Metric Report

Parameters

I trained four models. I used Deep Learning, Logistic Regression, Random Forrest, and XGboost. For the Deep Learning model I constructed a 2 layer sequential network. I used Relu activation for the first layer and a sigmoid activation for the output layer. When I compiled the model, I used 'sgd' optimizer and for loss I used categorical cross entropy. The model compiled with 20 epochs and the precision and recall scores for predicting complications was 0.

For Random Forest I did CV GridSearch to find optimized parameters. I found that the following variables were optimized as max depth=80, max features=3, min samples_leaf=3, min samples split= 8, and n estimators= 1000.

For Logistic Regression I did hyperparameter tuning on max iterations. I found the value of 100 to be optimal(train score = 0.617).

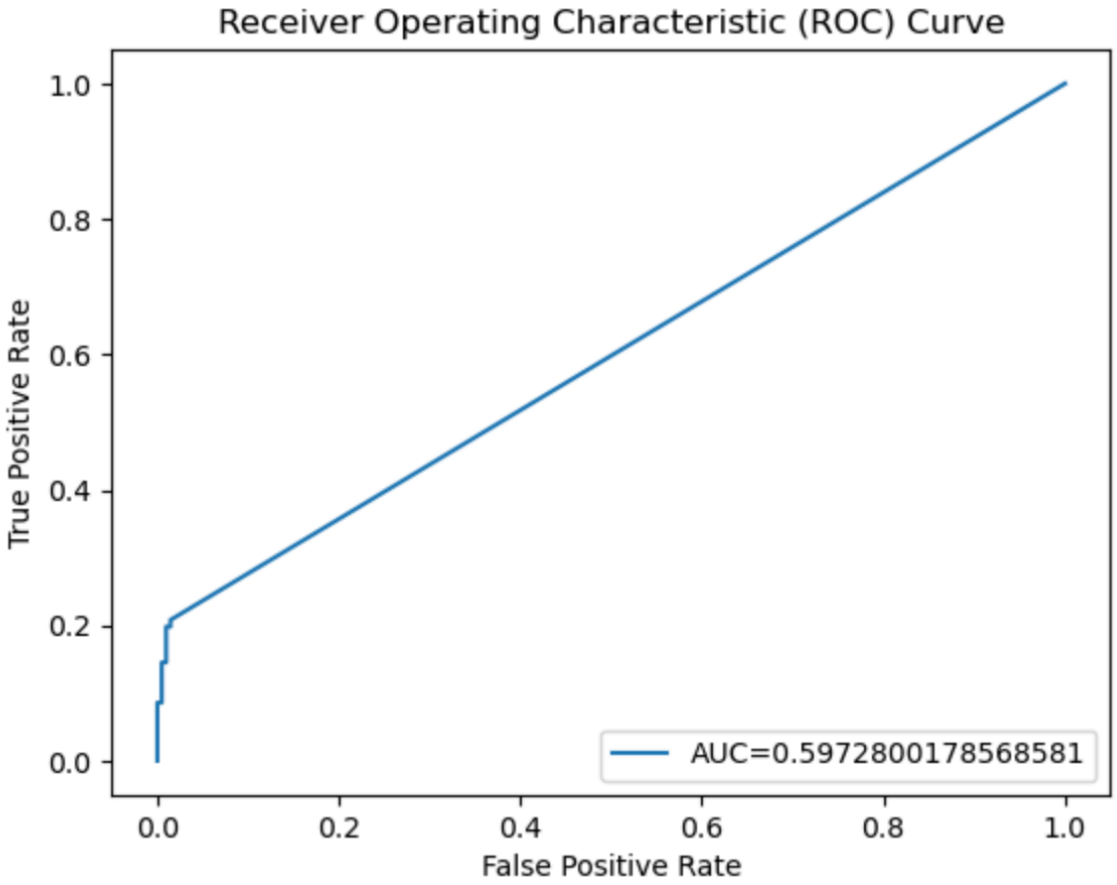
For XGboost I did hyperparameter tuning on Learning Rate and Estimator and found that the best learning rate was 0.5 and the best estimator was 500.

