Appendix 1. Hyperparameter search space and optimal hyperparameters for machine learning models

Model	Search space	n_iter	CV	Best hyperparameters
XGBoost	n_estimators = [100,150,200], max_depth = [3,5,7], learning_rate = [0.01,0.05,0.1], subsample = [0.8,1.0], colsample_bytree = [0.8,1.0]	5	5-fold	n_estimators = 100, max_depth = 3, learning_rate = 0.01, subsample = 0.8, colsample_bytree = 0.8, scale_pos_weight = 1.2
CatBoost	iterations = [100,150,200], depth = [3,5,7], learning_rate = [0.01,0.05,0.1], l2_leaf_reg = [1,3,5]	5	5-fold	iterations = 200, depth = 3, learning_rate = 0.05, l2_leaf_reg = 3
LightGBM	n_estimators = [100,150,200], max_depth = [3,5,7,-1], learning_rate = [0.01,0.05,0.1], num_leaves = [15,31,63], class_weight = [None,balanced]	5	5-fold	n_estimators = 200, max_depth = -1, learning_rate = 0.05, num_leaves = 15, class_weight = balanced
Logistic regression	C = [0.01,0.1,1,10], penalty = [l2], solver = [lbfgs], class_weight = [None,balanced]	5	5-fold	C = 0.1, class_weight = None
Random forest	n_estimators = [100,150,200], max_depth = [5,7,10,None], min_samples_split = [2,5], min_samples_leaf = [1,2], class_weight = [None,balanced]	5	5-fold	n_estimators = 150, max_depth = 7, min_samples_split = 2, min_samples_leaf = 1, class_weight = None

XGBoost, eXtreme gradient boosting; CatBoost, categorical boosting; LightGBM, light gradient-boosting machine.