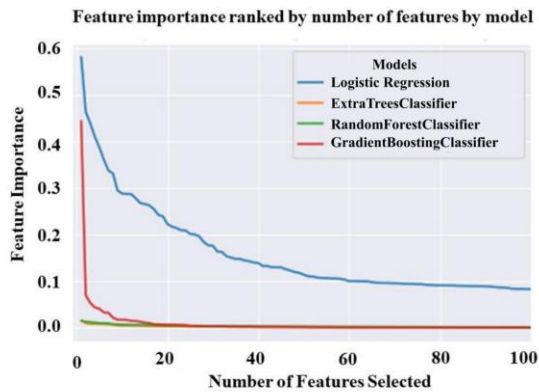
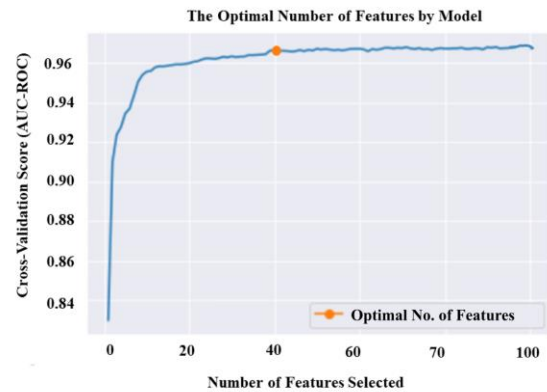


A**B**

Supplementary Figure S10. (A) This chart A shows the importance of the features (Y-axis) based on the number of features chosen (X-axis) by each model. The colored legends indicate the models used to select the features. The feature importance values are saturated with our best performing model (Gradient Boosting Classifier) when the top 100 most significant features are grouped. Thus, adding more features will not improve prediction accuracy any further. (B) In this figure, we can see an ideal RFECV curve. The curve jumps to an excellent AUC-ROC score when 40 informative features are captured, but the curve gradually becomes constant as non-informative features are incorporated.