The best performing model was a Random Forest as shown below.

Performance Metrics

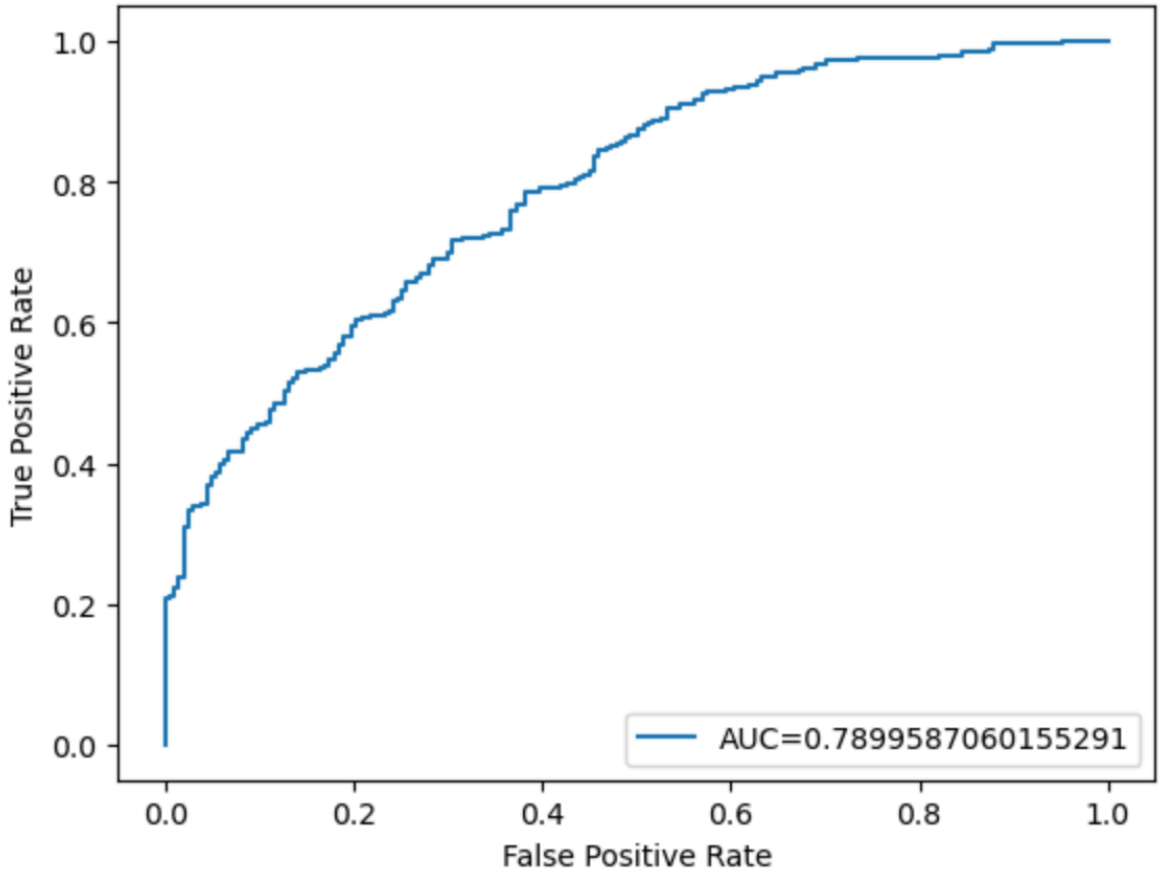
Deep Learning

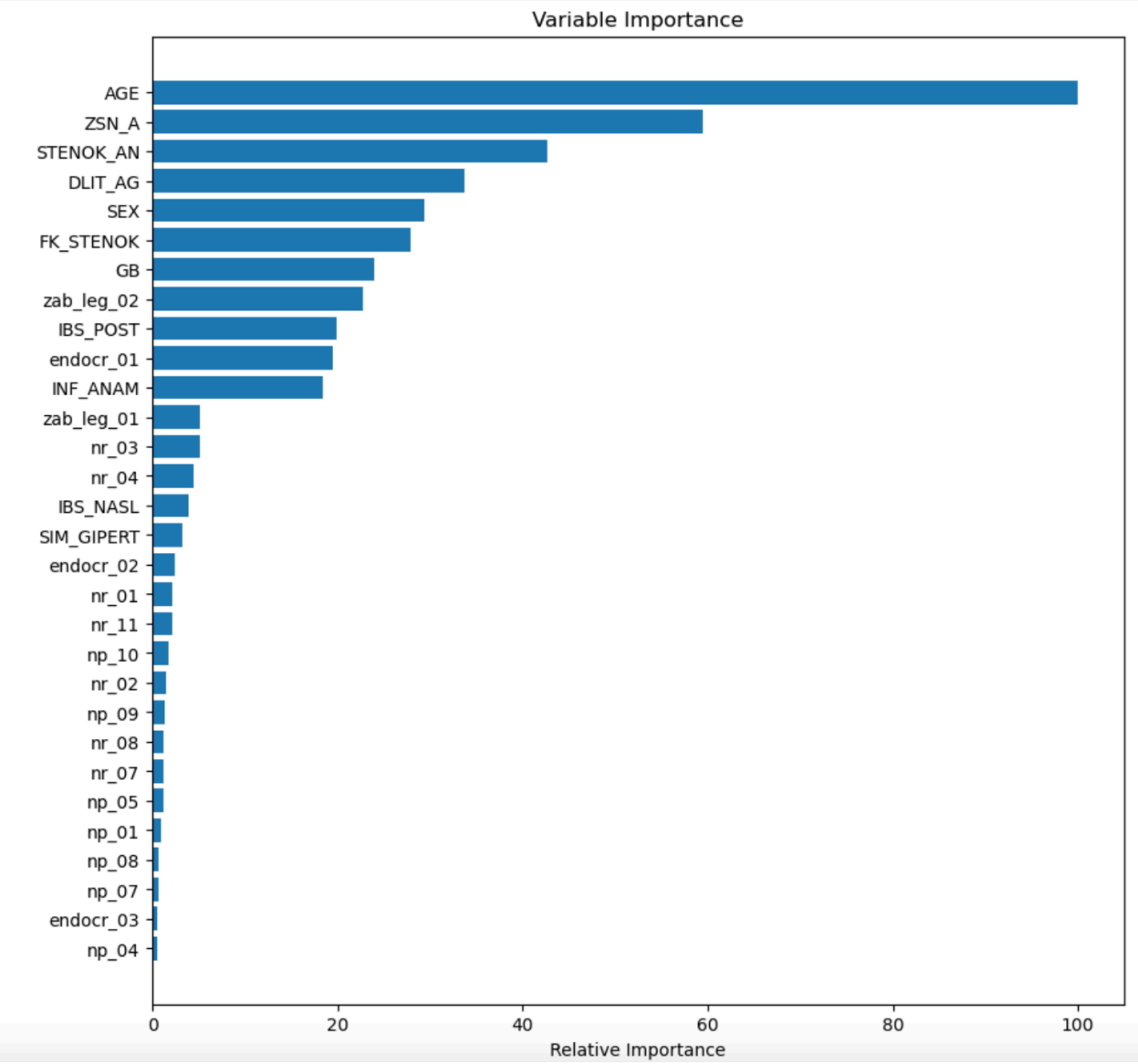
	precision	recall	f1-score	support
0	0.41	1.00	0.58	207
1	0.00	0.00	0.00	303
accuracy			0.41	510
macro avg	0.20	0.50	0.29	510
weighted avg	0.16	0.41	0.23	510

