Table SI. Hyperparameter tuning grid and optimal parameters for each machine learning model.

Model	Parameters tuned	Grid search range	Optimal parameters
Logistic regression	Regularization strength (C); Penalty; Solver	C: [0.001, 0.01, 0.1, 1, 10, 100]; Penalty: ['l1', 'l2', 'elasticnet', 'none']; Solver: ['newton-cg', 'lbfgs', 'liblinear', 'sag', 'saga']	C = 100; Penalty = l2; Solver = sag
Random forest	Number of estimators; Maximum depth; Minimum samples per split; Minimum samples per leaf; Maximum features	n_estimators: [10, 50, 100, 200]; max_depth: [None, 5, 10, 20, 30]; min_samples_split: [2, 5, 10]; min_samples_leaf: [1, 2, 5]; max_features: ['sqrt', 'log2', None]	max_depth = 10; max_features = sqrt; min_samples_leaf = 1; min_samples_split = 2; n_estimators = 50
XGBoost	Number of estimators; Maximum depth; Learning rate; Subsample; Column sample by tree; Gamma	n_estimators: [50, 100, 200]; max_depth: [3, 5, 7, 10]; learning_rate: [0.01, 0.1, 0.2]; subsample: [0.6, 0.8, 1.0]; colsample_bytree: [0.6, 0.8, 1.0]; gamma: [0, 0.1, 0.2]	colsample_bytree = 0.6; gamma = 0.2; learning_rate = 0.01; max_depth = 5; n_estimators = 50; subsample = 0.6
Support vector machine	Regularization strength (C); Kernel type; Gamma	C: [0.01, 0.1, 1, 10, 100]; Kernel: ['linear', 'poly', 'rbf', 'sigmoid']; Gamma: ['scale', 'auto', 0.1, 1, 10]	C = 1; Gamma = 1; Kernel = rbf
Naïve Bayes	var_smoothing	var_smoothing: [1e-10, 1e-9, 1e-8, 1e-7, 1e-6]	var_smoothing = 1e-10

XGBoost, eXtreme Gradient Boosting.