STREAMLINE Training Summary Report: 2022-06-15 06:16:43.548609

General Pipeline Settings:

Data Path: /home/ryanurb/idata/datasets/GAMETES_set
Output Path: /home/ryanurb/idata/output
Experiment Name: GAMETES_Set
Class Label: Class
Instance Label: None
Ignored Features: None
Specified Categorical Features: None
CV Partitions: 10
Partition Method: S
Match Label: None
Categorical Cutoff: 10
Statistical Significance Cutoff: 0.05
Export Feature Correlations: True
Export Univariate Plots: False
Random Seed: 42
Run From Jupyter Notebook: False
Use Data Scaling: True
Use Data Imputation: True
Use Multivariate Imputation: True
Use Mutual Information: True
Use MultiSURF: True
Use TURF: False
TURF Cutoff: 0.5
MultiSURF Instance Subset: 2000
Max Features to Keep: 2000
Filter Poor Features: True
Top Features to Display: 40
Export Feature Importance Plot: True
Overwrite CV Datasets: True
Primary Metric: balanced_accuracy
Training Subsample for KNN,ANN,SVM,and XGB: 0
Uniform Feature Importance Estimation (Models): True
Hyperparameter Sweep Number of Trials: 200
Hyperparameter Timeout: 900
Export Hyperparameter Sweep Plots: True
Export ROC Plot: True
Export PRC Plot: True
Export Metric Boxplots: True
Export Feature Importance Boxplots: True
Metric Weighting Composite FI Plots: balanced_accuracy
Top Model Features To Display: 40

ML Modeling Algorithms:

Naive Bayes: True
Logistic Regression: True
Decision Tree: True
Random Forest: True
Gradient Boosting: True
Extreme Gradient Boosting: True
Light Gradient Boosting: True
Category Gradient Boosting: True
Support Vector Machine: True
Artificial Neural Network: True
K-Nearest Neightbors: True
Genetic Programming: True
eLCS: False
XCS: False
ExSTraCS: True

LCS Settings (eLCS,XCS,ExSTraCS):

Do LCS Hyperparameter Sweep: False
nu: 1
Training Iterations: 200000
N (Rule Population Size): 2000
LCS Hyperparameter Sweep Timeout: 1200

Datasets:

D1 = A_gametes_univariate_L_1_H_0
D2 = B_gametes_uni_4add_L_1_H_0
D3 = C_gametes_uni_4het_L_1_H_0
D4 = D_gametes_2way_epistasis_L_2_H_0
D5 = E_gametes_2way_epi_2het_L_2_H_0
D6 = F_gametes_3way_epistasis_L_3_H_0

Univariate Analysis of Each Dataset (Top 10 Features for Each): Page 1

D1 = A_gametes_univariate_L_1_H_0

Feature: P-Value
M0P0: 3.1354809610620264e-153
N93: 0.0018155781298386
N85: 0.0185080636550741
N67: 0.0598127418382017
N18: 0.0680671826646687
N11: 0.0760873954359647
N66: 0.0870112019874774
N59: 0.0913493991212986
N34: 0.0921194966839936
N15: 0.1073435732820547

D2 = B_gametes_uni_4add_L_1_H_0

Feature: P-Value
M0P0: 1.11064715131141122e-164
M2P0: 1.18667349514186658e-157
M1P0: 2.2401547020040568e-154
M3P0: 5.040770678366382e-135
N20: 0.0259070859943441
N37: 0.0474800954604112
N5: 0.0533088603442028
N87: 0.0576055460913057
N43: 0.059260439873388
N10: 0.0651911642607531

D3 = C_gametes_uni_4het_L_1_H_0

Feature: P-Value
M2P0: 6.426090536165597e-14
M1P0: 3.583636293759726e-10
M0P0: 5.343816294212595e-09
M3P0: 8.690076282752275e-08
N86: 0.0143947288220938
N60: 0.0270282262454729
N32: 0.0292941753656844
N30: 0.084886706454954
N42: 0.0887611657002169
N81: 0.0986478983120729

D4 = D_gametes_2way_epistasis_L_2_H_0

Feature: P-Value
N27: 0.017443903516463
N60: 0.030238939964301
N9: 0.0363104970744642
N49: 0.045974643967979
N39: 0.0570573723322331
N51: 0.0958228904576739
N94: 0.1238904670297999
N41: 0.1276792520324695
N10: 0.1302368572103056
N13: 0.1326730450150889

D5 = E_gametes_2way_epi_2het_L_2_H_0

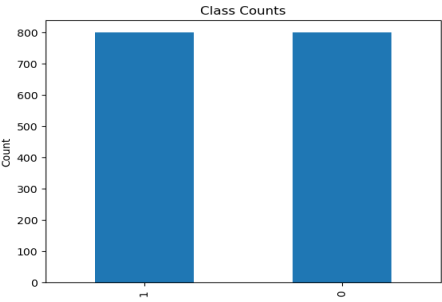| Feature: P-Value |
| --- |
| N8: 0.0022378710091608 |
| N10: 0.0033282354218843 |
| N1: 0.023956678010501 |
| N90: 0.0409688621078246 |
| N16: 0.0475567352335571 |
| N36: 0.0483697186665984 |
| N45: 0.05076830219 |
| N49: 0.0536871730192233 |
| N9: 0.0598766172399342 |
| N61: 0.076713380936202 |

Univariate Analysis of Each Dataset (Top 10 Features for Each): Page 2

D6 = F_gametes_3way_epistasis_L_3_H_0

Feature: P-Value

N32: 0.0603455725673375
N81: 0.0701416847012984
N79: 0.0741037719317539
N41: 0.076331904851009
N8: 0.0862935864993705
N29: 0.1141351219582369
N6: 0.1445612533667589
N77: 0.1449669067148887
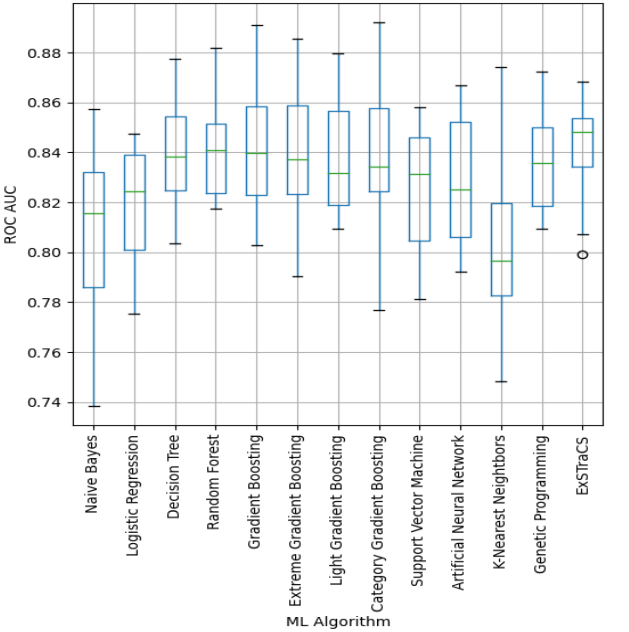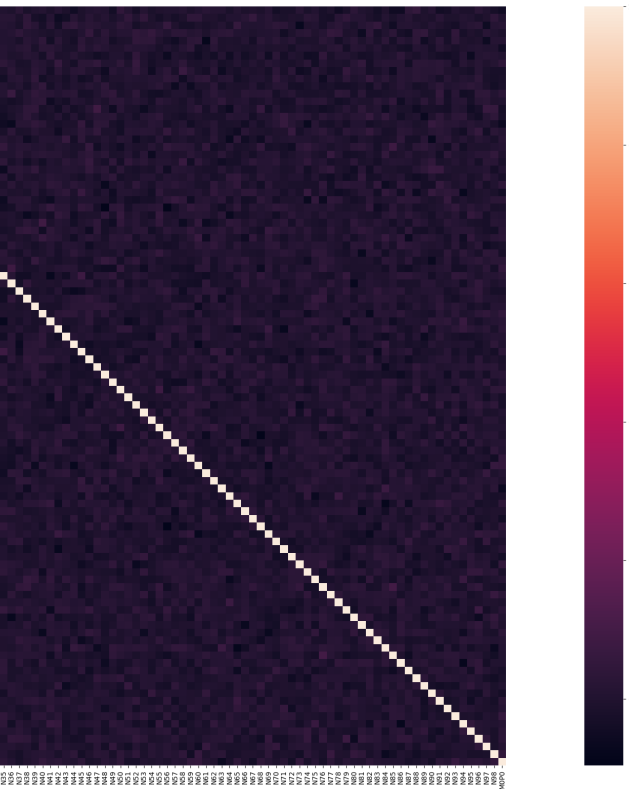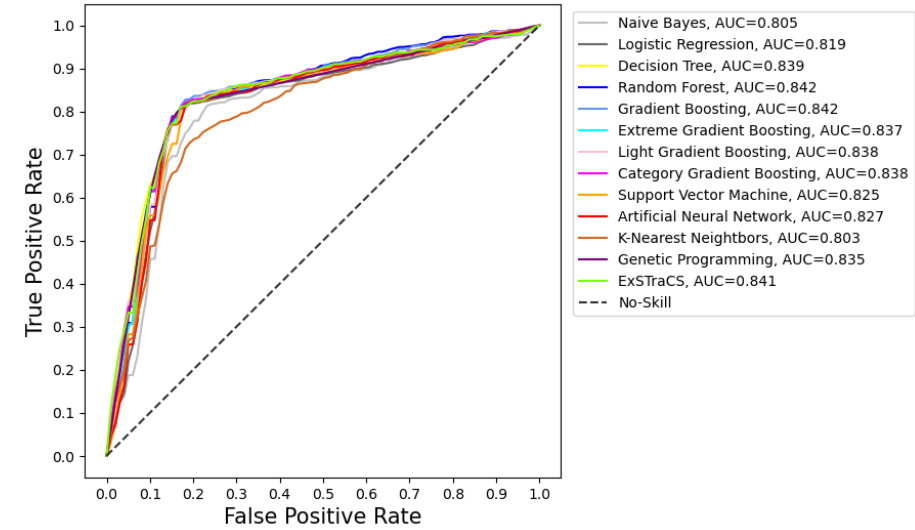N23: 0.1526194924016151
N91: 0.1567681464069809

Dataset and Model Prediction Summary: D1 = A_gametes_univariate_L_1_H_0

Class Counts

Dataset Counts Summary:

instances: 1600.0

features: 100.0

categorical_features: 100.0

quantitative_features: 0.0

missing_values: 0.0

missing_percent: 0.0

Top ML Algorithm Results (Averaged Over CV Runs):

Best (ROC_AUC): Random Forest = 0.842

Best (Balanced Acc.): ExSTraCS = 0.830

Best (F1 Score): ExSTraCS = 0.827

Best (PRC AUC): Genetic Programming = 0.875

Best (PRC APS): ExSTraCS = 0.833

ROC

Naive Bayes, AUC=0.805

Logistic Regression, AUC=0.819

Decision Tree, AUC=0.839

Random Forest, AUC=0.842

Gradient Boosting, AUC=0.842

Extreme Gradient Boosting, AUC=0.837

Light Gradient Boosting, AUC=0.838

Category Gradient Boosting, AUC=0.838

Support Vector Machine, AUC=0.825

Artificial Neural Network, AUC=0.827

K-Nearest Neightbors, AUC=0.803

Genetic Programming, AUC=0.835

ExSTraCS, AUC=0.841

No-Skill

PRC

Naive Bayes, AUC=0.765, APS=0.770

Logistic Regression, AUC=0.763, APS=0.782

Decision Tree, AUC=0.831, APS=0.818

Random Forest, AUC=0.811, APS=0.815

Gradient Boosting, AUC=0.812, APS=0.815

Extreme Gradient Boosting, AUC=0.823, APS=0.810

Light Gradient Boosting, AUC=0.808, APS=0.812

Category Gradient Boosting, AUC=0.816, APS=0.820

Support Vector Machine, AUC=0.787, APS=0.791

Artificial Neural Network, AUC=0.784, APS=0.789

K-Nearest Neightbors, AUC=0.771, APS=0.775

Genetic Programming, AUC=0.875, APS=0.789

ExSTraCS, AUC=0.830, APS=0.833

No-Skill

Generated with the URBS-Lab STREAMLINE: (https://github.com/UrbsLab/STREAMLINE)

