

Glaucus自动化监督性学习模块实验数据结果

I. 测试方案

使用了Balanced Accuracy进行计算评测，对于自动化预测模型来讲，与500颗规模的随机森林进行比较，在166个数据集（监督性学习数据集）上面，对比的为500颗子树的随机森林算法，算法base在sklearn基础算法包上，其他参数均使用算法默认。

II. 测试结果：

其中26组数据集上全自动化预测模型表现好于500传统RF算法，136组数据与RF算法并无太大区别，4组数据表现比RF算法差。

III. 详细结果：

A. 平衡准确率实验数据及训练选择的最佳管道

数据集	Glaucus	500-tree RF	Pipeline
GAMETES_Epistas is 2-Way_1000atts_0.4H_EDM-1_EDM-1_1	53.3%	53.0%	KNeighborsClassifier(n_neighbors=62, p=1, weights="uniform")
GAMETES_Epistas is 2-Way_20atts_0.1H_EDM-1_1	60.6%	60.4%	make_pipeline(FastICA(tol=0.1), Normalizer(norm="l1"), DecisionTreeClassifier(criterion="gini", max_depth=9, min_samples_leaf=3, min_samples_split=13))
GAMETES_Epistas is 2-Way_20atts_0.4H_EDM-1_1	67.6%	67.3%	make_pipeline(StackingEstimator(estimator=RandomForestClassifier(bootstrap=True, criterion="gini", max_features=0.5, min_samples_leaf=6, min_samples_split=10, n_estimators=100)), GaussianNB())
GAMETES_Epistas is 3-Way_20atts_0.2H_EDM-1_1	56.7%	50.3%	make_pipeline(PolynomialFeatures(degree=2, include_bias=False, interaction_only=False), GradientBoostingClassifier(learning_rate=0.01, max_depth=8, max_features=0.4, min_samples_leaf=14, min_samples_split=4, n_estimators=100, subsample=0.9000000000000001))

数据集	Glaucus	500-tree RF	Pipeline
GAMETES_Heterogeneity_20atts_1600_Het_0.4_0.2_50_EDM-2_001	71.9%	67.3%	<pre> make_pipeline(PolynomialFeatures(degree=2, include_bias=False, interaction_only=False), RandomForestClassifier(bootstrap=True, criterion="entropy", max_features=0.25, min_samples_leaf=18, min_samples_split=12, n_estimators=100)) </pre>
GAMETES_Heterogeneity_20atts_1600_Het_0.4_0.2_75_EDM-2_001	72.6%	66.5%	<pre> make_pipeline(PolynomialFeatures(degree=2, include_bias=False, interaction_only=False), DecisionTreeClassifier(criterion="entropy", max_depth=8, min_samples_leaf=19, min_samples_split=9)) </pre>
Hill_Valley_with_noise	99.8%	57.1%	<pre> make_pipeline(Normalizer(norm="max"), KNeighborsClassifier(n_neighbors=12, p=2, weights="distance")) </pre>
Hill_Valley_without_noise	96.6%	61.0%	<code>LogisticRegression(C=10.0, dual=False, penalty="l2")</code>
adult	86.5%	86.1%	<code>GradientBoostingClassifier(learning_rate=1.0, max_depth=1, max_features=0.55, min_samples_leaf=18, min_samples_split=18, n_estimators=100, subsample=0.90000000000000001)</code>
agaricus-lepiota	100%	99.7%	<code>ExtraTreesClassifier(bootstrap=True, criterion="entropy", max_features=0.75000000000000001, min_samples_leaf=10, min_samples_split=8, n_estimators=100)</code>
allbp	97.6%	97.1%	<code>GradientBoostingClassifier(learning_rate=0.5, max_depth=8, max_features=0.25, min_samples_leaf=14, min_samples_split=14, n_estimators=100, subsample=0.85000000000000001)</code>
allhyper	98.3%	98.8%	<pre> make_pipeline(StackingEstimator(estimator=ExtraTreesC lassifier(bootstrap=True, criterion="entropy", max_features=1.0, min_samples_leaf=10, min_samples_split=17, n_estimators=100)), ExtraTreesClassifier(bootstrap=False, criterion="entropy", max_features=0.70000000000000001, min_samples_leaf=11, min_samples_split=17, n_estimators=100)) </pre>

