8. Supplementary data

Table S. 1 Machine learning model parameters

Model	Parameters	AUROC	
		Train	Test
Gaussian Naïve Bayes	prior probabilities: 0.2 / 0.8	**0.72**	**0.79**
Neural Net	layers = [250, 150, 250]; loss_function=cross_entropy	**0.84**	**0.70**
Logistic regression	c=0.001, class_weight=balanced, penalty=None, solver=lbfgs	**0.71**	**0.76**
Random Forest	criterion: entropy, max_depth: 5, max_leaf_nodes: 5, min_samples_split: 0.33, n_estimators: 200	**0.82**	**0.75**
Decision Tree	max_depth: 5, max_features: auto, min_samples_split: 0.05	**0.81**	**0.57**

The table shows the chosen parameters for each model which were defined in hyperparameter tuning as described in the methods section. For Gaussian Naïve Bayes, Logistic regression, Random Forest and Decision Tree, the parameters correspond to the scikit-learn library while for the neural net they correspond to the fast.ai library. The last column provides the AUROC achieved on the training and test set respectively.