Dataset and Model Prediction Summary: D2 = B_gametes_uni_4add_L_1_H_0

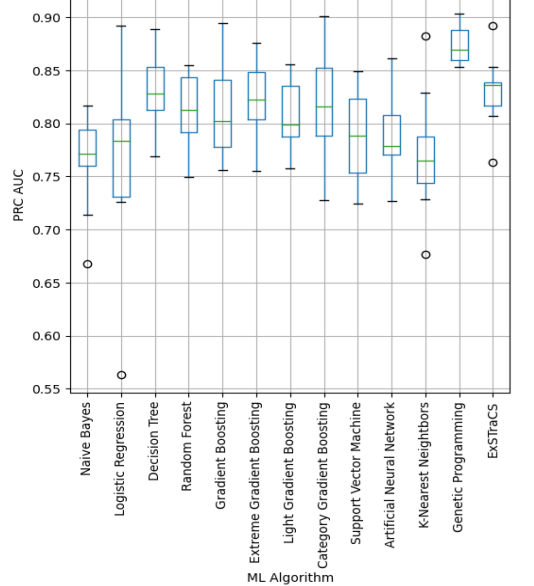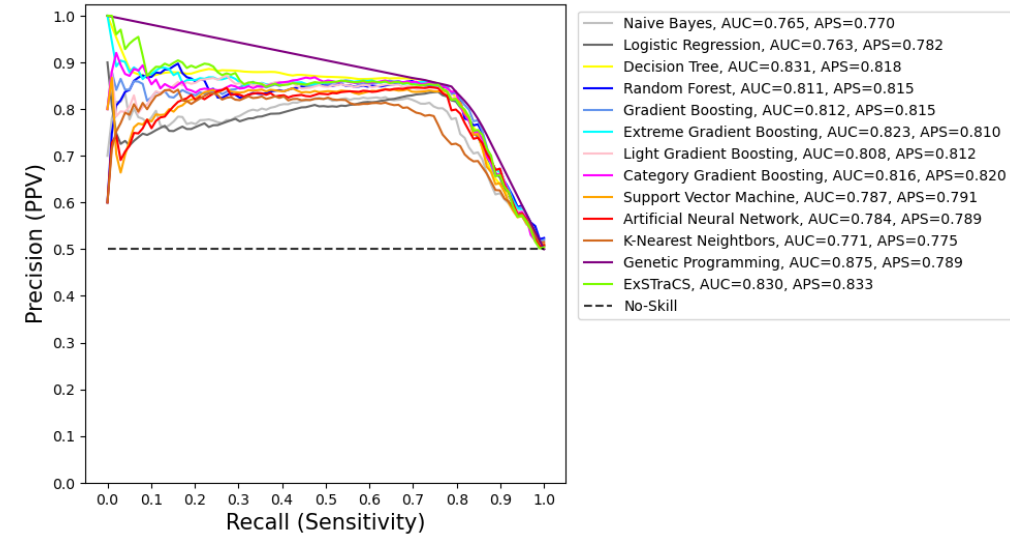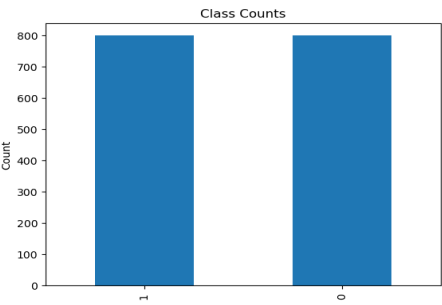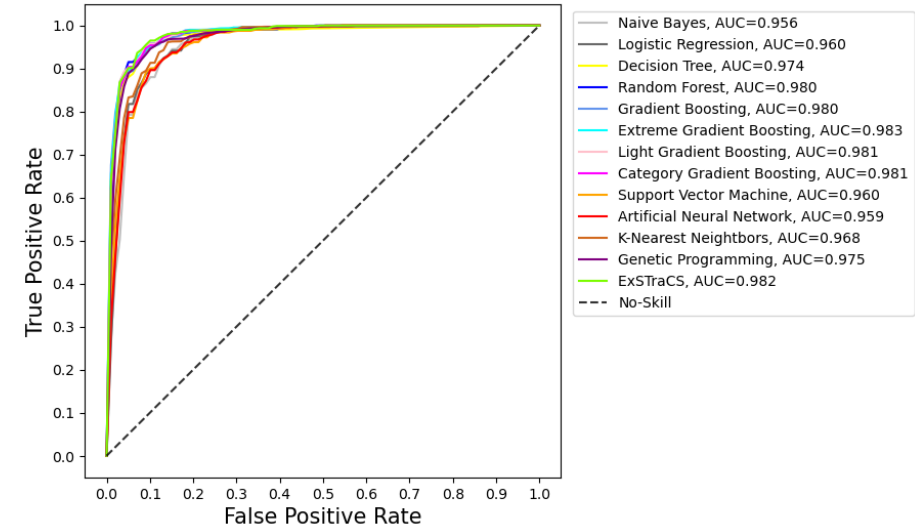### Class Counts

Dataset Counts Summary:

instances: 1600.0

features: 100.0

categorical_features: 100.0

quantitative_features: 0.0

missing_values: 0.0

missing_percent: 0.0

Top ML Algorithm Results (Averaged Over CV Runs):

Best (ROC_AUC): Extreme Gradient Boosting = 0.983

Best (Balanced Acc.): Extreme Gradient Boosting = 0.926

Best (F1 Score): Extreme Gradient Boosting = 0.926

Best (PRC AUC): Extreme Gradient Boosting = 0.982

Best (PRC APS): Extreme Gradient Boosting = 0.982

ROC

- Naive Bayes, AUC=0.956
- Logistic Regression, AUC=0.960
- Decision Tree, AUC=0.974
- Random Forest, AUC=0.980
- Gradient Boosting, AUC=0.980
- Extreme Gradient Boosting, AUC=0.983
- Light Gradient Boosting, AUC=0.981
- Category Gradient Boosting, AUC=0.981
- Support Vector Machine, AUC=0.960
- Artificial Neural Network, AUC=0.959
- K-Nearest Neightbors, AUC=0.968
- Genetic Programming, AUC=0.975
- ExSTraCS, AUC=0.982
- No-Skill

PRC

- Naive Bayes, AUC=0.939, APS=0.940
- Logistic Regression, AUC=0.943, APS=0.944
- Decision Tree, AUC=0.976, APS=0.967
- Random Forest, AUC=0.977, APS=0.977
- Gradient Boosting, AUC=0.981, APS=0.978
- Extreme Gradient Boosting, AUC=0.982, APS=0.982
- Light Gradient Boosting, AUC=0.980, APS=0.980
- Category Gradient Boosting, AUC=0.980, APS=0.980
- Support Vector Machine, AUC=0.949, APS=0.950
- Artificial Neural Network, AUC=0.944, APS=0.945
- K-Nearest Neightbors, AUC=0.962, APS=0.962
- Genetic Programming, AUC=0.967, APS=0.972
- ExSTraCS, AUC=0.981, APS=0.981
- No-Skill

Generated with the URBS-Lab STREAMLINE: (https://github.com/UrbsLab/STREAMLINE)

Dataset and Model Prediction Summary: D3 = C_gametes_uni_4het_L_1_H_0
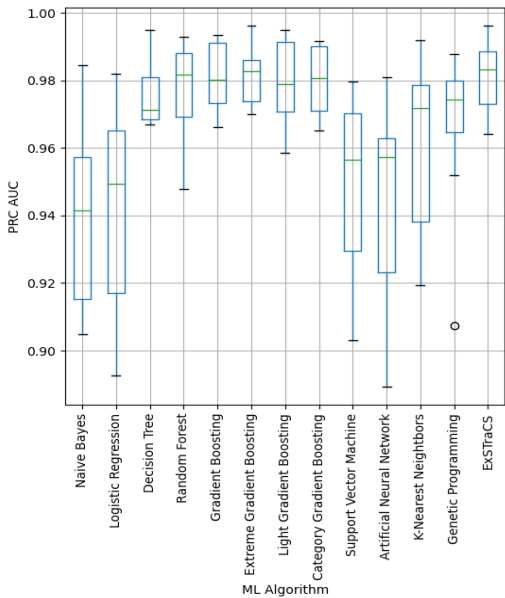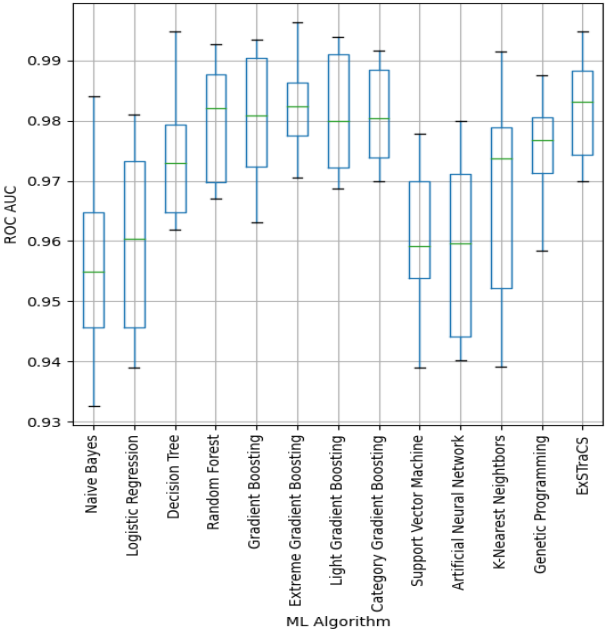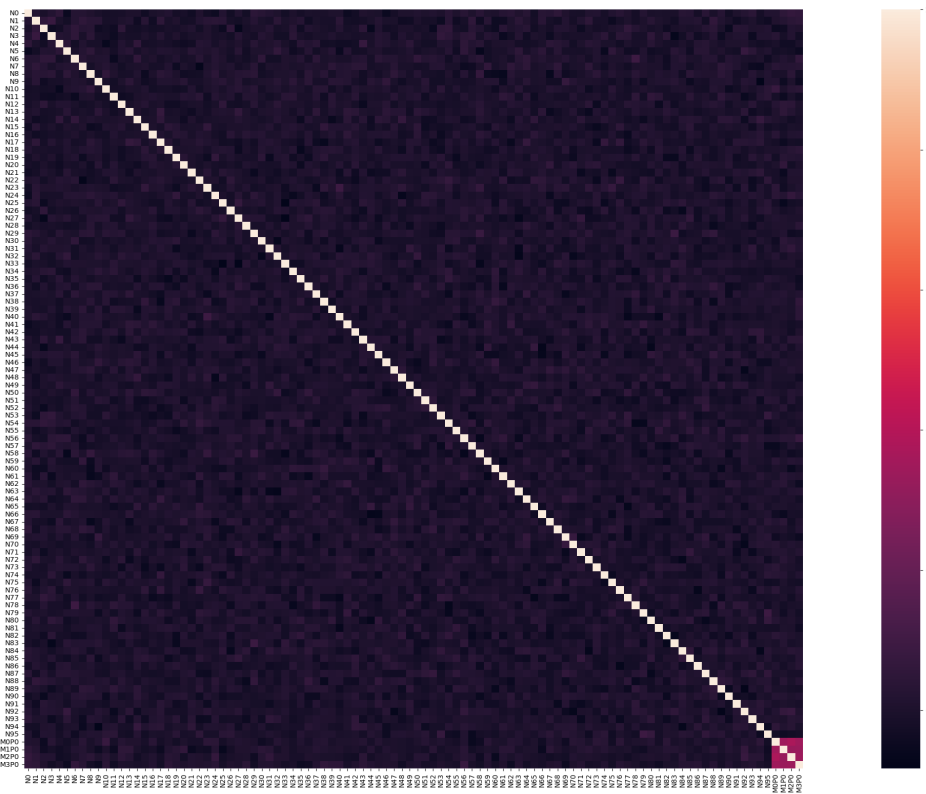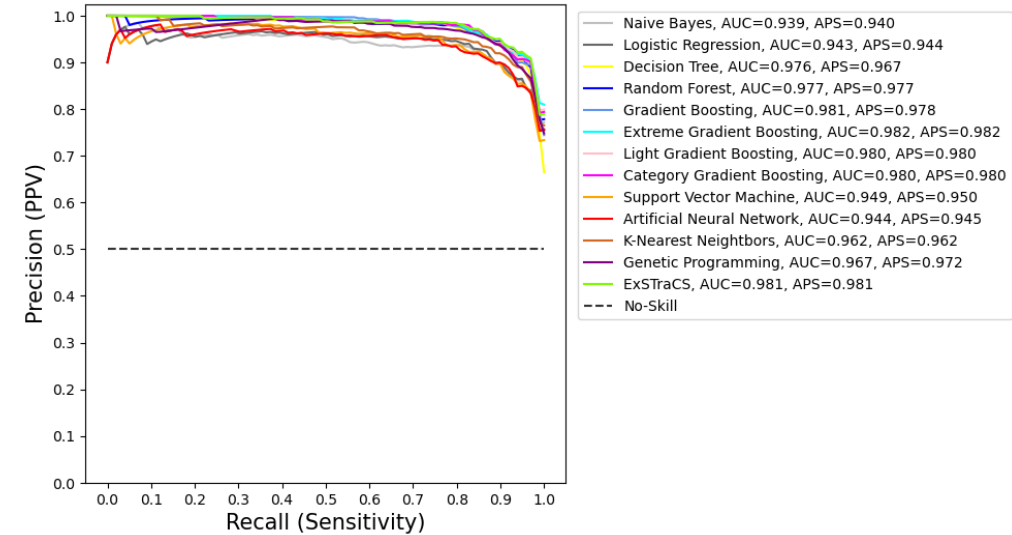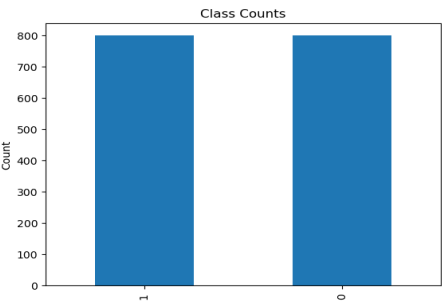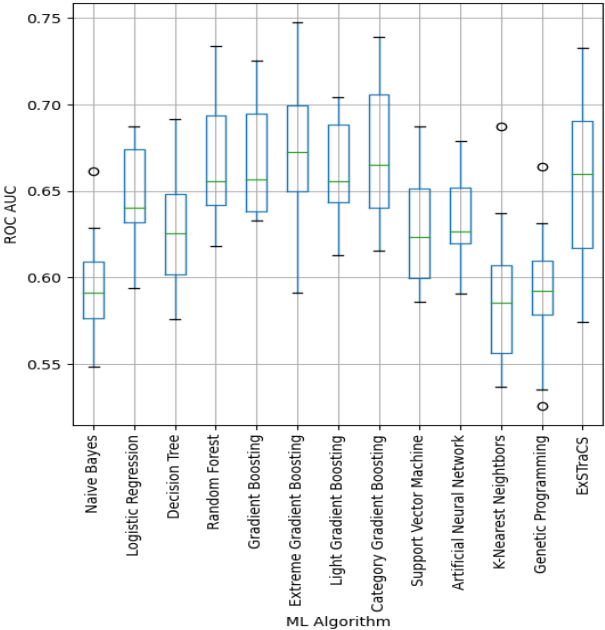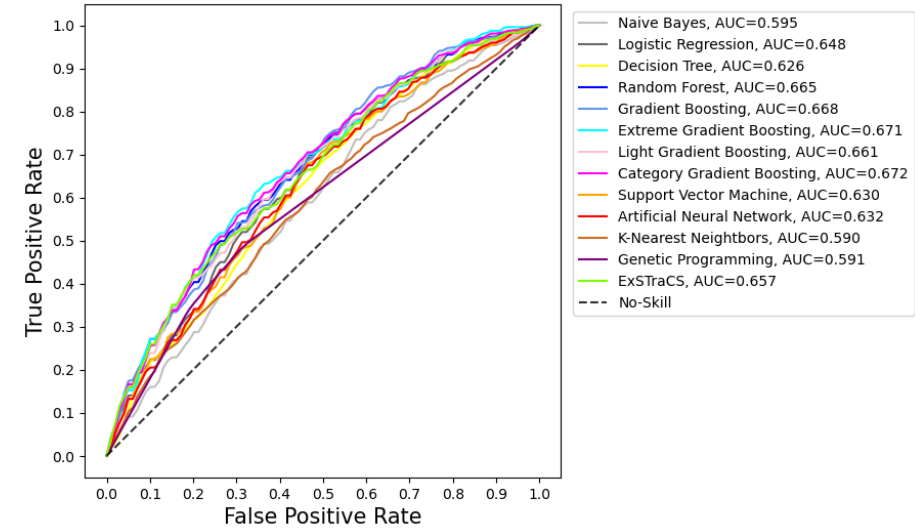
Class Counts