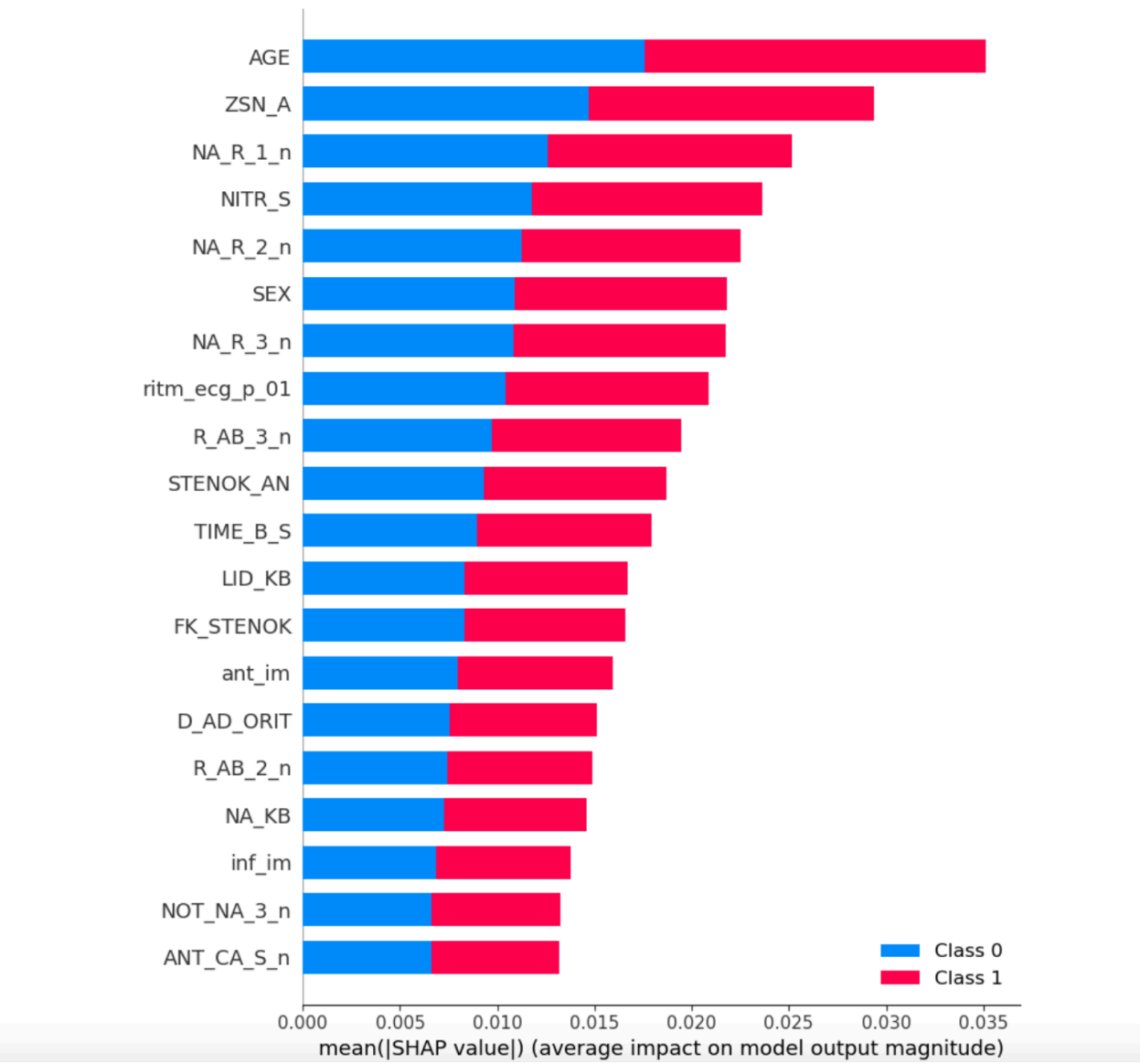


Random Forest

	precision	recall	f1-score	support
0	0.80	0.38	0.52	207
1	0.69	0.93	0.79	303
accuracy			0.71	510
macro avg	0.74	0.66	0.65	510
weighted avg	0.73	0.71	0.68	510