数据集	Glaucus	500-tree RF	Pipeline
allhypo	96.9%	81.6%	<pre> make_pipeline(PolynomialFeatures(degree=2, include_bias=False, interaction_only=False), StackingEstimator(estimator=ExtraTreesClassifier(bootstrap=True, criterion="entropy", max_features=0.6500000000000001, min_samples_leaf=3, min_samples_split=15, n_estimators=100)), LinearSVC(C=5.0, dual=False, loss="squared_hinge", penalty="l1", tol=0.1)) </pre>
allrep	98.6%	98.0%	<pre> GradientBoostingClassifier(learning_rate=0.5, max_depth=4, max_features=0.8, min_samples_leaf=19, min_samples_split=16, n_estimators=100, subsample=0.55) </pre>
analcatsdata_aids	74.5%	75.0%	<pre> LogisticRegression(C=0.1, dual=False, penalty="l1") </pre>
analcatsdata_asbestos	86.5%	87.0%	<pre> GaussianNB() </pre>
analcatsdata_authorsip	99.8%	99.2%	<pre> make_pipeline(StackingEstimator(estimator=ExtraTreesClassifier(bootstrap=True, criterion="entropy", max_features=0.35000000000000003, min_samples_leaf=4, min_samples_split=11, n_estimators=100)), LinearSVC(C=0.001, dual=False, loss="squared_hinge", penalty="l2", tol=0.0001)) </pre>
analcatsdata_bankruptcy	94.3%	94.1%	<pre> make_pipeline(RFE(estimator=ExtraTreesClassifier(criterion="gini", max_features=0.4, n_estimators=100), step=0.8), LinearSVC(C=0.01, dual=False, loss="squared_hinge", penalty="l1", tol=0.0001)) </pre>
analcatsdata_boxing1	75.4%	75.1%	<pre> make_pipeline(StackingEstimator(estimator=DecisionTreeClassifier(criterion="gini", max_depth=6, min_samples_leaf=7, min_samples_split=10)), GaussianNB()) </pre>
analcatsdata_boxing2	72.1%	72.6%	<pre> DecisionTreeClassifier(criterion="gini", max_depth=1, min_samples_leaf=11, min_samples_split=10) </pre>
analcatsdata_credit score	100%	100%	<pre> DecisionTreeClassifier(criterion="gini", max_depth=8, min_samples_leaf=5, min_samples_split=13) </pre>
analcatsdata_cycling8092	92.5%	92.7%	<pre> LogisticRegression(C=1.0, dual=False, penalty="l1") </pre>

数据集	Glaucus	500-tree RF	Pipeline
analcata_data_cyyoung9302	88.9%	88.5%	make_pipeline(StackingEstimator(estimator=RandomForestClassifier(bootstrap=False, criterion="gini", max_features=0.2, min_samples_leaf=5, min_samples_split=4, n_estimators=100)), RandomForestClassifier(bootstrap=False, criterion="gini", max_features=0.5, min_samples_leaf=8, min_samples_split=10, n_estimators=100))
analcata_data_dmft	24.2%	24.7%	GradientBoostingClassifier(learning_rate=0.1, max_depth=7, max_features=0.1, min_samples_leaf=16, min_samples_split=5, n_estimators=100, subsample=0.6500000000000001)
analcata_data_fraud	92.6%	92.4%	make_pipeline(make_union(FunctionTransformer(copy), FunctionTransformer(copy)), RandomForestClassifier(bootstrap=True, criterion="entropy", max_features=0.35000000000000003, min_samples_leaf=4, min_samples_split=3, n_estimators=100))
analcata_data_german_credit	41.5%	35.1%	DecisionTreeClassifier(criterion="gini", max_depth=9, min_samples_leaf=15, min_samples_split=16)
analcata_data_happiness	63.9%	45.1%	make_pipeline(RobustScaler(), GradientBoostingClassifier(learning_rate=0.5, max_depth=9, max_features=0.1, min_samples_leaf=4, min_samples_split=13, n_estimators=100, subsample=0.5))
analcata_data_japanese_vowel	88.9%	88.9%	RandomForestClassifier(bootstrap=False, criterion="gini", max_features=0.2, min_samples_leaf=14, min_samples_split=19, n_estimators=100)
analcata_data_lawsuit	98.4%	97.9%	make_pipeline(FeatureAgglomeration(affinity="precomputed", linkage="average"), GaussianNB())
ann-thyroid	99.5%	99.1%	RandomForestClassifier(bootstrap=False, criterion="entropy", max_features=0.6500000000000001, min_samples_leaf=5, min_samples_split=20, n_estimators=100)
appendicitis	87.9%	87.9%	RandomForestClassifier(bootstrap=False, criterion="entropy", max_features=0.05, min_samples_leaf=5, min_samples_split=19, n_estimators=100)

数据集	Glaucus	500-tree RF	Pipeline
australian	87.3%	87.7%	GradientBoostingClassifier(learning_rate=0.1, max_depth=4, max_features=0.15000000000000002, min_samples_leaf=11, min_samples_split=6, n_estimators=100, subsample=0.65000000000000001)
auto	72.9%	72.7%	GradientBoostingClassifier(learning_rate=0.1, max_depth=1, max_features=0.60000000000000001, min_samples_leaf=1, min_samples_split=7, n_estimators=100, subsample=0.90000000000000001)
backache	84.8%	84.5%	KNeighborsClassifier(n_neighbors=58, p=1, weights="distance")
balance-scale	90.9%	67.2%	make_pipeline(PolynomialFeatures(degree=2, include_bias=False, interaction_only=False), LogisticRegression(C=0.01, dual=False, penalty="l1"))
banana	90.2%	89.2%	make_pipeline(ZeroCount(), ExtraTreesClassifier(bootstrap=True, criterion="entropy", max_features=0.90000000000000001, min_samples_leaf=2, min_samples_split=5, n_estimators=100))
biomed	93.1%	92.9%	make_pipeline(StackingEstimator(estimator=KNeighborsClassifier(n_neighbors=1, p=1, weights="uniform")), RandomForestClassifier(bootstrap=False, criterion="gini", max_features=0.95000000000000001, min_samples_leaf=8, min_samples_split=12, n_estimators=100))
breast-cancer-wisconsin	97.0%	96.7%	make_pipeline(MaxAbsScaler(), LogisticRegression(C=10.0, dual=False, penalty="l2"))
breast-cancer	75.0%	80.2%	make_pipeline(StackingEstimator(estimator=LinearSVC(C=10.0, dual=False, loss="squared_hinge", penalty="l1", tol=0.0001)), FastICA(tol=0.5), RandomForestClassifier(bootstrap=True, criterion="entropy", max_features=0.75000000000000001, min_samples_leaf=9, min_samples_split=4, n_estimators=100))
breast-w	97.1%	96.8%	make_pipeline(StackingEstimator(estimator=GaussianNB()), KNeighborsClassifier(n_neighbors=24, p=1, weights="distance"))