Dataset Counts Summary:

instances: 1600.0

features: 100.0

categorical_features: 100.0

quantitative_features: 0.0

missing_values: 0.0

missing_percent: 0.0

Top ML Algorithm Results (Averaged Over CV Runs):

Best (ROC_AUC): Category Gradient Boosting = 0.672

Best (Balanced Acc.): Category Gradient Boosting = 0.629

Best (F1 Score): Category Gradient Boosting = 0.618

Best (PRC AUC): Genetic Programming = 0.674

Best (PRC APS): Category Gradient Boosting = 0.659

ROC

Naive Bayes, AUC=0.595
Logistic Regression, AUC=0.648
Decision Tree, AUC=0.626
Random Forest, AUC=0.665
Gradient Boosting, AUC=0.668
Extreme Gradient Boosting, AUC=0.671
Light Gradient Boosting, AUC=0.661
Category Gradient Boosting, AUC=0.672
Support Vector Machine, AUC=0.630
Artificial Neural Network, AUC=0.632
K-Nearest Neightbors, AUC=0.590
Genetic Programming, AUC=0.591
ExSTraCS, AUC=0.657
No-Skill

PRC

Naive Bayes, AUC=0.578, APS=0.584
Logistic Regression, AUC=0.629, APS=0.634
Decision Tree, AUC=0.605, APS=0.601
Random Forest, AUC=0.643, APS=0.649
Gradient Boosting, AUC=0.652, APS=0.657
Extreme Gradient Boosting, AUC=0.649, APS=0.654
Light Gradient Boosting, AUC=0.645, APS=0.651
Category Gradient Boosting, AUC=0.654, APS=0.659
Support Vector Machine, AUC=0.615, APS=0.621
Artificial Neural Network, AUC=0.611, APS=0.617
K-Nearest Neightbors, AUC=0.582, APS=0.581
Genetic Programming, AUC=0.674, APS=0.562
ExSTraCS, AUC=0.645, APS=0.651
No-Skill

Dataset and Model Prediction Summary: D4 = D_gametes_2way_epistasis_L_2_H_0
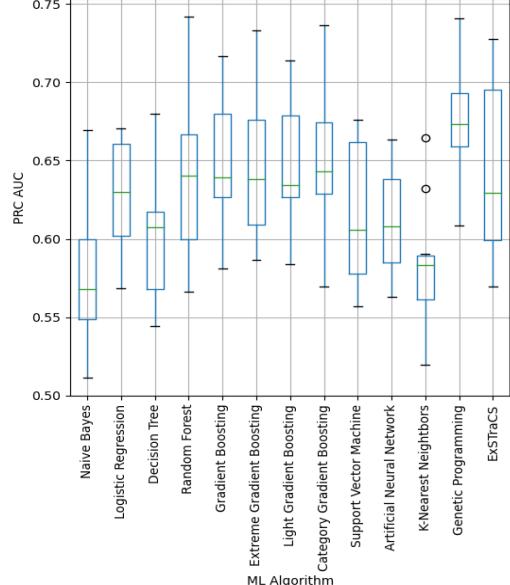
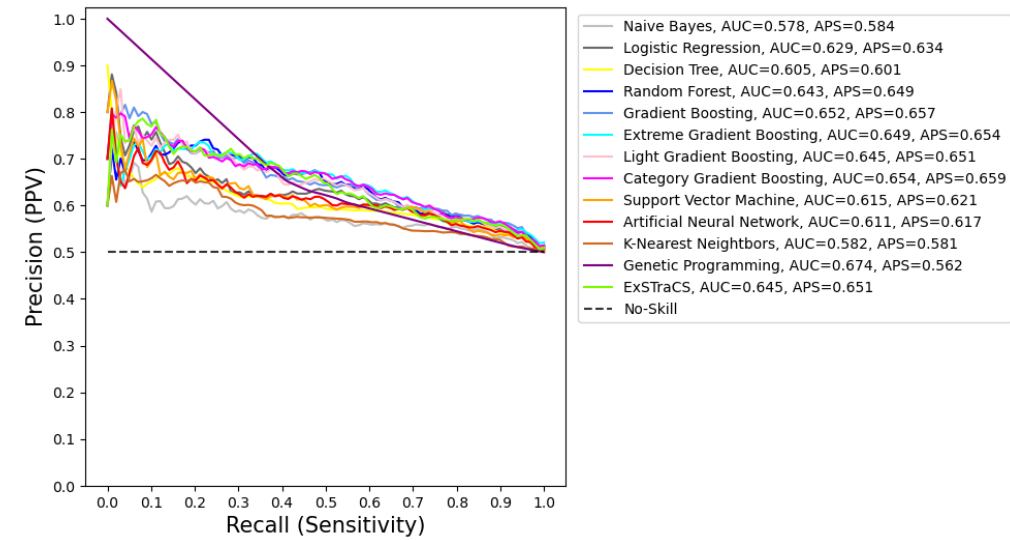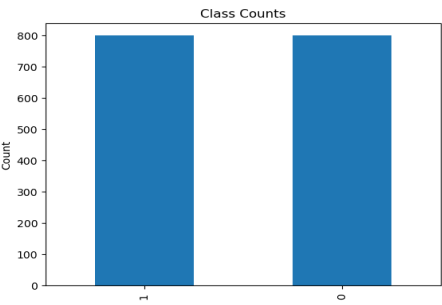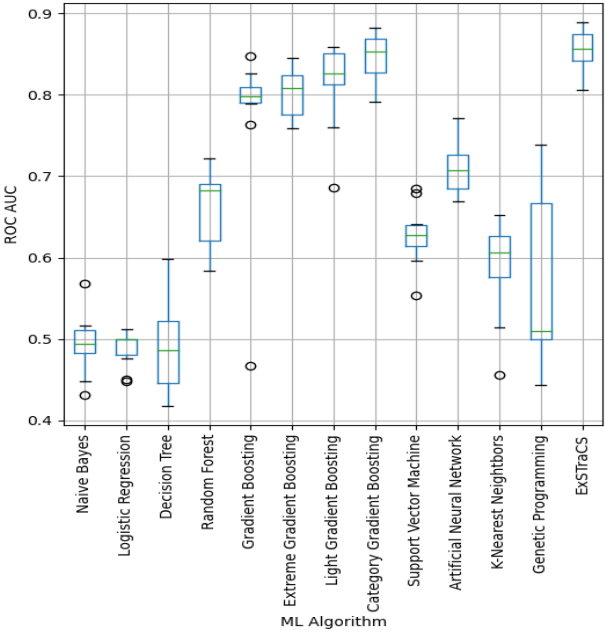**Class Counts**

Dataset Counts Summary:

instances: 1600.0

features: 100.0

categorical_features: 100.0

quantitative_features: 0.0

missing_values: 0.0

missing_percent: 0.0

Top ML Algorithm Results (Averaged Over CV Runs):

Best (ROC_AUC): ExSTraCS = 0.854
Best (Balanced Acc.): ExSTraCS = 0.802
Best (F1 Score): ExSTraCS = 0.825
Best (PRC AUC): ExSTraCS = 0.818
Best (PRC APS): ExSTraCS = 0.820

ROC

Naive Bayes, AUC=0.494
Logistic Regression, AUC=0.488
Decision Tree, AUC=0.489
Random Forest, AUC=0.658
Gradient Boosting, AUC=0.770
Extreme Gradient Boosting, AUC=0.802
Light Gradient Boosting, AUC=0.814
Category Gradient Boosting, AUC=0.845
Support Vector Machine, AUC=0.627
Artificial Neural Network, AUC=0.709
K-Nearest Neightbors, AUC=0.588
Genetic Programming, AUC=0.568
ExSTraCS, AUC=0.854
No-Skill

PRC

Naive Bayes, AUC=0.492, APS=0.500
Logistic Regression, AUC=0.593, APS=0.499
Decision Tree, AUC=0.512, APS=0.505
Random Forest, AUC=0.641, APS=0.646
Gradient Boosting, AUC=0.725, APS=0.730
Extreme Gradient Boosting, AUC=0.751, APS=0.756
Light Gradient Boosting, AUC=0.772, APS=0.776
Category Gradient Boosting, AUC=0.802, APS=0.806
Support Vector Machine, AUC=0.617, APS=0.623
Artificial Neural Network, AUC=0.684, APS=0.690
K-Nearest Neightbors, AUC=0.579, APS=0.586
Genetic Programming, AUC=0.667, APS=0.554
ExSTraCS, AUC=0.818, APS=0.820
No-Skill

Dataset and Model Prediction Summary: D5 = E_gametes_2way_epi_2het_L_2_H_0
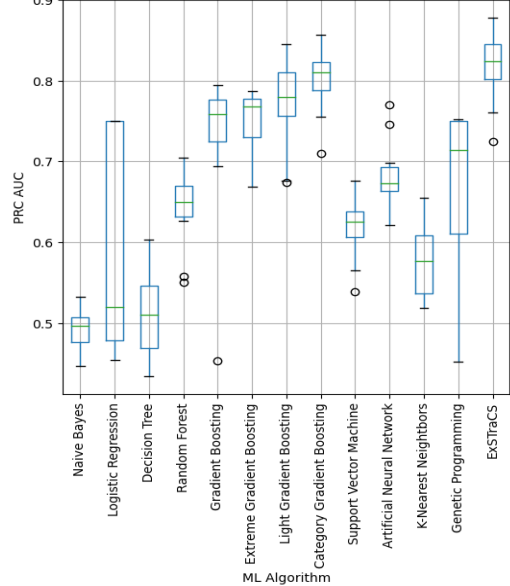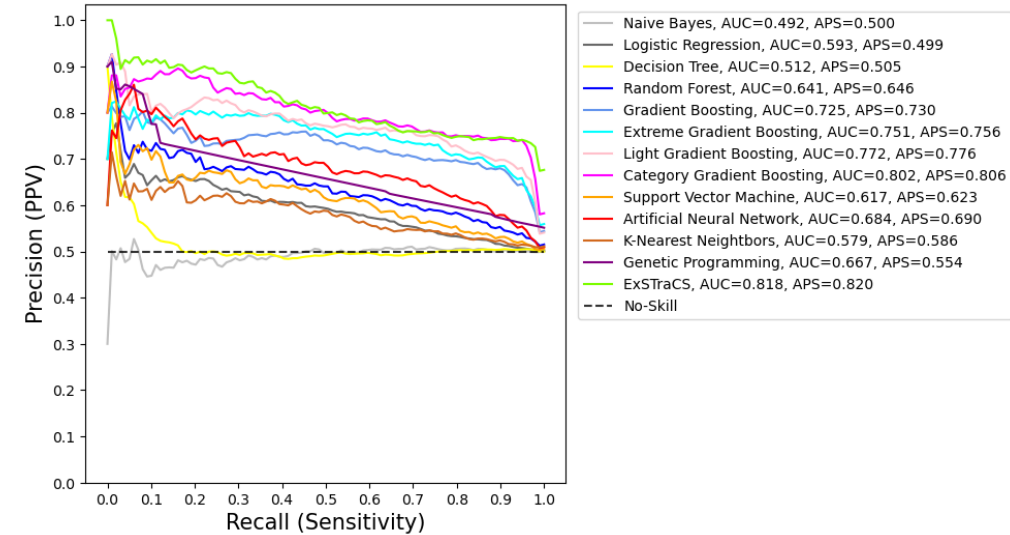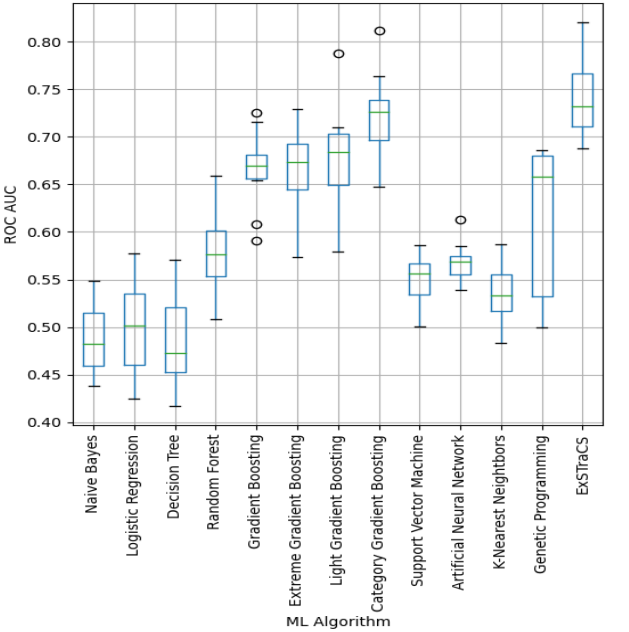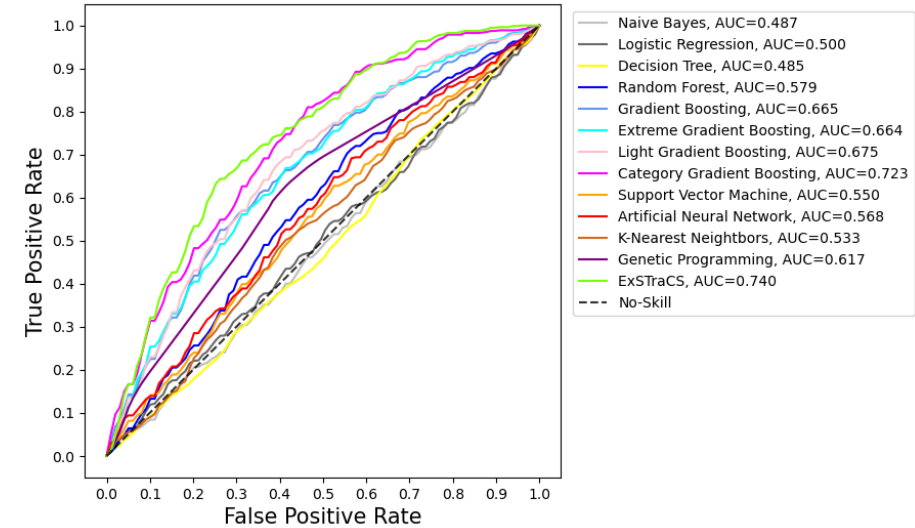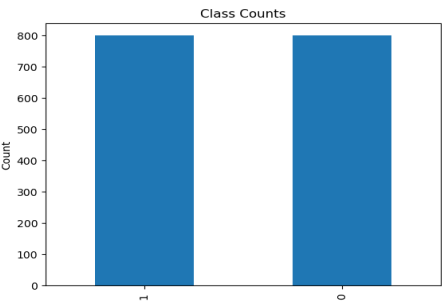
Class Counts