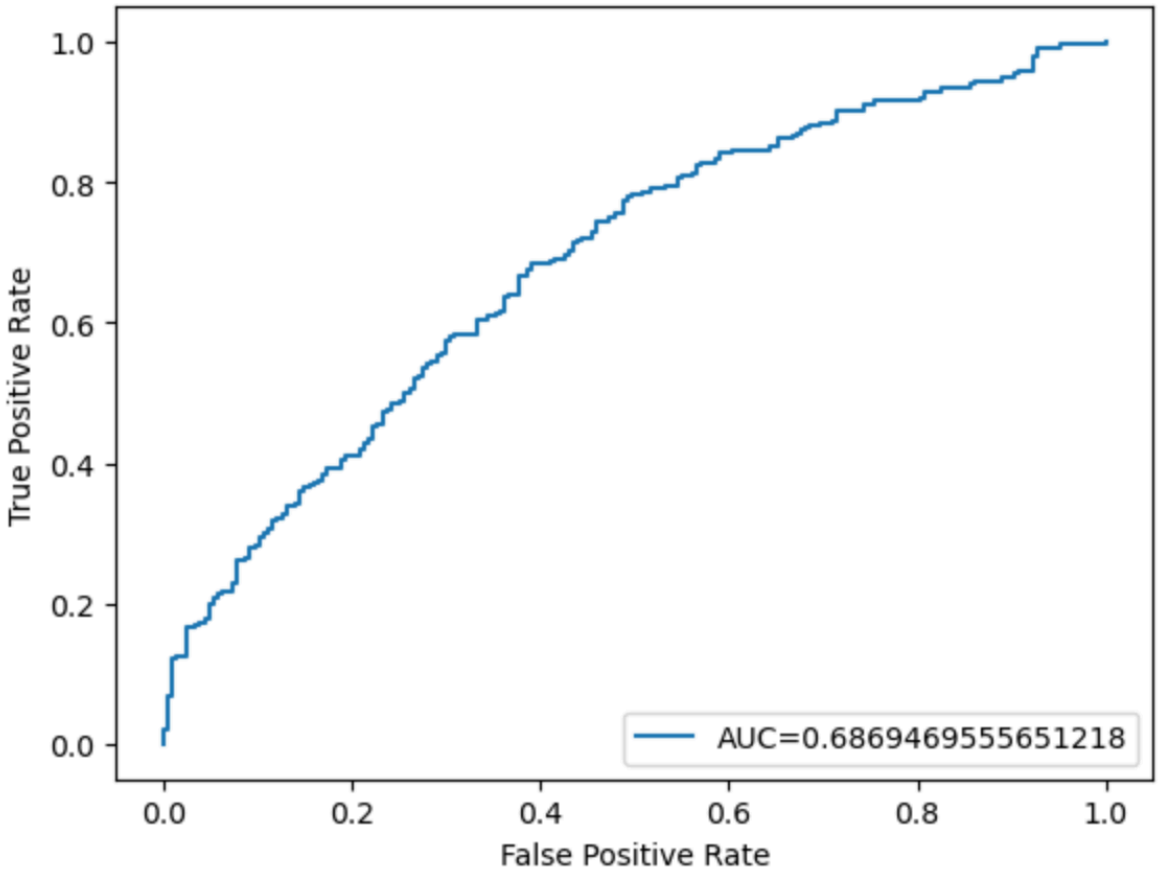

1.0
0.8
0.6
0.4
0.2
0.0
True Positive Rate
0.0
0.2
0.4
0.6
0.8
1.0
False Positive Rate
AUC=0.7899587060155291