数据集	Glaucus	500-tree RF	Pipeline
breast	97.3%	97.8%	GradientBoostingClassifier(learning_rate=0.1, max_depth=9, max_features=1.0, min_samples_leaf=5, min_samples_split=18, n_estimators=100, subsample=0.6500000000000001)
buggyCrx	88.0%	88.1%	GradientBoostingClassifier(learning_rate=0.01, max_depth=10, max_features=0.3, min_samples_leaf=2, min_samples_split=6, n_estimators=100, subsample=0.6500000000000001)
bupa	69.7%	69.4%	ExtraTreesClassifier(bootstrap=False, criterion="entropy", max_features=0.55, min_samples_leaf=4, min_samples_split=6, n_estimators=100)
calendarDOW	61.7%	62.1%	make_pipeline(VarianceThreshold(threshold=0.25), MaxAbsScaler(), KNeighborsClassifier(n_neighbors=9, p=2, weights="distance"))
car-evaluation	97.8%	92.9%	GradientBoostingClassifier(learning_rate=0.5, max_depth=6, max_features=0.6500000000000001, min_samples_leaf=5, min_samples_split=3, n_estimators=100, subsample=0.9000000000000001)
car	98.7%	98.2%	make_pipeline(StackingEstimator(estimator=GradientBoostingClassifier(learning_rate=0.5, max_depth=4, max_features=0.7500000000000001, min_samples_leaf=1, min_samples_split=8, n_estimators=100, subsample=0.7000000000000001)), GaussianNB())
cars	90.9%	90.9%	RandomForestClassifier(bootstrap=True, criterion="gini", max_features=1.0, min_samples_leaf=3, min_samples_split=6, n_estimators=100)
cars1	83.2%	83.3%	RandomForestClassifier(bootstrap=False, criterion="entropy", max_features=0.7500000000000001, min_samples_leaf=4, min_samples_split=8, n_estimators=100)
chess	99.2%	99.7%	ExtraTreesClassifier(bootstrap=False, criterion="entropy", max_features=0.4, min_samples_leaf=2, min_samples_split=3, n_estimators=100)
churn	95.7%	95.2%	ExtraTreesClassifier(bootstrap=False, criterion="entropy", max_features=0.6500000000000001, min_samples_leaf=1, min_samples_split=2, n_estimators=100)
clean1	100%	100%	LogisticRegression(C=0.01, dual=False, penalty="l1")

数据集	Glaucus	500-tree RF	Pipeline
clean2	100%	100%	LogisticRegression(C=0.5, dual=False, penalty="l2")
cleve	83.5%	83.1%	make_pipeline(ZeroCount(), OneHotEncoder(minimum_fraction=0.2, sparse=False), BernoulliNB(alpha=10.0, fit_prior=False))
cleveland-nominal	59.1%	59.5%	KNeighborsClassifier(n_neighbors=27, p=2, weights="distance")
cleveland	61.3%	61.8%	make_pipeline(StackingEstimator(estimator=GaussianNB()), DecisionTreeClassifier(criterion="gini", max_depth=2, min_samples_leaf=13, min_samples_split=2))
cloud	63.8%	49.5%	KNeighborsClassifier(n_neighbors=53, p=2, weights="distance")
cmc	55.8%	55.5%	ExtraTreesClassifier(bootstrap=False, criterion="entropy", max_features=0.9000000000000001, min_samples_leaf=18, min_samples_split=18, n_estimators=100)
coil2000	94.0%	93.6%	DecisionTreeClassifier(criterion="entropy", max_depth=3, min_samples_leaf=14, min_samples_split=5)
colic	86.8%	86.9%	DecisionTreeClassifier(criterion="gini", max_depth=4, min_samples_leaf=18, min_samples_split=18)
collins	100%	97%	GaussianNB()
confidence	81.9%	81.7%	make_pipeline(PolynomialFeatures(degree=2, include_bias=False, interaction_only=False), KNeighborsClassifier(n_neighbors=2, p=2, weights="distance"))
connect-4	83.6%	83.4%	GradientBoostingClassifier(learning_rate=0.5, max_depth=10, max_features=0.35000000000000003, min_samples_leaf=16, min_samples_split=12, n_estimators=100, subsample=0.7500000000000001)
contraceptive	56.8%	56.5%	make_pipeline(OneHotEncoder(minimum_fraction=0.05, sparse=False), GradientBoostingClassifier(learning_rate=0.1, max_depth=2, max_features=0.7000000000000001, min_samples_leaf=10, min_samples_split=20, n_estimators=100, subsample=0.6000000000000001))
corral	100%	100%	GradientBoostingClassifier(learning_rate=0.1, max_depth=9, max_features=0.6500000000000001, min_samples_leaf=13, min_samples_split=15, n_estimators=100, subsample=0.9000000000000001)