Dataset Counts Summary:

instances: 1600.0

features: 100.0

categorical_features: 100.0

quantitative_features: 0.0

missing_values: 0.0

missing_percent: 0.0

Top ML Algorithm Results (Averaged Over CV Runs):

Best (ROC_AUC): ExSTraCS = 0.740	
Best (Balanced Acc.): ExSTraCS = 0.688	
Best (F1 Score): ExSTraCS = 0.696	
Best (PRC AUC): Category Gradient Boosting = 0.695	
Best (PRC APS): ExSTraCS = 0.699	

ROC

- Naive Bayes, AUC=0.487
- Logistic Regression, AUC=0.500
- Decision Tree, AUC=0.485
- Random Forest, AUC=0.579
- Gradient Boosting, AUC=0.665
- Extreme Gradient Boosting, AUC=0.664
- Light Gradient Boosting, AUC=0.675
- Category Gradient Boosting, AUC=0.723
- Support Vector Machine, AUC=0.550
- Artificial Neural Network, AUC=0.568
- K-Nearest Neightbors, AUC=0.533
- Genetic Programming, AUC=0.617
- ExSTraCS, AUC=0.740
- No-Skill

PRC

- Naive Bayes, AUC=0.497, APS=0.504
- Logistic Regression, AUC=0.511, APS=0.518
- Decision Tree, AUC=0.495, APS=0.495
- Random Forest, AUC=0.550, APS=0.558
- Gradient Boosting, AUC=0.646, APS=0.651
- Extreme Gradient Boosting, AUC=0.638, APS=0.643
- Light Gradient Boosting, AUC=0.641, APS=0.648
- Category Gradient Boosting, AUC=0.695, APS=0.699
- Support Vector Machine, AUC=0.545, APS=0.552
- Artificial Neural Network, AUC=0.564, APS=0.570
- K-Nearest Neightbors, AUC=0.528, APS=0.534
- Genetic Programming, AUC=0.686, APS=0.588
- ExSTraCS, AUC=0.693, APS=0.699
- No-Skill

Generated with the URBS-Lab STREAMLINE: (https://github.com/UrbsLab/STREAMLINE)

Dataset and Model Prediction Summary: D6 = F_gametes_3way_epistasis_L_3_H_0
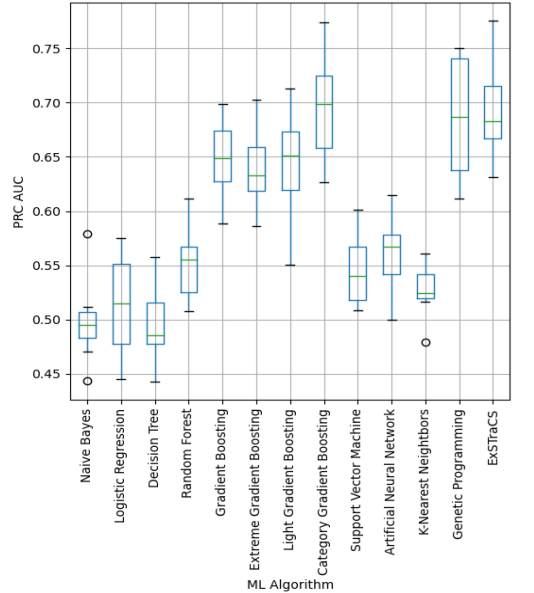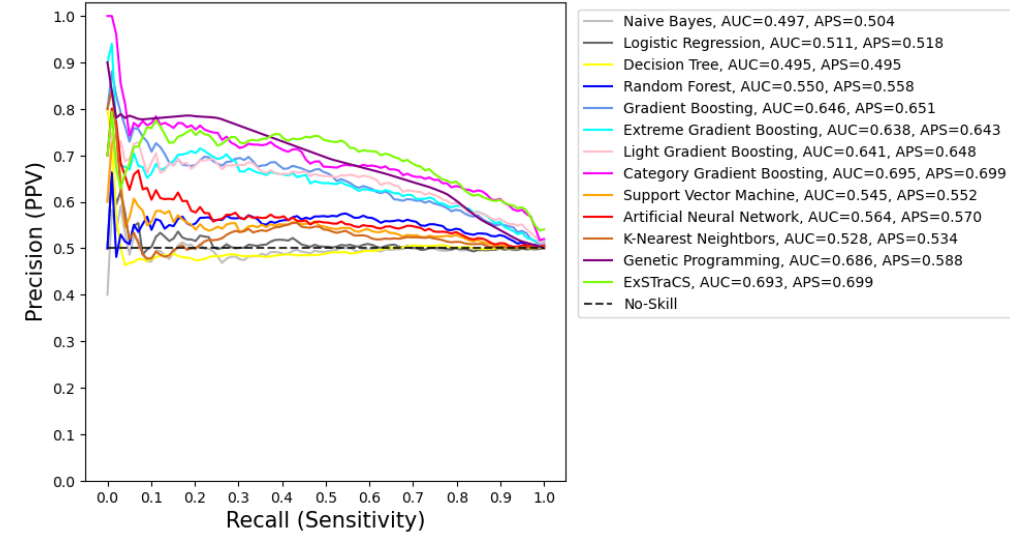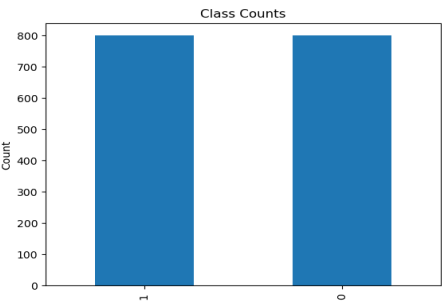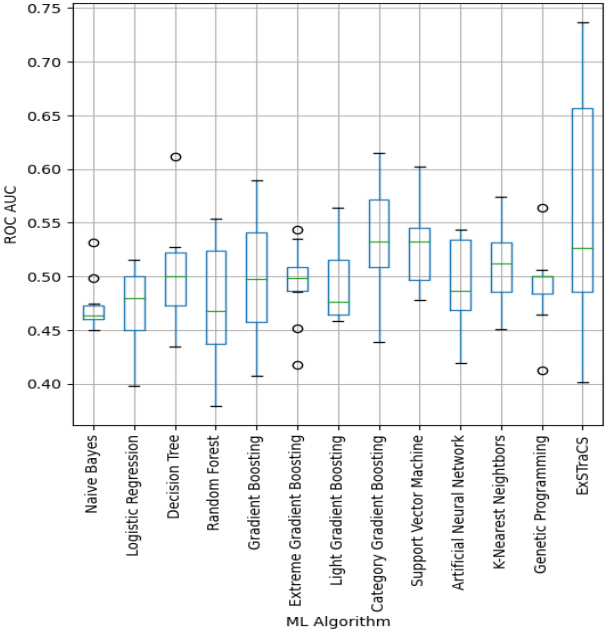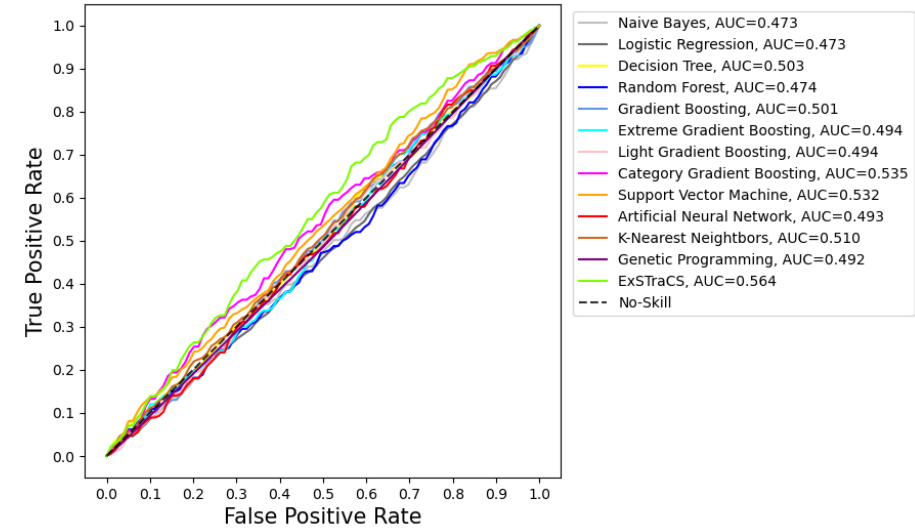
Class Counts

Dataset Counts Summary:

instances: 1600.0

features: 100.0

categorical_features: 100.0

quantitative_features: 0.0

missing_values: 0.0

missing_percent: 0.0

Top ML Algorithm Results (Averaged Over CV Runs):

Best (ROC_AUC): ExSTraCS = 0.564
Best (Balanced Acc.): ExSTraCS = 0.548
Best (F1 Score): ExSTraCS = 0.560
Best (PRC AUC): Genetic Programming = 0.604
Best (PRC APS): ExSTraCS = 0.554

ROC

Naive Bayes, AUC=0.473
Logistic Regression, AUC=0.473
Decision Tree, AUC=0.503
Random Forest, AUC=0.474
Gradient Boosting, AUC=0.501
Extreme Gradient Boosting, AUC=0.494
Light Gradient Boosting, AUC=0.494
Category Gradient Boosting, AUC=0.535
Support Vector Machine, AUC=0.532
Artificial Neural Network, AUC=0.493
K-Nearest Neightbors, AUC=0.510
Genetic Programming, AUC=0.492
ExSTraCS, AUC=0.564
No-Skill

PRC

Naive Bayes, AUC=0.488, APS=0.496
Logistic Regression, AUC=0.586, APS=0.490
Decision Tree, AUC=0.524, APS=0.515
Random Forest, AUC=0.491, APS=0.499
Gradient Boosting, AUC=0.497, APS=0.504
Extreme Gradient Boosting, AUC=0.497, APS=0.504
Light Gradient Boosting, AUC=0.489, APS=0.498
Category Gradient Boosting, AUC=0.529, APS=0.536
Support Vector Machine, AUC=0.527, APS=0.531
Artificial Neural Network, AUC=0.491, APS=0.498
K-Nearest Neightbors, AUC=0.546, APS=0.511
Genetic Programming, AUC=0.604, APS=0.497
ExSTraCS, AUC=0.548, APS=0.554
No-Skill

Generated with the URBS-Lab STREAMLINE: (https://github.com/UrbsLab/STREAMLINE)

Average Model Prediction Statistics (Rounded to 3 Decimal Points): Page 1

D1 = A_gametes_univariate_L_1_H_0

ML Algorithm	Balanced Accuracy	Accuracy	F1 Score	Sensitivity (Recall)	Specificity	Precision (PPV)	TP	TN	FP	FN	NPV	LR+	LR-	ROC AUC	PRC AUC	PRC APS
Naive Bayes	0.794	0.794	0.791	0.779	0.81	0.804	62.3	64.8	15.2	17.7	0.786	4.196	0.273	0.805	0.765	0.77
Logistic Regression	0.829	0.829	0.826	0.811	0.848	0.842	64.9	67.8	12.2	15.1	0.819	5.501	0.223	0.819	0.763	0.782
Decision Tree	0.829	0.829	0.826	0.811	0.848	0.842	64.9	67.8	12.2	15.1	0.819	5.501	0.223	0.839	0.831	0.818
Random Forest	0.829	0.829	0.826	0.811	0.848	0.842	64.9	67.8	12.2	15.1	0.819	5.501	0.223	0.842	0.811	0.815
Gradient Boosting	0.829	0.829	0.826	0.811	0.846	0.841	64.9	67.7	12.3	15.1	0.818	5.474	0.223	0.842	0.812	0.81
Extreme Gradient Boosting	0.829	0.829	0.825	0.81	0.848	0.842	64.8	67.8	12.2	15.2	0.818	5.493	0.224	0.837	0.823	0.81
Light Gradient Boosting	0.829	0.829	0.826	0.811	0.848	0.842	64.9	67.8	12.2	15.1	0.819	5.501	0.223	0.838	0.808	0.812
Category Gradient Boosting	0.829	0.829	0.826	0.811	0.848	0.842	64.9	67.8	12.2	15.1	0.819	5.501	0.223	0.838	0.816	0.82
Support Vector Machine	0.829	0.829	0.826	0.811	0.848	0.842	64.9	67.8	12.2	15.1	0.819	5.501	0.223	0.825	0.787	0.791
Artificial Neural Network	0.826	0.826	0.823	0.809	0.844	0.839	64.7	67.5	12.5	15.3	0.816	5.356	0.227	0.827	0.784	0.789
K-Nearest Neightbors	0.759	0.759	0.734	0.678	0.84	0.808	54.2	67.2	12.8	25.8	0.727	4.463	0.384	0.803	0.771	0.775
Genetic Programming	0.829	0.829	0.826	0.811	0.848	0.842	64.9	67.8	12.2	15.1	0.819	5.501	0.223	0.835	0.875	0.789
ExSTraCS	0.83	0.83	0.827	0.811	0.849	0.843	64.9	67.9	12.1	15.1	0.819	5.542	0.222	0.841	0.83	0.833