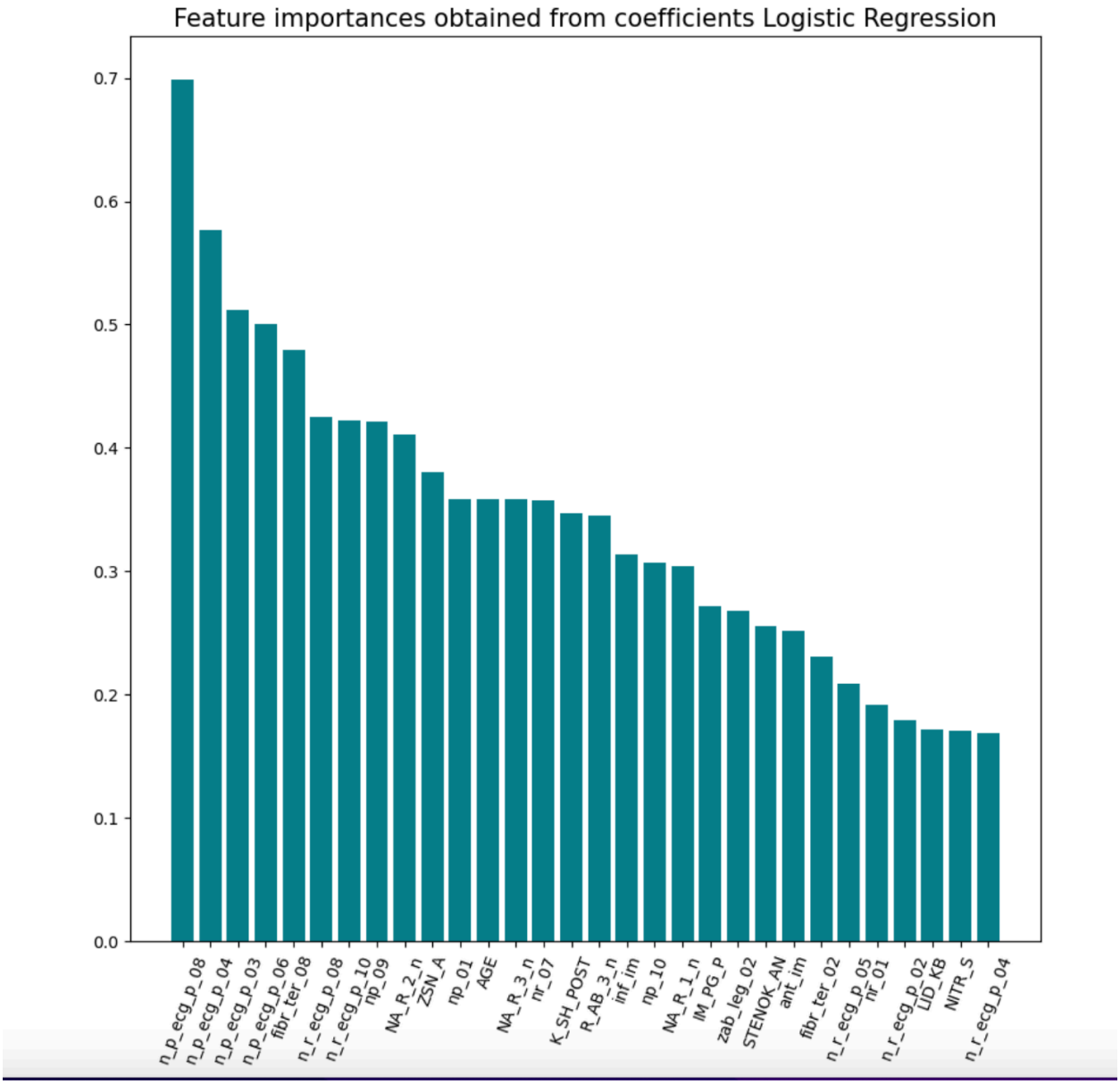

Variable Importance
AGE
ZSN_A
STENOK_AN
DLIT_AG
SEX
FK_STENOK
GB
zab_leg_02
IBS_POST
endocr_01
INF_ANAM
zab_leg_01
nr_03
nr_04
IBS_NASL
SIM_GIPERT
endocr_02
nr_01
nr_11
np_10
nr_02
np_09
nr_08
nr_07
np_05
np_01
np_08
np_07
endocr_03
np_04
0
20
40
60
80
100
Relative Importance