数据集	Glaucus	500-tree RF	Pipeline
credit-a	88.6%	88.2%	LogisticRegression(C=10.0, dual=False, penalty="l1")
credit-g	74.7%	74.2%	RandomForestClassifier(bootstrap=True, criterion="gini", max_features=0.7500000000000001, min_samples_leaf=5, min_samples_split=19, n_estimators=100)
crx	86.1%	85.7%	make_pipeline(StackingEstimator(estimator=BernoulliNB(alpha=100.0, fit_prior=False)), GradientBoostingClassifier(learning_rate=0.01, max_depth=4, max_features=0.2, min_samples_leaf=8, min_samples_split=19, n_estimators=100, subsample=0.45))
dermatology	98%	98.1%	RandomForestClassifier(bootstrap=False, criterion="entropy", max_features=0.1, min_samples_leaf=2, min_samples_split=8, n_estimators=100)
diabetes	78.6%	78.1%	ExtraTreesClassifier(bootstrap=True, criterion="entropy", max_features=0.9000000000000001, min_samples_leaf=4, min_samples_split=17, n_estimators=100)
dis	99.0%	74.2%	DecisionTreeClassifier(criterion="entropy", max_depth=9, min_samples_leaf=14, min_samples_split=11)
dna	95.6%	95.2%	make_pipeline(SelectFwe(score_func=f_classif, alpha=0.048), StackingEstimator(estimator=LinearSVC(C=0.0001, dual=False, loss="squared_hinge", penalty="l1", tol=0.01)), BernoulliNB(alpha=0.001, fit_prior=True))
ecoli	88.2%	92.3%	make_pipeline(PolynomialFeatures(degree=2, include_bias=False, interaction_only=False), LogisticRegression(C=15.0, dual=False, penalty="l2"))
fars	79.9%	80.3%	make_pipeline(make_union(FunctionTransformer(copy), FunctionTransformer(copy)), ExtraTreesClassifier(bootstrap=False, criterion="gini", max_features=0.8500000000000001, min_samples_leaf=20, min_samples_split=15, n_estimators=100))
flags	52.1%	51.8%	GradientBoostingClassifier(learning_rate=0.01, max_depth=10, max_features=0.6000000000000001, min_samples_leaf=3, min_samples_split=11, n_estimators=100, subsample=0.9000000000000001)

数据集	Glaucus	500-tree RF	Pipeline
flare	83.6%	83.2%	<pre> make_pipeline(OneHotEncoder(minimum_fraction=0.1, sparse=False), Binarizer(threshold=0.9500000000000001), ExtraTreesClassifier(bootstrap=True, criterion="entropy", max_features=0.6500000000000001, min_samples_leaf=18, min_samples_split=2, n_estimators=100)) </pre>
german	74.3%	74.7%	<pre> make_pipeline(make_union(FunctionTransformer(copy), FunctionTransformer(copy)), RandomForestClassifier(bootstrap=True, criterion="entropy", max_features=0.6500000000000001, min_samples_leaf=1, min_samples_split=16, n_estimators=100)) </pre>
glass	69.1%	69.2%	<pre> ExtraTreesClassifier(bootstrap=True, criterion="entropy", max_features=0.5, min_samples_leaf=6, min_samples_split=7, n_estimators=100) </pre>
glass2	80.8%	80.7%	<pre> ExtraTreesClassifier(bootstrap=False, criterion="entropy", max_features=0.45, min_samples_leaf=5, min_samples_split=3, n_estimators=100) </pre>
haberman	79.4%	79.7%	<pre> make_pipeline(ZeroCount(), GaussianNB()) </pre>
hayes-roth	83.9%	83.7%	<pre> ExtraTreesClassifier(bootstrap=False, criterion="entropy", max_features=0.5, min_samples_leaf=1, min_samples_split=5, n_estimators=100) </pre>
heart-c	86.3%	85.9%	<pre> LogisticRegression(C=15.0, dual=False, penalty="l2") </pre>
heart-h	83.4%	83.9%	<pre> make_pipeline(RobustScaler(), KNeighborsClassifier(n_neighbors=5, p=2, weights="uniform")) </pre>
heart-statlog	86.8%	86.8%	<pre> RandomForestClassifier(bootstrap=False, criterion="gini", max_features=0.3, min_samples_leaf=14, min_samples_split=4, n_estimators=100) </pre>
hepatitis	88.2%	87.8%	<pre> GradientBoostingClassifier(learning_rate=0.1, max_depth=6, max_features=0.2, min_samples_leaf=6, min_samples_split=17, n_estimators=100, subsample=0.9500000000000001) </pre>
horse-colic	86.8%	86.4%	<pre> make_pipeline(StandardScaler(), DecisionTreeClassifier(criterion="gini", max_depth=2, min_samples_leaf=19, min_samples_split=17)) </pre>