D2 = B_gametes_uni_4add_L_1_H_0

ML Algorithm	Balanced Accuracy	Accuracy	F1 Score	Sensitivity (Recall)	Specificity	Precision (PPV)	TP	TN	FP	FN	NPV	LR+	LR-	ROC AUC	PRC AUC	PRC APS
Naive Bayes	0.893	0.893	0.894	0.9	0.886	0.889	72.0	70.9	9.1	8.0	0.899	8.973	0.113	0.956	0.939	0.94
Logistic Regression	0.899	0.899	0.899	0.895	0.904	0.904	71.6	72.3	7.7	8.4	0.896	10.655	0.116	0.96	0.943	0.944
Decision Tree	0.921	0.921	0.921	0.921	0.921	0.923	73.7	73.7	6.3	6.3	0.922	10.809	0.085	0.974	0.976	0.967
Random Forest	0.92	0.92	0.921	0.936	0.904	0.909	74.9	72.3	7.7	5.1	0.936	12.427	0.07	0.98	0.977	0.977
Gradient Boosting	0.916	0.916	0.917	0.926	0.906	0.909	74.1	72.5	7.5	5.9	0.925	12.519	0.082	0.98	0.981	0.978
Extreme Gradient Boosting	0.926	0.926	0.926	0.932	0.919	0.921	74.6	73.5	6.5	5.4	0.932	18.066	0.074	0.983	0.982	0.982
Light Gradient Boosting	0.921	0.921	0.921	0.924	0.918	0.919	73.9	73.4	6.6	6.1	0.925	14.068	0.083	0.981	0.98	0.98
Category Gradient Boosting	0.921	0.921	0.922	0.932	0.91	0.914	74.6	72.8	7.2	5.4	0.932	12.303	0.074	0.981	0.98	0.98
Support Vector Machine	0.892	0.892	0.894	0.902	0.882	0.887	72.2	70.6	9.4	7.8	0.901	8.859	0.11	0.96	0.949	0.95
Artificial Neural Network	0.898	0.898	0.899	0.914	0.881	0.886	73.1	70.5	9.5	6.9	0.911	8.896	0.099	0.959	0.944	0.945
K-Nearest Neightbors	0.909	0.909	0.908	0.889	0.93	0.93	71.1	74.4	5.6	8.9	0.894	20.688	0.119	0.968	0.962	0.962
Genetic Programming	0.919	0.919	0.918	0.908	0.93	0.929	72.6	74.4	5.6	7.4	0.91	13.67	0.099	0.975	0.967	0.972
ExSTraCS	0.924	0.924	0.925	0.939	0.909	0.913	75.1	72.7	7.3	4.9	0.938	12.013	0.067	0.982	0.981	0.981

D3 = C_gametes_uni_4het_L_1_H_0

ML Algorithm	Balanced Accuracy	Accuracy	F1 Score	Sensitivity (Recall)	Specificity	Precision (PPV)	TP	TN	FP	FN	NPV	LR+	LR-	ROC AUC	PRC AUC	PRC APS
Naive Bayes	0.564	0.564	0.55	0.535	0.594	0.568	42.8	47.5	32.5	37.2	0.561	1.325	0.785	0.595	0.578	0.584
Logistic Regression	0.606	0.606	0.591	0.571	0.641	0.618	45.7	51.3	28.7	34.3	0.6	1.648	0.67	0.648	0.629	0.634
Decision Tree	0.59	0.59	0.605	0.632	0.548	0.582	50.6	43.8	36.2	29.4	0.601	1.403	0.671	0.626	0.605	0.601
Random Forest	0.617	0.617	0.608	0.599	0.635	0.623	47.9	50.8	29.2	32.1	0.614	1.679	0.633	0.665	0.643	0.649
Gradient Boosting	0.615	0.615	0.605	0.592	0.638	0.623	47.4	51.0	29.0	32.6	0.611	1.673	0.639	0.668	0.652	0.657
Extreme Gradient Boosting	0.626	0.626	0.615	0.6	0.652	0.635	48.0	52.2	27.8	32.0	0.622	1.781	0.616	0.671	0.649	0.654
Light Gradient Boosting	0.615	0.615	0.604	0.589	0.641	0.623	47.1	51.3	28.7	32.9	0.61	1.675	0.642	0.661	0.645	0.651
Category Gradient Boosting	0.629	0.629	0.618	0.602	0.655	0.638	48.2	52.4	27.6	31.8	0.623	1.789	0.608	0.672	0.654	0.659
Support Vector Machine	0.594	0.594	0.584	0.575	0.612	0.599	46.0	49.0	31.0	34.0	0.592	1.516	0.697	0.63	0.615	0.621
Artificial Neural Network	0.601	0.601	0.592	0.584	0.619	0.605	46.7	49.5	30.5	33.3	0.6	1.545	0.673	0.632	0.611	0.617
K-Nearest Neightbors	0.553	0.553	0.492	0.441	0.665	0.569	35.3	53.2	26.8	44.7	0.545	1.343	0.84	0.59	0.582	0.581
Genetic Programming	0.591	0.591	0.53	0.464	0.719	0.624	37.1	57.5	22.5	42.9	0.573	1.699	0.748	0.591	0.674	0.562
ExSTraCS	0.603	0.603	0.575	0.541	0.665	0.619	43.3	53.2	26.8	36.7	0.594	1.654	0.691	0.657	0.645	0.651

D4 = D_gametes_2way_epistasis_L_2_H_0

ML Algorithm	Balanced Accuracy	Accuracy	F1 Score	Sensitivity (Recall)	Specificity	Precision (PPV)	TP	TN	FP	FN	NPV	LR+	LR-	ROC AUC	PRC AUC	PRC APS
Naive Bayes	0.493	0.493	0.495	0.499	0.488	0.493	39.9	39.0	41.0	40.1	0.493	0.979	1.036	0.494	0.492	0.5
Logistic Regression	0.494	0.494	0.253	0.252	0.735	0.287	20.2	58.8	21.2	59.8	0.494	0.554	1.026	0.488	0.593	0.499
Decision Tree	0.489	0.489	0.485	0.484	0.495	0.489	38.7	39.6	40.4	41.3	0.49	0.966	1.054	0.489	0.512	0.505
Random Forest	0.609	0.609	0.605	0.6	0.619	0.611	48.0	49.5	30.5	32.0	0.609	1.597	0.651	0.658	0.641	0.646
Gradient Boosting	0.711	0.711	0.717	0.734	0.689	0.702	58.7	55.1	24.9	21.3	0.723	2.496	0.406	0.77	0.725	0.73
Extreme Gradient Boosting	0.735	0.735	0.739	0.754	0.716	0.727	60.3	57.3	22.7	19.7	0.746	2.696	0.344	0.802	0.751	0.756
Light Gradient Boosting	0.746	0.746	0.753	0.778	0.715	0.733	62.2	57.2	22.8	17.8	0.767	2.856	0.313	0.814	0.772	0.776
Category Gradient Boosting	0.786	0.786	0.804	0.875	0.698	0.744	70.0	55.8	24.2	10.0	0.849	2.934	0.18	0.845	0.802	0.806
Support Vector Machine	0.599	0.599	0.598	0.599	0.599	0.6	47.9	47.9	32.1	32.1	0.599	1.523	0.675	0.627	0.617	0.623
Artificial Neural Network	0.656	0.656	0.651	0.645	0.666	0.659	51.6	53.3	26.7	28.4	0.653	1.952	0.534	0.709	0.684	0.69
K-Nearest Neightbors	0.56	0.56	0.533	0.509	0.611	0.569	40.7	48.9	31.1	39.3	0.555	1.353	0.823	0.588	0.579	0.586
Genetic Programming	0.571	0.571	0.388	0.49	0.652	0.407	39.2	52.2	27.8	40.8	0.516	1.069	0.538	0.568	0.667	0.554
ExSTraCS	0.802	0.802	0.825	0.932	0.672	0.741	74.6	53.8	26.2	5.4	0.913	2.886	0.099	0.854	0.818	0.82

D5 = E_gametes_2way_epi_2het_L_2_H_0

ML Algorithm	Balanced Accuracy	Accuracy	F1 Score	Sensitivity (Recall)	Specificity	Precision (PPV)	TP	TN	FP	FN	NPV	LR+	LR-	ROC AUC	PRC AUC	PRC APS
Naive Bayes	0.491	0.491	0.493	0.5	0.482	0.49	40.0	38.6	41.4	40.0	0.493	0.969	1.039	0.487	0.497	0.504
Logistic Regression	0.507	0.507	0.494	0.489	0.525	0.506	39.1	42.0	38.0	40.9	0.509	1.042	0.982	0.5	0.511	0.518
Decision Tree	0.479	0.479	0.454	0.439	0.52	0.477	35.1	41.6	38.4	44.9	0.481	0.931	1.098	0.485	0.495	0.495
Random Forest	0.561	0.561	0.566	0.575	0.548	0.559	46.0	43.8	36.2	34.0	0.564	1.286	0.782	0.579	0.55	0.558
Gradient Boosting	0.619	0.619	0.621	0.628	0.61	0.618	50.2	48.8	31.2	29.8	0.622	1.653	0.615	0.665	0.646	0.651
Extreme Gradient Boosting	0.621	0.621	0.626	0.634	0.608	0.62	50.7	48.6	31.4	29.3	0.623	1.701	0.621	0.664	0.638	0.643
Light Gradient Boosting	0.639	0.639	0.643	0.652	0.625	0.635	52.2	50.0	30.0	27.8	0.644	1.791	0.564	0.675	0.641	0.648
Category Gradient Boosting	0.663	0.663	0.665	0.668	0.659	0.663	53.4	52.7	27.3	26.6	0.664	2.032	0.51	0.723	0.695	0.699
Support Vector Machine	0.545	0.545	0.526	0.509	0.581	0.547	40.7	46.5	33.5	39.3	0.544	1.218	0.844	0.55	0.545	0.552
Artificial Neural Network	0.552	0.552	0.55	0.546	0.559	0.553	43.7	44.7	35.3	36.3	0.552	1.242	0.813	0.568	0.564	0.57
K-Nearest Neightbors	0.532	0.532	0.521	0.511	0.554	0.537	40.9	44.3	35.7	39.1	0.529	1.169	0.897	0.533	0.528	0.534
Genetic Programming	0.611	0.611	0.614	0.726	0.495	0.543	58.1	39.6	40.4	21.9	0.538	1.415	0.407	0.617	0.686	0.588
ExSTraCS	0.688	0.688	0.696	0.712	0.664	0.68	57.0	53.1	26.9	23.0	0.698	2.192	0.438	0.74	0.693	0.699

D6 = F_gametes_3way_epistasis_L_3_H_0

ML Algorithm	Balanced Accuracy	Accuracy	F1 Score	Sensitivity (Recall)	Specificity	Precision (PPV)	TP	TN	FP	FN	NPV	LR+	LR-	ROC AUC	PRC AUC	PRC APS
Naive Bayes	0.478	0.478	0.467	0.461	0.495	0.476	36.9	39.6	40.4	43.1	0.48	0.911	1.087	0.473	0.488	0.496
Logistic Regression	0.483	0.483	0.26	0.25	0.716	0.28	20.0	57.3	22.7	60.0	0.484	0.53	1.071	0.473	0.586	0.49
Decision Tree	0.505	0.505	0.486	0.47	0.54	0.508	37.6	43.2	36.8	42.4	0.503	1.042	0.996	0.503	0.524	0.515
Random Forest	0.479	0.479	0.47	0.462	0.496	0.481	37.0	39.7	40.3	43.0	0.477	0.944	1.121	0.474	0.491	0.499
Gradient Boosting	0.512	0.512	0.519	0.526	0.499	0.512	42.1	39.9	40.1	37.9	0.513	1.073	0.968	0.501	0.497	0.504
Extreme Gradient Boosting	0.492	0.492	0.483	0.478	0.507	0.492	38.2	40.6	39.4	41.8	0.493	0.974	1.04	0.494	0.497	0.504
Light Gradient Boosting	0.498	0.498	0.495	0.492	0.504	0.498	39.4	40.3	39.7	40.6	0.498	0.999	1.013	0.494	0.489	0.498
Category Gradient Boosting	0.521	0.521	0.527	0.532	0.51	0.523	42.6	40.8	39.2	37.4	0.52	1.113	0.94	0.535	0.529	0.536
Support Vector Machine	0.529	0.529	0.547	0.578	0.48	0.526	46.2	38.4	41.6	33.8	0.534	1.127	0.887	0.532	0.527	0.531
Artificial Neural Network	0.504	0.504	0.495	0.494	0.514	0.502	39.5	41.1	38.9	40.5	0.505	1.021	0.99	0.493	0.491	0.498
K-Nearest Neightbors	0.507	0.507	0.495	0.484	0.531	0.507	38.7	42.5	37.5	41.3	0.508	1.039	0.977	0.51	0.546	0.511
Genetic Programming	0.491	0.491	0.358	0.426	0.555	0.385	34.1	44.4	35.6	45.9	0.413	0.751	0.783	0.492	0.604	0.497
ExSTraCS	0.548	0.548	0.56	0.589	0.507	0.539	47.1	40.6	39.4	32.9	0.563	1.208	0.82	0.564	0.548	0.554