Logistic Regression

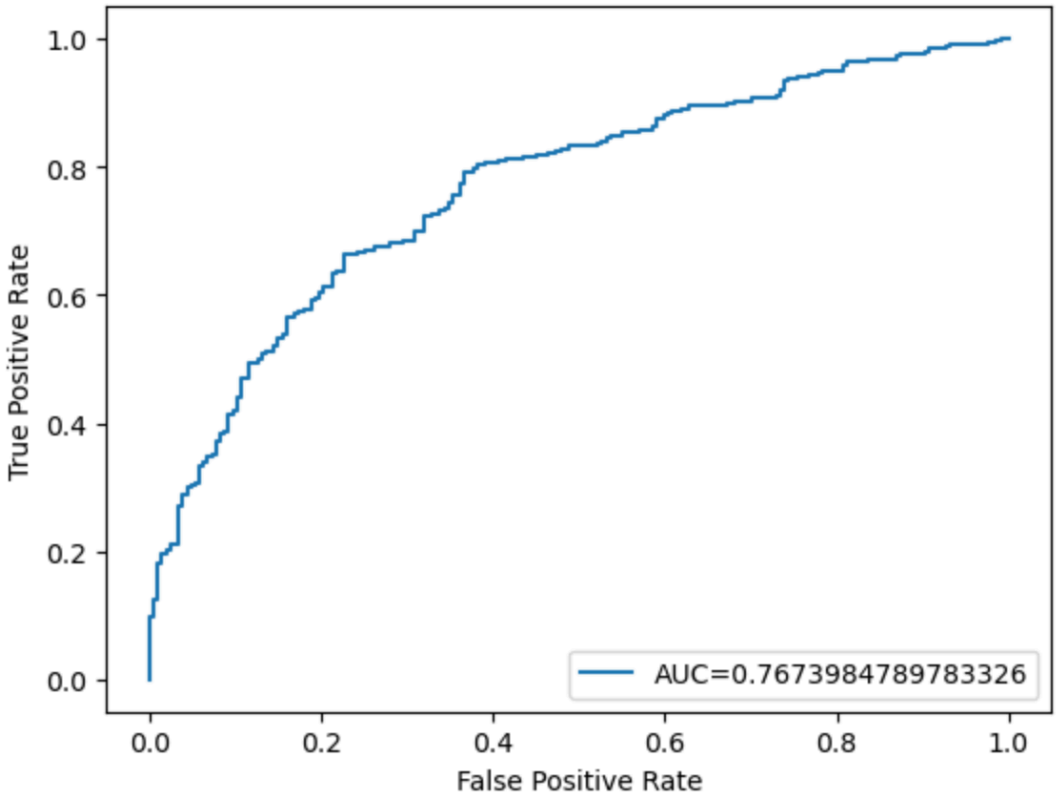
	precision	recall	f1-score	support
0	0.67	0.62	0.64	207
1	0.75	0.79	0.77	303
accuracy			0.72	510
macro avg	0.71	0.70	0.71	510
weighted avg	0.72	0.72	0.72	510


1.0
0.8
0.6
0.4
0.2
0.0
True Positive Rate
0.0
0.2
0.4
0.6
0.8
1.0
False Positive Rate
AUC=0.6869469555651218



XgBoost

	precision	recall	f1-score	support
NonExpired	0.65	0.64	0.64	207
Expired	0.76	0.77	0.76	303
accuracy			0.71	510
macro avg	0.70	0.70	0.70	510
weighted avg	0.71	0.71	0.71	510


1.0
0.8
0.6
0.4
0.2
0.0
True Positive Rate
0.0
0.2
0.4
0.6
0.8
1.0
False Positive Rate
AUC=0.7673984789783326