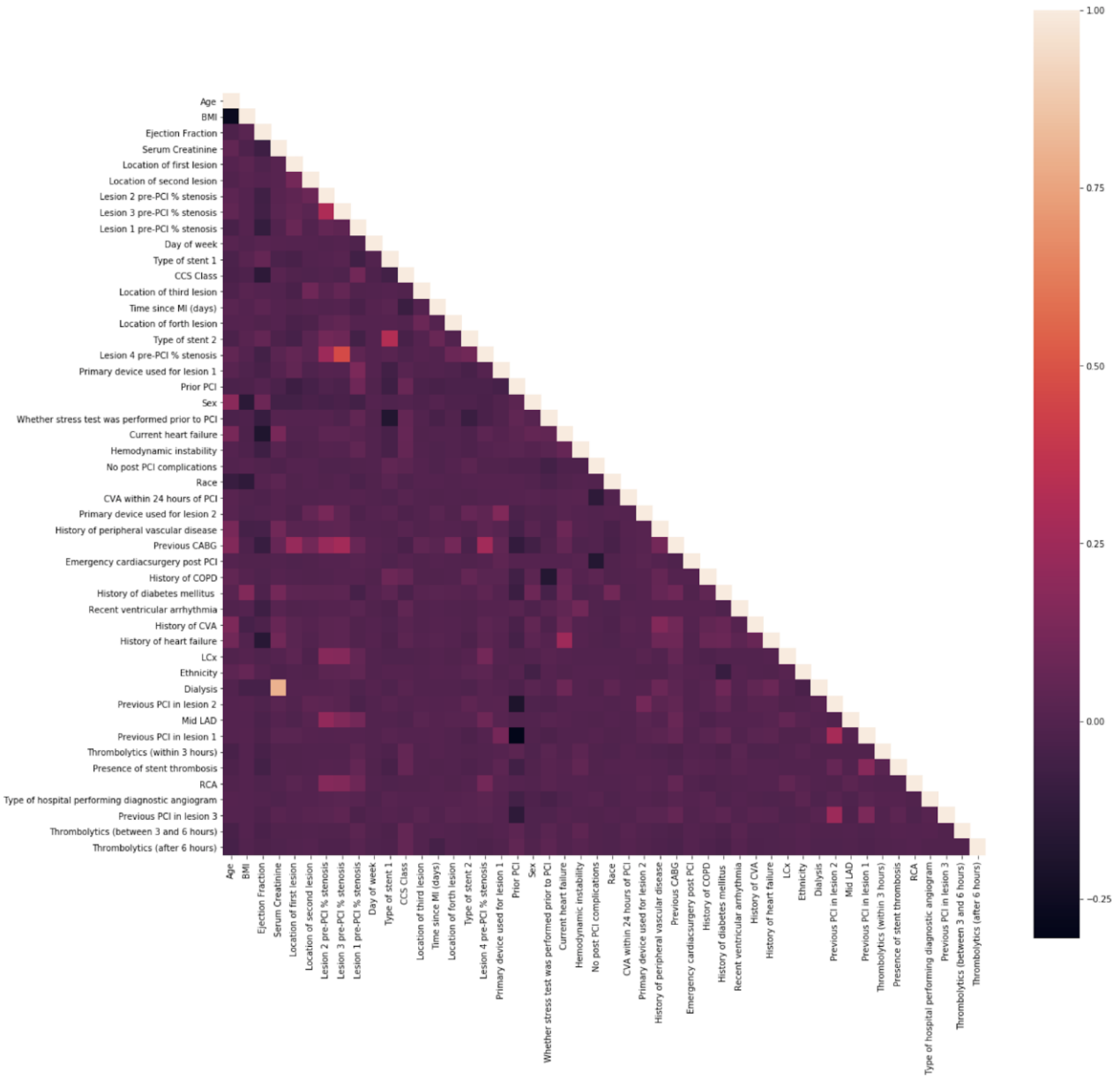

Blue bars represent the proportion of variables meeting inclusion criteria stratified by missingness percentage (i.e. when variables with ≤90% missing values are included in the analysis, 66% of the variables meet the inclusion criteria). In the present analysis, 70% missingness was used as the cutoff value since it represented a balance between overall model performance as well inclusion of a larger proportion of variables (yellow highlight).