Median Model Prediction Statistics (Rounded to 3 Decimal Points): Page 1

D1 = A_gametes_univariate_L_1_H_0

ML Algorithm	Balanced Accuracy	Accuracy	F1 Score	Sensitivity (Recall)	Specificity	Precision (PPV)	TP	TN	FP	FN	NPV	LR+	LR-	ROC AUC	PRC AUC	PRC APS
Naive Bayes	0.794	0.794	0.792	0.775	0.812	0.807	62.0	65.0	15.0	18.0	0.784	4.174	0.275	0.816	0.772	0.775
Logistic Regression	0.834	0.834	0.831	0.806	0.844	0.838	64.5	67.5	12.5	15.5	0.813	5.163	0.23	0.825	0.783	0.785
Decision Tree	0.834	0.834	0.831	0.806	0.844	0.838	64.5	67.5	12.5	15.5	0.813	5.163	0.23	0.838	0.828	0.816
Random Forest	0.834	0.834	0.831	0.806	0.844	0.838	64.5	67.5	12.5	15.5	0.813	5.163	0.23	0.841	0.813	0.818
Gradient Boosting	0.834	0.834	0.831	0.806	0.844	0.838	64.5	67.5	12.5	15.5	0.812	5.163	0.231	0.84	0.802	0.805
Extreme Gradient Boosting	0.834	0.834	0.831	0.806	0.844	0.838	64.5	67.5	12.5	15.5	0.813	5.163	0.23	0.837	0.823	0.815
Light Gradient Boosting	0.834	0.834	0.831	0.806	0.844	0.838	64.5	67.5	12.5	15.5	0.813	5.163	0.23	0.832	0.799	0.802
Category Gradient Boosting	0.834	0.834	0.831	0.806	0.844	0.838	64.5	67.5	12.5	15.5	0.813	5.163	0.23	0.834	0.816	0.818
Support Vector Machine	0.834	0.834	0.831	0.806	0.844	0.838	64.5	67.5	12.5	15.5	0.813	5.163	0.23	0.831	0.788	0.795
Artificial Neural Network	0.828	0.828	0.82	0.8	0.844	0.839	64.0	67.5	12.5	16.0	0.806	5.202	0.24	0.825	0.778	0.784
K-Nearest Neightbors	0.769	0.769	0.745	0.675	0.844	0.814	54.0	67.5	12.5	26.0	0.726	4.375	0.377	0.797	0.765	0.768
Genetic Programming	0.834	0.834	0.831	0.806	0.844	0.838	64.5	67.5	12.5	15.5	0.813	5.163	0.23	0.836	0.87	0.779
ExSTraCS	0.834	0.834	0.831	0.806	0.85	0.84	64.5	68.0	12.0	15.5	0.813	5.25	0.23	0.848	0.836	0.838

D2 = B_gametes_uni_4add_L_1_H_0

ML Algorithm	Balanced Accuracy	Accuracy	F1 Score	Sensitivity (Recall)	Specificity	Precision (PPV)	TP	TN	FP	FN	NPV	LR+	LR-	ROC AUC	PRC AUC	PRC APS
Naive Bayes	0.891	0.891	0.893	0.888	0.894	0.893	71.0	71.5	8.5	9.0	0.892	8.382	0.122	0.955	0.942	0.942
Logistic Regression	0.9	0.9	0.899	0.888	0.9	0.901	71.0	72.0	8.0	9.0	0.893	9.063	0.119	0.96	0.949	0.952
Decision Tree	0.919	0.919	0.919	0.925	0.925	0.924	74.0	74.0	6.0	6.0	0.924	10.976	0.082	0.973	0.971	0.96
Random Forest	0.912	0.912	0.915	0.944	0.912	0.914	75.5	73.0	7.0	4.5	0.942	10.896	0.062	0.982	0.982	0.982
Gradient Boosting	0.916	0.916	0.915	0.938	0.906	0.909	75.0	72.5	7.5	5.0	0.932	10.417	0.073	0.981	0.98	0.98
Extreme Gradient Boosting	0.922	0.922	0.922	0.931	0.919	0.92	74.5	73.5	6.5	5.5	0.933	11.631	0.072	0.982	0.983	0.983
Light Gradient Boosting	0.912	0.912	0.915	0.931	0.919	0.915	74.5	73.5	6.5	5.5	0.93	10.833	0.075	0.98	0.979	0.978
Category Gradient Boosting	0.922	0.922	0.924	0.925	0.919	0.919	74.0	73.5	6.5	6.0	0.926	11.44	0.079	0.98	0.981	0.981
Support Vector Machine	0.9	0.9	0.901	0.906	0.888	0.89	72.5	71.0	9.0	7.5	0.901	8.111	0.11	0.959	0.957	0.957
Artificial Neural Network	0.897	0.897	0.9	0.925	0.869	0.876	74.0	69.5	10.5	6.0	0.921	7.064	0.086	0.96	0.957	0.958
K-Nearest Neightbors	0.909	0.909	0.906	0.888	0.944	0.939	71.0	75.5	4.5	9.0	0.891	15.675	0.122	0.974	0.972	0.97
Genetic Programming	0.919	0.919	0.918	0.912	0.925	0.924	73.0	74.0	6.0	7.0	0.914	12.167	0.095	0.977	0.974	0.974
ExSTraCS	0.919	0.919	0.919	0.944	0.919	0.922	75.5	73.5	6.5	4.5	0.939	11.929	0.065	0.983	0.983	0.983

D3 = C_gametes_uni_4het_L_1_H_0

| ML Algorithm	Balanced Accuracy	Accuracy	F1 Score	Sensitivity (Recall)	Specificity	Precision (PPV)	TP	TN	FP	FN	NPV	LR+	LR-	ROC AUC	PRC AUC	PRC APS
Naive Bayes	0.559	0.559	0.55	0.531	0.594	0.57	42.5	47.5	32.5	37.5	0.558	1.325	0.794	0.591	0.568	0.575
Logistic Regression	0.603	0.603	0.578	0.556	0.638	0.604	44.5	51.0	29.0	35.5	0.596	1.525	0.679	0.64	0.63	0.637
Decision Tree	0.594	0.594	0.622	0.669	0.55	0.586	53.5	44.0	36.0	26.5	0.61	1.416	0.64	0.626	0.608	0.598
Random Forest	0.628	0.628	0.634	0.606	0.625	0.623	48.5	50.0	30.0	31.5	0.636	1.656	0.573	0.656	0.64	0.646
Gradient Boosting	0.609	0.609	0.609	0.594	0.625	0.623	47.5	50.0	30.0	32.5	0.612	1.656	0.634	0.657	0.639	0.643
Extreme Gradient Boosting	0.625	0.625	0.604	0.581	0.656	0.637	46.5	52.5	27.5	33.5	0.615	1.754	0.626	0.672	0.638	0.645
Light Gradient Boosting	0.619	0.619	0.618	0.581	0.631	0.625	46.5	50.5	29.5	33.5	0.625	1.67	0.601	0.655	0.634	0.639
Category Gradient Boosting	0.628	0.628	0.627	0.606	0.644	0.643	48.5	51.5	28.5	31.5	0.626	1.802	0.596	0.665	0.643	0.648
Support Vector Machine	0.584	0.584	0.567	0.569	0.606	0.587	45.5	48.5	31.5	34.5	0.581	1.421	0.721	0.623	0.606	0.611
Artificial Neural Network	0.591	0.591	0.598	0.619	0.631	0.594	49.5	50.5	29.5	30.5	0.591	1.464	0.693	0.627	0.608	0.614
K-Nearest Neightbors	0.547	0.547	0.486	0.438	0.694	0.567	35.0	55.5	24.5	45.0	0.54	1.308	0.852	0.585	0.583	0.577
Genetic Programming	0.591	0.591	0.533	0.462	0.694	0.623	37.0	55.5	24.5	43.0	0.57	1.653	0.754	0.592	0.673	0.559
ExSTraCS	0.584	0.584	0.569	0.538	0.644	0.604	43.0	51.5	28.5	37.0	0.579	1.527	0.727	0.66	0.63	0.637

D4 = D_gametes_2way_epistasis_L_2_H_0

ML Algorithm	Balanced Accuracy	Accuracy	F1 Score	Sensitivity (Recall)	Specificity	Precision (PPV)	TP	TN	FP	FN	NPV	LR+	LR-	ROC AUC	PRC AUC	PRC APS
Naive Bayes	0.481	0.481	0.494	0.5	0.5	0.481	40.0	40.0	40.0	40.0	0.482	0.926	1.075	0.494	0.497	0.506
Logistic Regression	0.5	0.5	0.262	0.244	0.731	0.445	19.5	58.5	21.5	60.5	0.5	0.804	1.0	0.5	0.521	0.5
Decision Tree	0.494	0.494	0.491	0.494	0.512	0.493	39.5	41.0	39.0	40.5	0.494	0.974	1.024	0.487	0.511	0.503
Random Forest	0.622	0.622	0.623	0.619	0.619	0.621	49.5	49.5	30.5	30.5	0.623	1.643	0.607	0.682	0.649	0.655
Gradient Boosting	0.731	0.731	0.732	0.75	0.706	0.722	60.0	56.5	23.5	20.0	0.737	2.597	0.358	0.798	0.758	0.761
Extreme Gradient Boosting	0.734	0.734	0.739	0.769	0.712	0.722	61.5	57.0	23.0	18.5	0.748	2.592	0.337	0.808	0.768	0.774
Light Gradient Boosting	0.756	0.756	0.768	0.806	0.694	0.724	64.5	55.5	24.5	15.5	0.788	2.62	0.27	0.826	0.78	0.782
Category Gradient Boosting	0.794	0.794	0.811	0.875	0.688	0.746	70.0	55.0	25.0	10.0	0.852	2.939	0.174	0.854	0.81	0.814
Support Vector Machine	0.6	0.6	0.608	0.612	0.606	0.59	49.0	48.5	31.5	31.0	0.608	1.437	0.645	0.628	0.625	0.629
Artificial Neural Network	0.659	0.659	0.662	0.663	0.669	0.656	53.0	53.5	26.5	27.0	0.663	1.911	0.509	0.707	0.673	0.677
K-Nearest Neightbors	0.578	0.578	0.516	0.488	0.631	0.591	39.0	50.5	29.5	41.0	0.562	1.444	0.78	0.606	0.577	0.584
Genetic Programming	0.5	0.5	0.42	0.494	0.762	0.5	39.5	61.0	19.0	40.5	0.5	1.0	0.586	0.509	0.715	0.505
ExSTraCS	0.806	0.806	0.83	0.95	0.675	0.748	76.0	54.0	26.0	4.0	0.933	2.964	0.071	0.857	0.824	0.826

D5 = E_gametes_2way_epi_2het_L_2_H_0

| ML Algorithm	Balanced Accuracy	Accuracy	F1 Score	Sensitivity (Recall)	Specificity	Precision (PPV)	TP	TN	FP	FN	NPV	LR+	LR-	ROC AUC	PRC AUC	PRC APS
Naive Bayes	0.488	0.488	0.481	0.475	0.462	0.488	38.0	37.0	43.0	42.0	0.49	0.959	1.045	0.482	0.495	0.501
Logistic Regression	0.512	0.512	0.497	0.481	0.506	0.517	38.5	40.5	39.5	41.5	0.509	1.074	0.965	0.502	0.514	0.52
Decision Tree	0.462	0.462	0.435	0.419	0.512	0.462	33.5	41.0	39.0	46.5	0.466	0.86	1.148	0.472	0.485	0.485
Random Forest	0.566	0.566	0.575	0.581	0.531	0.56	46.5	42.5	37.5	33.5	0.569	1.273	0.758	0.576	0.555	0.563
Gradient Boosting	0.612	0.612	0.621	0.65	0.581	0.612	52.0	46.5	33.5	28.0	0.618	1.58	0.619	0.67	0.649	0.655
Extreme Gradient Boosting	0.616	0.616	0.619	0.625	0.612	0.614	50.0	49.0	31.0	30.0	0.618	1.592	0.618	0.673	0.633	0.638
Light Gradient Boosting	0.638	0.638	0.644	0.656	0.631	0.633	52.5	50.5	29.5	27.5	0.643	1.725	0.556	0.684	0.651	0.659
Category Gradient Boosting	0.65	0.65	0.65	0.669	0.656	0.656	53.5	52.5	27.5	26.5	0.65	1.909	0.539	0.726	0.699	0.703
Support Vector Machine	0.559	0.559	0.532	0.506	0.581	0.559	40.5	46.5	33.5	39.5	0.557	1.268	0.796	0.557	0.54	0.547
Artificial Neural Network	0.553	0.553	0.55	0.538	0.556	0.552	43.0	44.5	35.5	37.0	0.553	1.234	0.809	0.569	0.567	0.573
K-Nearest Neightbors	0.541	0.541	0.526	0.506	0.581	0.547	40.5	46.5	33.5	39.5	0.536	1.206	0.865	0.533	0.524	0.532
Genetic Programming	0.644	0.644	0.684	0.775	0.556	0.618	62.0	44.5	35.5	18.0	0.691	1.617	0.412	0.658	0.687	0.614
ExSTraCS	0.684	0.684	0.689	0.712	0.656	0.673	57.0	52.5	27.5	23.0	0.691	2.056	0.447	0.732	0.683	0.689

D6 = F_gametes_3way_epistasis_L_3_H_0

| ML Algorithm	Balanced Accuracy	Accuracy	F1 Score	Sensitivity (Recall)	Specificity	Precision (PPV)	TP	TN	FP	FN	NPV	LR+	LR-	ROC AUC	PRC AUC	PRC APS
Naive Bayes	0.478	0.478	0.475	0.475	0.488	0.474	38.0	39.0	41.0	42.0	0.48	0.902	1.085	0.464	0.472	0.481
Logistic Regression	0.491	0.491	0.339	0.294	0.656	0.434	23.5	52.5	27.5	56.5	0.494	0.769	1.025	0.48	0.508	0.498
Decision Tree	0.5	0.5	0.479	0.469	0.562	0.499	37.5	45.0	35.0	42.5	0.5	0.998	0.999	0.5	0.519	0.515
Random Forest	0.497	0.497	0.488	0.469	0.506	0.496	37.5	40.5	39.5	42.5	0.497	0.984	1.01	0.468	0.479	0.486
Gradient Boosting	0.5	0.5	0.516	0.538	0.5	0.5	43.0	40.0	40.0	37.0	0.5	1.0	1.0	0.498	0.494	0.502
Extreme Gradient Boosting	0.494	0.494	0.495	0.494	0.525	0.494	39.5	42.0	38.0	40.5	0.494	0.976	1.026	0.499	0.491	0.499
Light Gradient Boosting	0.503	0.503	0.494	0.5	0.5	0.503	40.0	40.0	40.0	40.0	0.503	1.012	0.988	0.476	0.478	0.486
Category Gradient Boosting	0.534	0.534	0.542	0.531	0.512	0.533	42.5	41.0	39.0	37.5	0.535	1.143	0.869	0.533	0.533	0.541
Support Vector Machine	0.519	0.519	0.545	0.588	0.488	0.517	47.0	39.0	41.0	33.0	0.524	1.071	0.908	0.532	0.509	0.518
Artificial Neural Network	0.494	0.494	0.5	0.506	0.488	0.494	40.5	39.0	41.0	39.5	0.493	0.977	1.027	0.487	0.483	0.491
K-Nearest Neightbors	0.503	0.503	0.494	0.481	0.544	0.503	38.5	43.5	36.5	41.5	0.503	1.013	0.988	0.512	0.507	0.508
Genetic Programming	0.5	0.5	0.37	0.319	0.644	0.495	25.5	51.5	28.5	54.5	0.498	0.979	1.0	0.5	0.606	0.5
ExSTraCS	0.512	0.512	0.533	0.575	0.519	0.512	46.0	41.5	38.5	34.0	0.514	1.048	0.946	0.527	0.552	0.559