数据集	Glaucus	500-tree RF	Pipeline
house-votes-84	95.7%	95.9%	DecisionTreeClassifier(criterion="gini", max_depth=7, min_samples_leaf=14, min_samples_split=16)
hungarian	84.4%	84.8%	make_pipeline(ZeroCount(), StackingEstimator(estimator=GaussianNB()), ExtraTreesClassifier(bootstrap=True, criterion="gini", max_features=0.5, min_samples_leaf=1, min_samples_split=12, n_estimators=100))
hypothyroid	98.3%	98.4%	make_pipeline(StackingEstimator(estimator=ExtraTreesClassifier(bootstrap=False, criterion="entropy", max_features=0.65000000000000001, min_samples_leaf=1, min_samples_split=14, n_estimators=100)), ExtraTreesClassifier(bootstrap=True, criterion="entropy", max_features=0.8, min_samples_leaf=7, min_samples_split=7, n_estimators=100))
ionosphere	93.5%	93.3%	make_pipeline(StackingEstimator(estimator=RandomForestClassifier(bootstrap=True, criterion="entropy", max_features=0.55, min_samples_leaf=1, min_samples_split=4, n_estimators=100)), DecisionTreeClassifier(criterion="gini", max_depth=6, min_samples_leaf=15, min_samples_split=17))
iris	98.1%	98.3%	make_pipeline(Normalizer(norm="l1"), GaussianNB())
irish	100%	100%	DecisionTreeClassifier(criterion="gini", max_depth=9, min_samples_leaf=18, min_samples_split=12)
kddcup	99.96%	99.9%	RandomForestClassifier(bootstrap=True, criterion="gini", max_features=0.15000000000000002, min_samples_leaf=2, min_samples_split=10, n_estimators=100)
kr-vs-kp	99.2%	99.6%	GradientBoostingClassifier(learning_rate=0.1, max_depth=6, max_features=0.60000000000000001, min_samples_leaf=17, min_samples_split=8, n_estimators=100, subsample=0.85000000000000001)
krkopt	76.6%	76.6%	RandomForestClassifier(bootstrap=True, criterion="gini", max_features=0.35000000000000003, min_samples_leaf=1, min_samples_split=5, n_estimators=100)

数据集	Glaucus	500-tree RF	Pipeline
labor	93.1%	92.1%	ExtraTreesClassifier(bootstrap=False, criterion="gini", max_features=0.90000000000000001, min_samples_leaf=4, min_samples_split=20, n_estimators=100)
led24	73.6%	73.7%	GradientBoostingClassifier(learning_rate=0.001, max_depth=8, max_features=0.45, min_samples_leaf=19, min_samples_split=3, n_estimators=100, subsample=0.15000000000000002)
led7	74.1%	73.7%	make_pipeline(SelectFwe(score_func=f_classif, alpha=0.012), StackingEstimator(estimator=GradientBoostingClassifier(l earning_rate=0.1, max_depth=5, max_features=0.3, min_samples_leaf=4, min_samples_split=20, n_estimators=100, subsample=0.90000000000000001)), BernoulliNB(alpha=100.0, fit_prior=False))
letter	90.7%	90.6%	make_pipeline(PolynomialFeatures(degree=2, include_bias=False, interaction_only=False), StackingEstimator(estimator=GradientBoostingClassifier(l earning_rate=0.5, max_depth=1, max_features=0.75000000000000001, min_samples_leaf=13, min_samples_split=11, n_estimators=100, subsample=0.15000000000000002)), RandomForestClassifier(bootstrap=False, criterion="entropy", max_features=0.35000000000000003, min_samples_leaf=20, min_samples_split=4, n_estimators=100))
liver-disorder	73.1%	73.2%	KNeighborsClassifier(n_neighbors=19, p=2, weights="distance")
lupus	78.3%	78.5%	GradientBoostingClassifier(learning_rate=0.01, max_depth=4, max_features=0.75000000000000001, min_samples_leaf=9, min_samples_split=3, n_estimators=100, subsample=0.5)
lymphography	81.7%	80.9%	make_pipeline(StackingEstimator(estimator=RandomForestClassifier(bo otstrap=True, criterion="gini", max_features=0.35000000000000003, min_samples_leaf=5, min_samples_split=16, n_estimators=100)), ExtraTreesClassifier(bootstrap=False, criterion="gini", max_features=0.90000000000000001, min_samples_leaf=14, min_samples_split=7, n_estimators=100))

数据集	Glaucus	500-tree RF	Pipeline
magic	87.7%	87.6%	GradientBoostingClassifier(learning_rate=0.1, max_depth=4, max_features=0.7500000000000001, min_samples_leaf=17, min_samples_split=6, n_estimators=100, subsample=0.9500000000000001)
mfeat-factors	97.3%	97.5%	LogisticRegression(C=5.0, dual=False, penalty="l1")
mfeat-fourier	84.2%	83.7%	make_pipeline(SelectFwe(score_func=f_classif, alpha=0.015), ExtraTreesClassifier(bootstrap=False, criterion="entropy", max_features=0.7000000000000001, min_samples_leaf=1, min_samples_split=12, n_estimators=100))
mfeat-karhunen	95.6%	95.2%	make_pipeline(StackingEstimator(estimator=ExtraTreesClassifier(bootstrap=False, criterion="entropy", max_features=0.15000000000000002, min_samples_leaf=4, min_samples_split=10, n_estimators=100)), GaussianNB())
mfeat-morphological	74.2%	74.7%	make_pipeline(StackingEstimator(estimator=GradientBoostingClassifier(learning_rate=0.01, max_depth=9, max_features=0.2, min_samples_leaf=14, min_samples_split=2, n_estimators=100, subsample=0.05)), LogisticRegression(C=20.0, dual=False, penalty="l1"))
mfeat-pixel	96.7%	96.8%	KNeighborsClassifier(n_neighbors=7, p=1, weights="distance")
mfeat-zernike	81.4%	81.1%	KNeighborsClassifier(n_neighbors=16, p=2, weights="uniform")
mnist	95.8%	95.7%	make_pipeline(MaxAbsScaler(), ExtraTreesClassifier(bootstrap=False, criterion="entropy", max_features=0.1, min_samples_leaf=9, min_samples_split=16, n_estimators=100))
mofn-3-7-10	100%	96.3%	XGBClassifier(learning_rate=0.5, max_depth=2, min_child_weight=11, n_estimators=100, nthread=1, subsample=0.8500000000000001)

