

1 Supplementary Information

Table S1: The fourteen traditional machine learning algorithms evaluated for the classification of AMP.

Algorithm	Description
RF	Random Forest [1]
SVM	Support Vector Machine [2]
Ada	Ada boost classifier [3]
LightGBM	Light Gradient Boosting Machine [4]
GBC	Gradient Boosting Classifier [5]
ET	Extra Trees classifier [6]
LDA	Linear discriminant analysis [7]
Ridge	Ridge classifier [8]
QDA	Quadratic discriminant analysis [9]
DT	Decision Tree classifier [10]
LR	Logistic Regression [11]
NB	Naïve Bayes [12]
KNN	K Nearest Neighbor [13]
Dummy	Dummy [14]

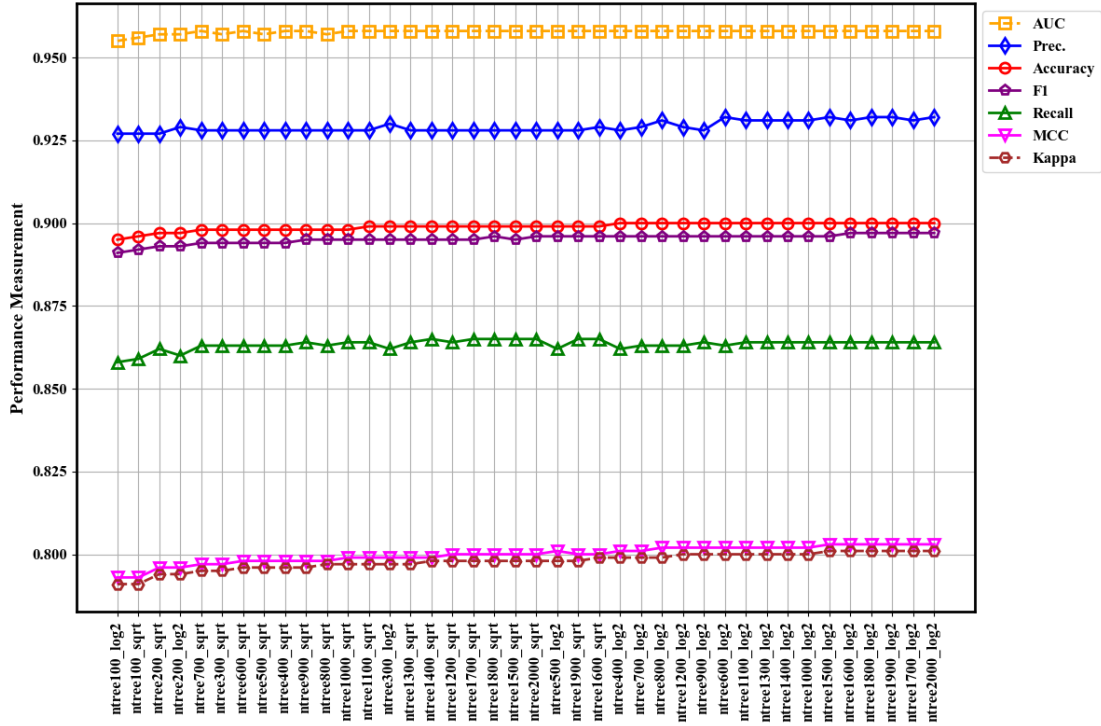


Figure S1: Grid search evaluation of ntree and max_features hyperparameters by 10-fold cross-validation in random forest models using AMP and non-AMP datasets.

2 References

1. Breiman L. Random forests. *Machine learning* 2001;45:5–32.
2. Hearst MA, Dumais ST, Osuna E, Platt J, and Scholkopf B. Support vector machines. *IEEE Intelligent Systems and their applications* 1998;13:18–28.
3. Freund Y and Schapire RE. A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of computer and system sciences* 1997;55:119–39.
4. Ke G, Meng Q, Finley T, et al. Lightgbm: A highly efficient gradient boosting decision tree. *Advances in neural information processing systems* 2017;30:3146–54.
5. Friedman JH. Greedy function approximation: a gradient boosting machine. *Annals of statistics* 2001:1189–232.
6. Geurts P, Ernst D, and Wehenkel L. Extremely randomized trees. *Machine learning* 2006;63:3–42.
7. Friedman J, Hastie T, Tibshirani R, et al. *The elements of statistical learning*. Vol. 1. 10. Springer series in statistics New York, 2001.
8. Robert C. *Machine learning, a probabilistic perspective*. 2014.
9. Wang J, Plataniotis KN, Lu J, and Venetsanopoulos AN. Kernel quadratic discriminant analysis for small sample size problem. *Pattern Recognition* 2008;41:1528–38.
10. Safavian SR and Landgrebe D. A survey of decision tree classifier methodology. *IEEE transactions on systems, man, and cybernetics* 1991;21:660–74.
11. Wright RE. Logistic regression. *Reading and understanding multivariate statistics* 1995:217–44.
12. Murphy KP et al. *Naive bayes classifiers*. University of British Columbia 2006;18:1–8.
13. Fix E and Hodges JL. Discriminatory analysis. *Nonparametric discrimination: Consistency properties*. *International Statistical Review/Revue Internationale de Statistique* 1989;57:238–47.
14. Zhang Y and Wang Y. Dummy Classifier for Imbalanced Data. *Journal of Data Science* 2019;17:345–68.