Feature Importance Summary: D1 = A_gametes_univariate_L_1_H_0
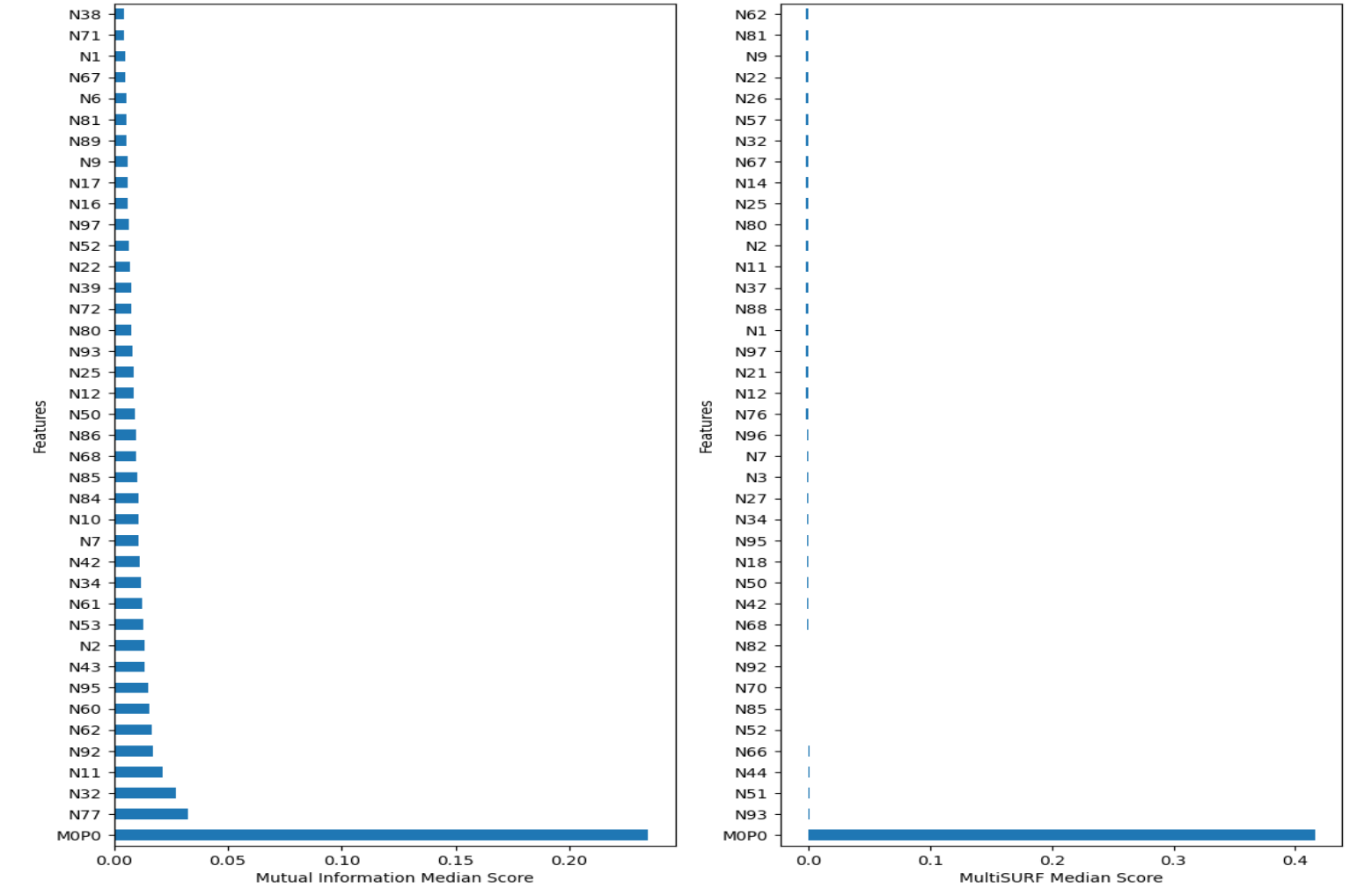
Composite Feature Importance Plot (Normalized and Performance Weighted)