数据集	Glaucus	500-tree RF	Pipeline
molecular- biology_promoters	100%	100%	<pre> make_pipeline(RFE(estimator=ExtraTreesClassifier(criterion="gini", max_features=0.1, n_estimators=100), step=0.9500000000000001), StackingEstimator(estimator=GradientBoostingClassifier(learning_rate=0.5, max_depth=10, max_features=0.6500000000000001, min_samples_leaf=10, min_samples_split=19, n_estimators=100, subsample=1.0)), RandomForestClassifier(bootstrap=True, criterion="gini", max_features=0.3500000000000003, min_samples_leaf=20, min_samples_split=11, n_estimators=100)) </pre>
monk1	99.7%	97.3%	<pre> make_pipeline(PolynomialFeatures(degree=2, include_bias=False, interaction_only=False), StackingEstimator(estimator=GradientBoostingClassifier(learning_rate=0.5, max_depth=6, max_features=0.15000000000000002, min_samples_leaf=4, min_samples_split=2, n_estimators=100, subsample=0.8)), GaussianNB()) </pre>
monk2	87.6%	83.2%	<pre> make_pipeline(StackingEstimator(estimator=LogisticRegression(C=15.0, dual=True, penalty="l2")), StackingEstimator(estimator=GradientBoostingClassifier(learning_rate=0.5, max_depth=7, max_features=0.7500000000000001, min_samples_leaf=11, min_samples_split=8, n_estimators=100, subsample=0.4)), ExtraTreesClassifier(bootstrap=True, criterion="entropy", max_features=0.45, min_samples_leaf=2, min_samples_split=15, n_estimators=100)) </pre>
monk3	98.7%	99.1%	<pre> make_pipeline(StackingEstimator(estimator=DecisionTreeClassifier(criteri on="gini", max_depth=7, min_samples_leaf=2, min_samples_split=10)), StackingEstimator(estimator=DecisionTreeClassifier(criteri on="gini", max_depth=3, min_samples_leaf=5, min_samples_split=13)), LogisticRegression(C=0.1, dual=False, penalty="l2")) </pre>
movement_libras	70.8%	71.1%	<pre> LogisticRegression(C=15.0, dual=False, penalty="l1") </pre>

数据集	Glaucus	500-tree RF	Pipeline
mushroom	100%	100%	GradientBoostingClassifier(learning_rate=1.0, max_depth=1, max_features=0.45, min_samples_leaf=18, min_samples_split=2, n_estimators=100, subsample=0.35000000000000003)
mux6	85.3%	85.5%	KNeighborsClassifier(n_neighbors=66, p=1, weights="distance")
new-thyroid	96.1%	95.8%	make_pipeline(StackingEstimator(estimator=LogisticRegression(C=0.1, dual=False, penalty="l2")), StackingEstimator(estimator=RandomForestClassifier(bootstrap=False, criterion="gini", max_features=0.35000000000000003, min_samples_leaf=20, min_samples_split=16, n_estimators=100)), StackingEstimator(estimator=RandomForestClassifier(bootstrap=True, criterion="entropy", max_features=0.4, min_samples_leaf=6, min_samples_split=10, n_estimators=100)), GaussianNB())
nursery	99.9%	100%	GradientBoostingClassifier(learning_rate=0.1, max_depth=6, max_features=0.9500000000000001, min_samples_leaf=19, min_samples_split=6, n_estimators=100, subsample=0.6000000000000001)
optdigits	97.5%	98.0%	make_pipeline(StandardScaler(), StackingEstimator(estimator=KNeighborsClassifier(n_neighbors=8, p=1, weights="distance")), RandomForestClassifier(bootstrap=True, criterion="gini", max_features=0.35000000000000003, min_samples_leaf=3, min_samples_split=7, n_estimators=100))
page-blocks	97.2%	97.2%	RandomForestClassifier(bootstrap=False, criterion="gini", max_features=0.55, min_samples_leaf=9, min_samples_split=12, n_estimators=100)
parity5+5	63.2%	32.3%	make_pipeline(PCA(iterated_power=3, svd_solver="randomized"), RandomForestClassifier(bootstrap=False, criterion="entropy", max_features=0.8500000000000001, min_samples_leaf=7, min_samples_split=7, n_estimators=100))

数据集	Glaucus	500-tree RF	Pipeline
parity5	65.9%	0.1%	<pre> make_pipeline(StackingEstimator(estimator=ExtraTreesClassifier(bootstrap=True, criterion="entropy", max_features=0.7000000000000001, min_samples_leaf=7, min_samples_split=8, n_estimators=100)), KNeighborsClassifier(n_neighbors=11, p=2, weights="uniform")) </pre>
pendigits	98.8%	99.1%	<pre> ExtraTreesClassifier(bootstrap=False, criterion="entropy", max_features=0.6000000000000001, min_samples_leaf=6, min_samples_split=14, n_estimators=100) </pre>
phoneme	88.8%	89%	<pre> make_pipeline(make_union(FunctionTransformer(copy), FunctionTransformer(copy)), ExtraTreesClassifier(bootstrap=False, criterion="entropy", max_features=0.8500000000000001, min_samples_leaf=5, min_samples_split=8, n_estimators=100)) </pre>
pima	78.4%	82.5%	<pre> make_pipeline(make_union(FunctionTransformer(copy), FunctionTransformer(copy)), LogisticRegression(C=1.0, dual=False, penalty="l1")) </pre>
poker	63.97%	63.5%	<pre> make_pipeline(MaxAbsScaler(), ExtraTreesClassifier(bootstrap=True, criterion="entropy", max_features=0.2, min_samples_leaf=8, min_samples_split=19, n_estimators=100)) </pre>
postoperative-patient-data	75.4%	75.6%	<pre> LinearSVC(C=0.0001, dual=False, loss="squared_hinge", penalty="l1", tol=0.001) </pre>
prnn_crabs	100%	95.7%	<pre> LogisticRegression(C=1.0, dual=False, penalty="l1") </pre>