**Figure S2. Nullity correlation Heat map.**



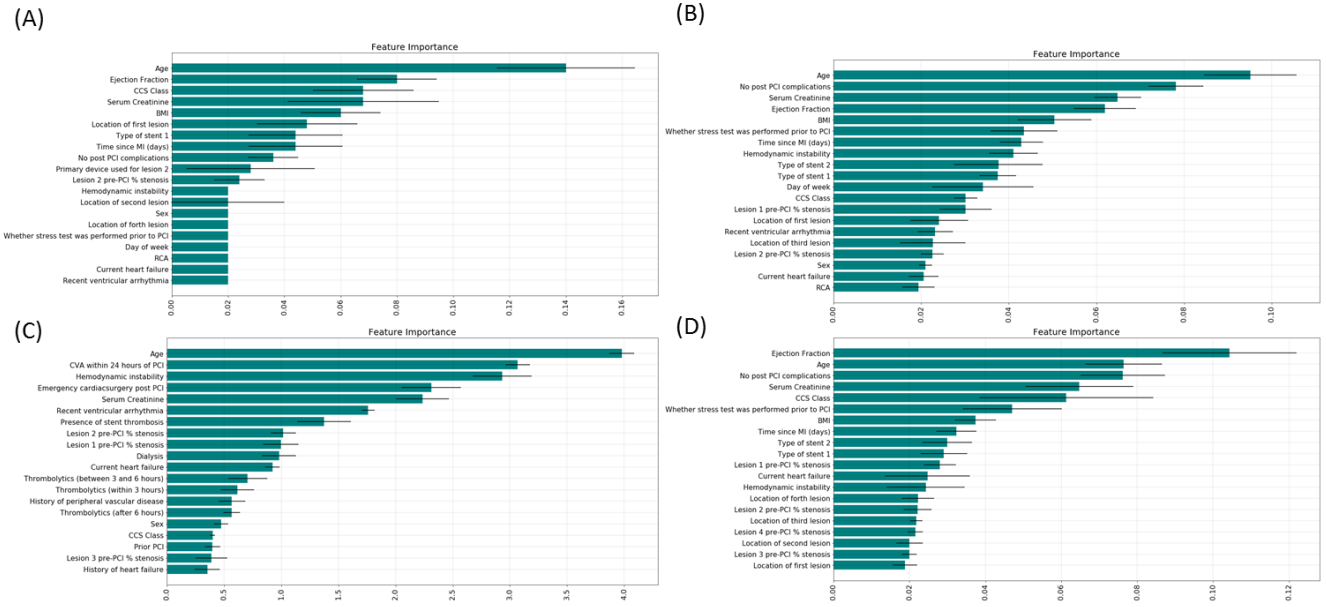

The nullity correlation ranges from -1 to 1. A value of -1 indicates that if one variable appears the other definitely does not. A value of '0' indicates that variables appearing or not appearing have no effect on one another whereas a value of '1' indicates that if one variable appears the other definitely also does. PCI: percutaneous coronary intervention; LAD: left anterior descending; RCA: right coronary artery; LCx: left circumflex; MI: myocardial infarction; EKG: electrocardiogram; TIMI: thrombolysis in myocardial infarction.

**Figure S3. Correlation analysis of the variables in the dataset.**




A simpler model is always more stable and hence preferred. Inclusion of highly correlated variables could contribute to numerical instability, mask interactions between different features, collude the interpretability of a machine learning model and also may cause overfitting. Thus, it is always preferred to exclude one of the two correlated variables. The heat map shown above indicates that dataset does not contain many correlated variables. BMI: body mass index; PCI: percutaneous coronary intervention; CCS: Canadian Cardiovascular Society; MI: myocardial infarction; CVA: cerebrovascular accident; CABG: coronary artery bypass grafting; COPD: chronic obstructive pulmonary disease; LCx: left circumflex; LAD: left anterior descending; RCA: right coronary artery.

**Figure S4. Feature importance ranking and the association standard deviations across 5-fold cross validation.**




Standard deviation, in black bars, is provided along with the feature importance ranking for **(A)** AdaBoost **(B)** XGBoost **(C)** Logistic Regression and **(D)** Random Forest. CCS: Canadian Cardiovascular Society; BMI: body mass index; MI: myocardial infarction; PCI: percutaneous coronary intervention; RCA: right coronary artery; CVA: cerebrovascular accident.