数据集	Glaucus	500-tree RF	Pipeline
prnn_fglass	80.5%	80.3%	<pre> make_pipeline(StackingEstimator(estimator=DecisionTreeClassifier(criteri on="gini", max_depth=9, min_samples_leaf=19, min_samples_split=11)), RandomForestClassifier(bootstrap=True, criterion="gini", max_features=0.9000000000000001, min_samples_leaf=1, min_samples_split=9, n_estimators=100)) </pre>
prnn_synth	87.4%	87.7%	<pre> KNeighborsClassifier(n_neighbors=35, p=2, weights="distance") </pre>
profb	68.5%	68.6%	<pre> RandomForestClassifier(bootstrap=False, criterion="gini", max_features=0.45, min_samples_leaf=19, min_samples_split=16, n_estimators=100) </pre>
promoters	98.8%	98.7%	<pre> RandomForestClassifier(bootstrap=False, criterion="entropy", max_features=0.55, min_samples_leaf=2, min_samples_split=3, n_estimators=100) </pre>
ring	98.1%	97.8%	<pre> make_pipeline(FastICA(tol=0.8), StackingEstimator(estimator=LinearSVC(C=0.1, dual=False, loss="squared_hinge", penalty="l1", tol=0.1)), GaussianNB()) </pre>
saheart	73.7%	73.7%	<pre> RandomForestClassifier(bootstrap=False, criterion="gini", max_features=0.25, min_samples_leaf=10, min_samples_split=20, n_estimators=100) </pre>
satimage	89.6%	89.8%	<pre> GradientBoostingClassifier(learning_rate=0.01, max_depth=9, max_features=0.8500000000000001, min_samples_leaf=5, min_samples_split=13, n_estimators=100, subsample=0.9500000000000001) </pre>
schizo	71.3%	71.1%	<pre> make_pipeline(StackingEstimator(estimator=DecisionTreeClassifier(criteri on="entropy", max_depth=7, min_samples_leaf=6, min_samples_split=20)), DecisionTreeClassifier(criterion="entropy", max_depth=7, min_samples_leaf=18, min_samples_split=9)) </pre>

数据集	Glaucus	500-tree RF	Pipeline
segmentation	96.0%	96.5%	<pre> make_pipeline(StackingEstimator(estimator=ExtraTreesClassifier(bootstrap=False, criterion="entropy", max_features=0.5, min_samples_leaf=18, min_samples_split=8, n_estimators=100)), RandomForestClassifier(bootstrap=False, criterion="entropy", max_features=0.55, min_samples_leaf=12, min_samples_split=12, n_estimators=100)) </pre>
shuttle	99.98%	99.91%	<pre> RandomForestClassifier(bootstrap=False, criterion="gini", max_features=0.55, min_samples_leaf=4, min_samples_split=6, n_estimators=100) </pre>
sleep	76.9%	76.4%	<pre> make_pipeline(PolynomialFeatures(degree=2, include_bias=False, interaction_only=False), ExtraTreesClassifier(bootstrap=False, criterion="entropy", max_features=0.25, min_samples_leaf=5, min_samples_split=4, n_estimators=100)) </pre>
solar-flare_1	73.7%	73.3%	<pre> RandomForestClassifier(bootstrap=True, criterion="entropy", max_features=0.5, min_samples_leaf=16, min_samples_split=3, n_estimators=100) </pre>
solar-flare_2	73.7%	73.5%	<pre> make_pipeline(VarianceThreshold(threshold=0.45), StackingEstimator(estimator=ExtraTreesClassifier(bootstrap=False, criterion="entropy", max_features=0.15000000000000002, min_samples_leaf=1, min_samples_split=19, n_estimators=100)), GradientBoostingClassifier(learning_rate=0.001, max_depth=8, max_features=0.1, min_samples_leaf=2, min_samples_split=12, n_estimators=100, subsample=0.7500000000000001)) </pre>
sonar	80.0%	79.9%	<pre> GradientBoostingClassifier(learning_rate=1.0, max_depth=6, max_features=0.3, min_samples_leaf=14, min_samples_split=15, n_estimators=100, subsample=0.8) </pre>
soybean	91.4%	91.7%	<pre> RandomForestClassifier(bootstrap=True, criterion="entropy", max_features=0.8500000000000001, min_samples_leaf=1, min_samples_split=11, n_estimators=100) </pre>

数据集	Glaucus	500-tree RF	Pipeline
spambase	95.2%	95.2%	<pre> make_pipeline(MinMaxScaler(), RandomForestClassifier(bootstrap=True, criterion="gini", max_features=0.15000000000000002, min_samples_leaf=1, min_samples_split=4, n_estimators=100)) </pre>
spect	83.3%	83.3%	<pre> make_pipeline(PCA(iterated_power=4, svd_solver="randomized"), RandomForestClassifier(bootstrap=False, criterion="gini", max_features=0.15000000000000002, min_samples_leaf=15, min_samples_split=18, n_estimators=100)) </pre>
spectf	86.5%	86.1%	<pre> make_pipeline(StackingEstimator(estimator=LinearSVC(C=0.0001, dual=True, loss="squared_hinge", penalty="l2", tol=0.001)), GradientBoostingClassifier(learning_rate=0.1, max_depth=4, max_features=0.3, min_samples_leaf=17, min_samples_split=4, n_estimators=100, subsample=0.90000000000000001)) </pre>
splice	94.9%	74.5%	<pre> ExtraTreesClassifier(bootstrap=False, criterion="gini", max_features=0.55, min_samples_leaf=7, min_samples_split=6, n_estimators=100) </pre>
tae	54.7%	55.1%	<pre> DecisionTreeClassifier(criterion="entropy", max_depth=4, min_samples_leaf=4, min_samples_split=15) </pre>
texture	99.8%	99.5%	<pre> make_pipeline(PolynomialFeatures(degree=2, include_bias=False, interaction_only=False), LogisticRegression(C=25.0, dual=False, penalty="l2")) </pre>
threeOf9	100%	99.7%	<pre> make_pipeline(StackingEstimator(estimator=GradientBoostingClassifier(learning_rate=0.1, max_depth=3, max_features=0.85000000000000001, min_samples_leaf=20, min_samples_split=18, n_estimators=100, subsample=0.90000000000000001)), LogisticRegression(C=15.0, dual=False, penalty="l1")) </pre>
tic-tac-toe	98.4%	94.7%	<pre> make_pipeline(ZeroCount(), ExtraTreesClassifier(bootstrap=True, criterion="entropy", max_features=0.90000000000000001, min_samples_leaf=5, min_samples_split=14, n_estimators=100)) </pre>

数据集	Glaucus	500-tree RF	Pipeline
titanic	79%	79.4%	make_pipeline(FastICA(tol=0.5), KNeighborsClassifier(n_neighbors=15, p=2, weights="uniform"))
tokyo1	93.1%	92.8%	make_pipeline(StackingEstimator(estimator=ExtraTreesClassifier(bootstrap=False, criterion="gini", max_features=0.35000000000000003, min_samples_leaf=4, min_samples_split=16, n_estimators=100)), StandardScaler(), GradientBoostingClassifier(learning_rate=0.5, max_depth=7, max_features=0.5, min_samples_leaf=7, min_samples_split=5, n_estimators=100, subsample=0.8))
twonorm	97.6%	97.3%	make_pipeline(PCA(iterated_power=4, svd_solver="randomized"), GradientBoostingClassifier(learning_rate=0.1, max_depth=6, max_features=0.4, min_samples_leaf=10, min_samples_split=10, n_estimators=100, subsample=1.0))
vehicle	80.2%	79.7%	make_pipeline(StackingEstimator(estimator=DecisionTreeClassifier(criterion="gini", max_depth=2, min_samples_leaf=18, min_samples_split=8)), LogisticRegression(C=0.5, dual=False, penalty="l2"))
vote	96.4%	96.8%	LogisticRegression(C=10.0, dual=False, penalty="l1")
vowel	91.4%	90.9%	make_pipeline(StackingEstimator(estimator=GradientBoostingClassifier(learning_rate=0.1, max_depth=7, max_features=0.35000000000000003, min_samples_leaf=4, min_samples_split=12, n_estimators=100, subsample=0.7000000000000001)), ExtraTreesClassifier(bootstrap=True, criterion="gini", max_features=0.7000000000000001, min_samples_leaf=7, min_samples_split=8, n_estimators=100))
waveform-21	87.0%	87.3%	LogisticRegression(C=1.0, dual=False, penalty="l1")
waveform-40	86.4%	86.7%	make_pipeline(SelectFwe(score_func=f_classif, alpha=0.045), KNeighborsClassifier(n_neighbors=100, p=2, weights="distance"))

数据集	Glaucus	500-tree RF	Pipeline
wdbc	96.0%	95.9%	make_pipeline(PolynomialFeatures(degree=2, include_bias=False, interaction_only=False), RandomForestClassifier(bootstrap=False, criterion="gini", max_features=0.35000000000000003, min_samples_leaf=10, min_samples_split=7, n_estimators=100))
wine-quality-red	65.2%	65.4%	make_pipeline(ZeroCount(), RandomForestClassifier(bootstrap=True, criterion="entropy", max_features=0.05, min_samples_leaf=1, min_samples_split=6, n_estimators=100))
wine-quality-white	63.5%	63.7%	GradientBoostingClassifier(learning_rate=0.1, max_depth=7, max_features=0.7000000000000001, min_samples_leaf=5, min_samples_split=12, n_estimators=100, subsample=0.9500000000000001)
wine-recognition	97.6%	99.7%	ExtraTreesClassifier(bootstrap=True, criterion="gini", max_features=0.9500000000000001, min_samples_leaf=13, min_samples_split=16, n_estimators=100)
xd6	100%	100%	GradientBoostingClassifier(learning_rate=0.5, max_depth=9, max_features=1.0, min_samples_leaf=8, min_samples_split=8, n_estimators=100, subsample=0.6500000000000001)
yeast	61.9%	61.9%	RandomForestClassifier(bootstrap=False, criterion="entropy", max_features=0.4, min_samples_leaf=11, min_samples_split=9, n_estimators=100)

*绿色表示优于500颗子树的随机森林算法，红色表示弱，黑色表示无太大差别（0.5%以内）

B. 大体折线图趋势