Generated with the URBS-Lab STREAMLINE: (https://github.com/UrbsLab/STREAMLINE)
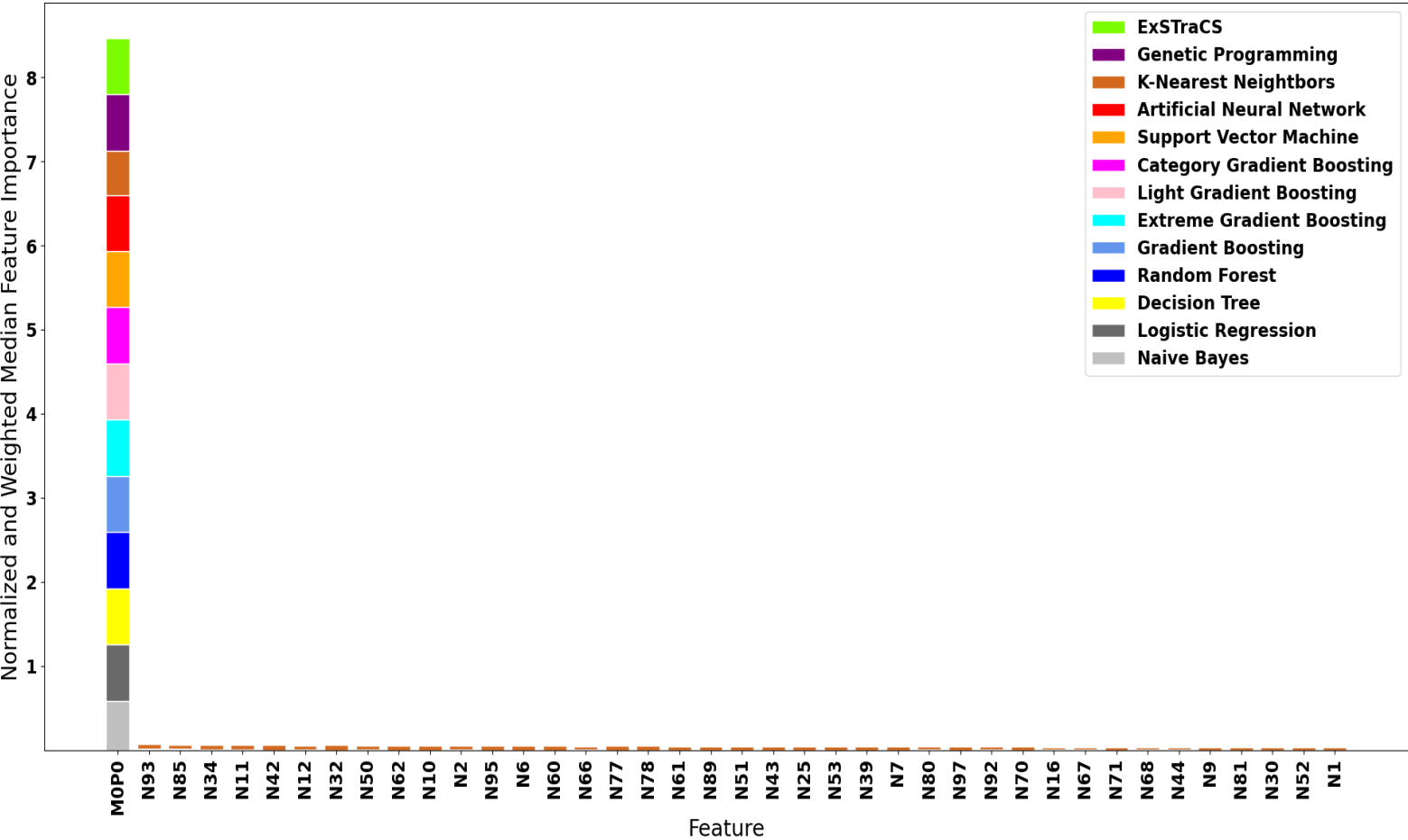
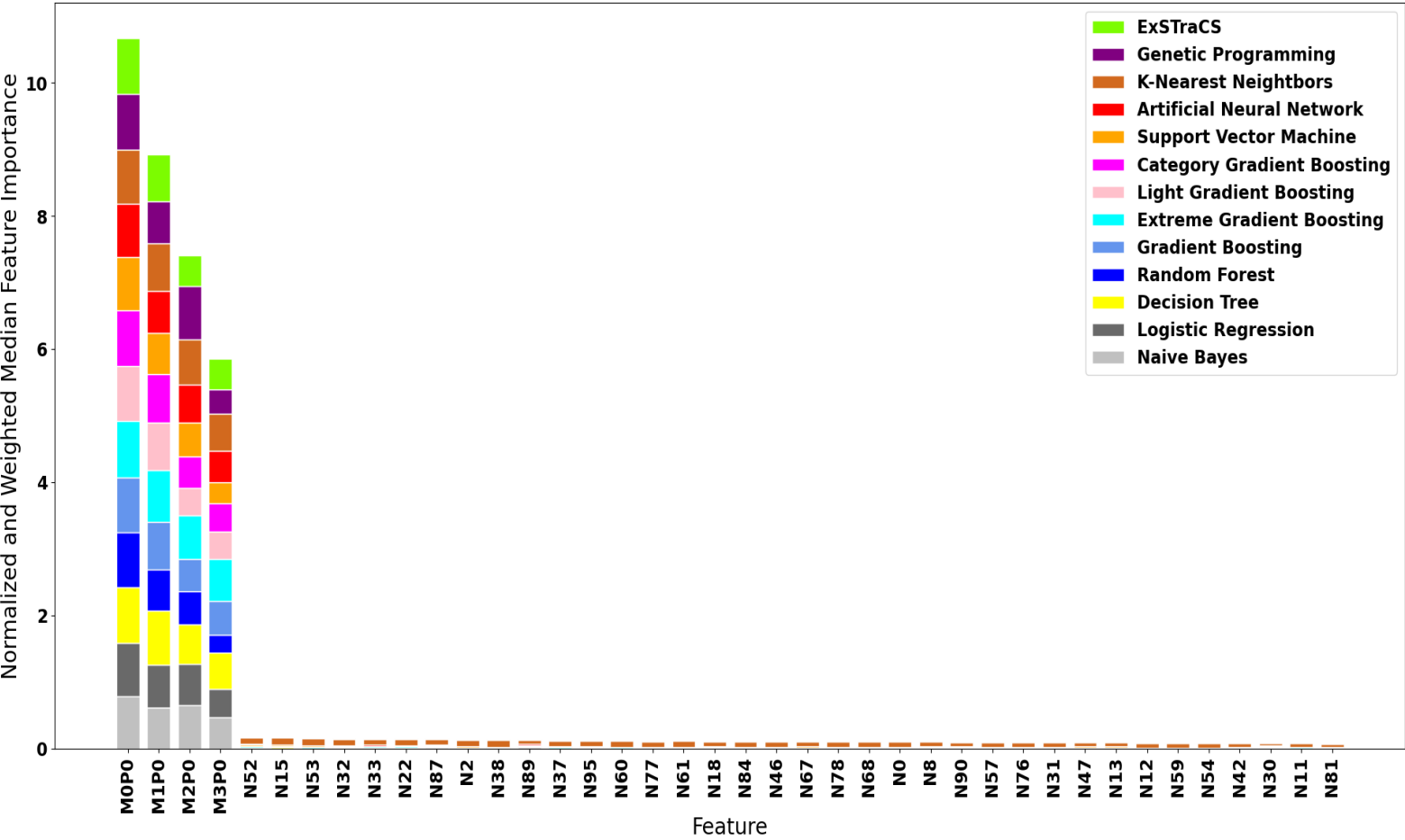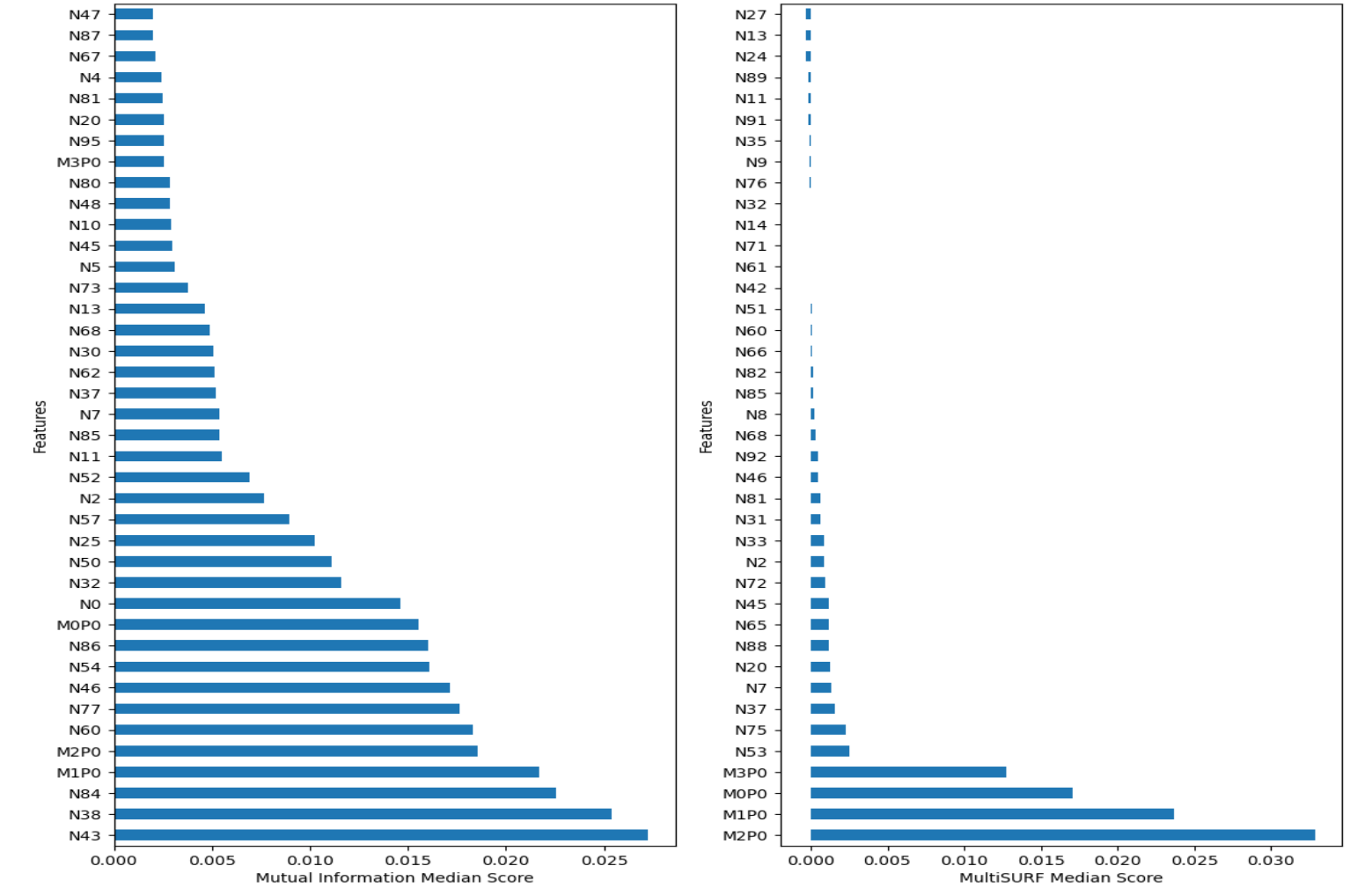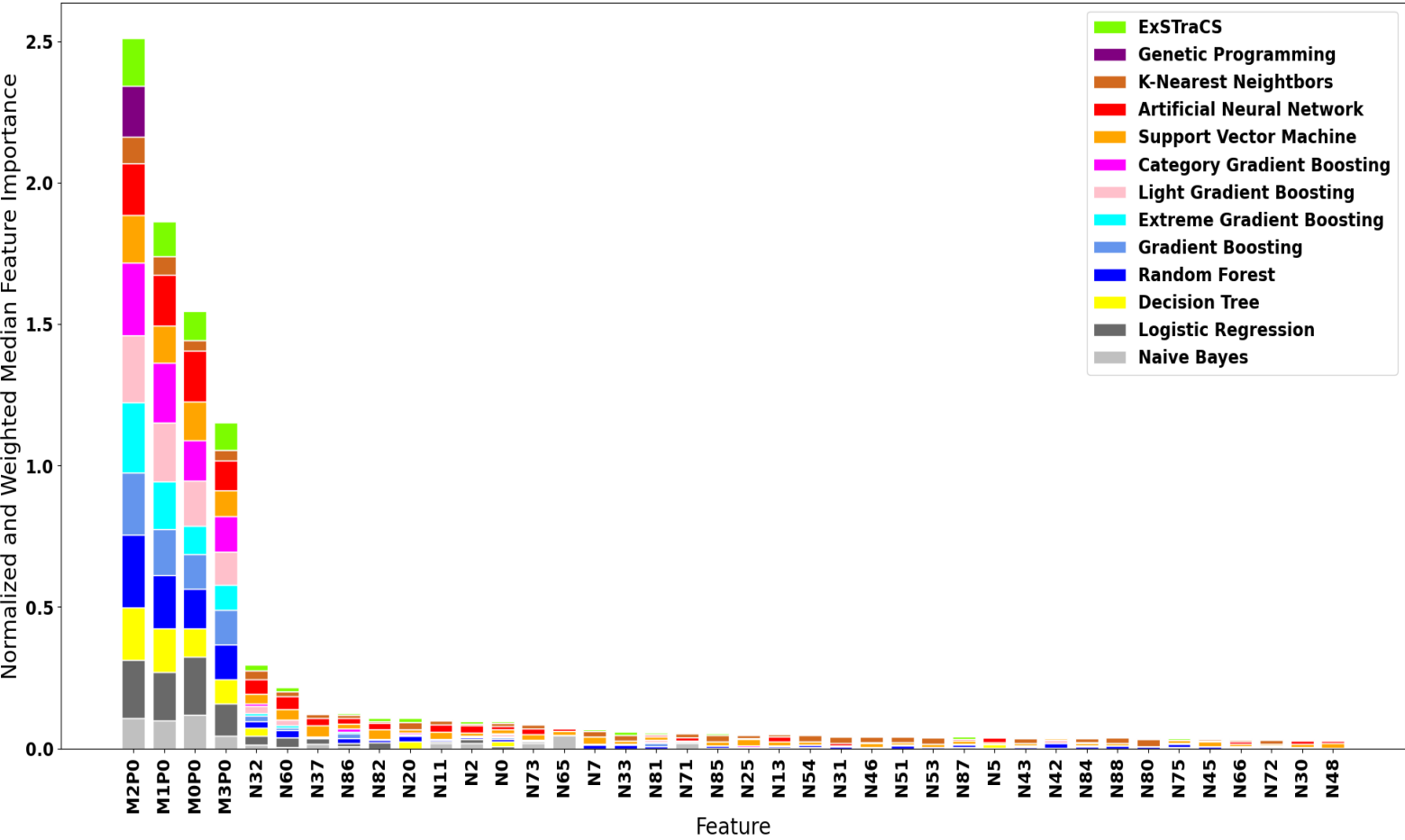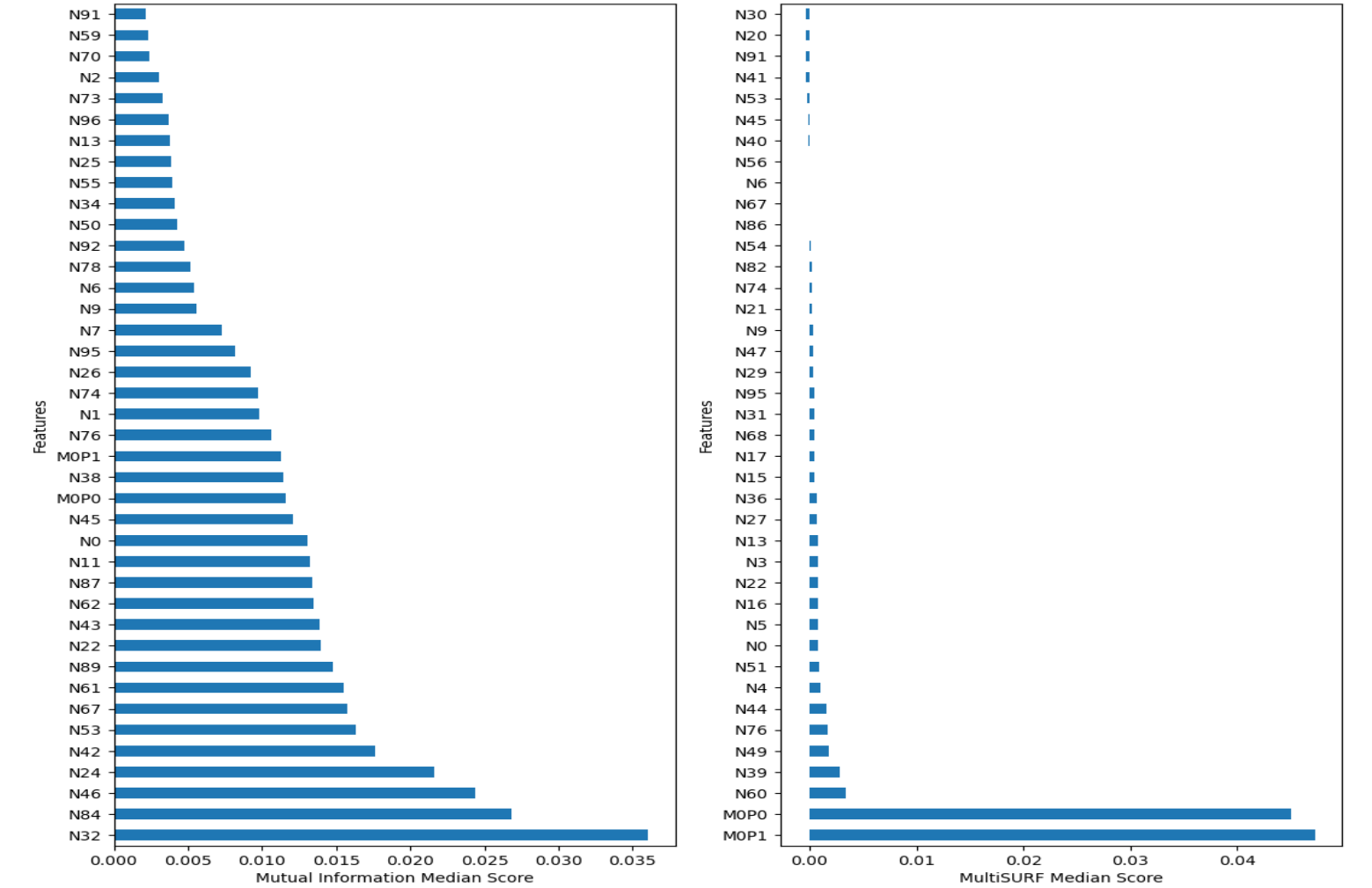# Feature Importance Summary: D2 = B_gametes_uni_4add_L_1_H_0
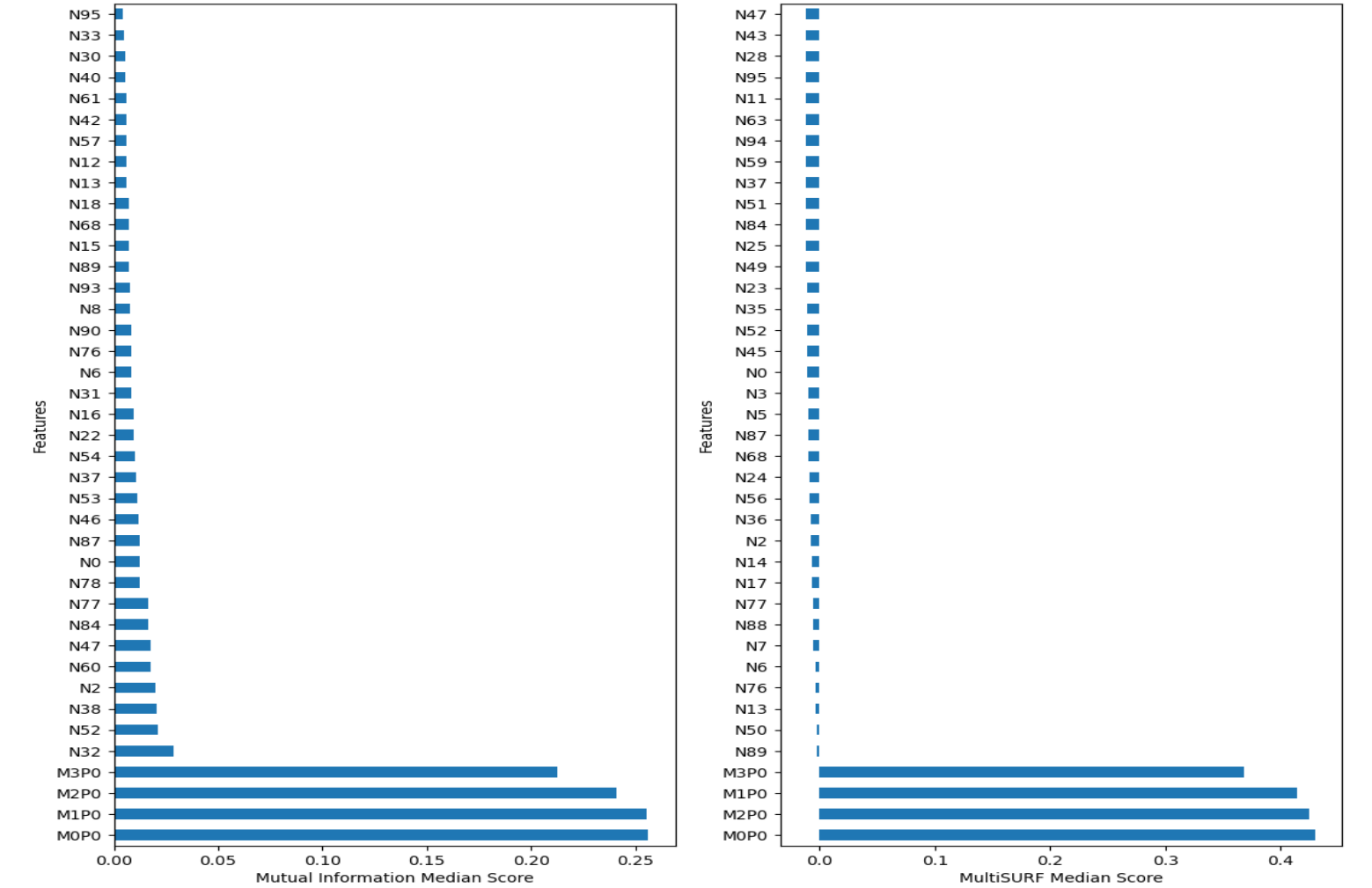
Composite Feature Importance Plot (Normalized and Performance Weighted)

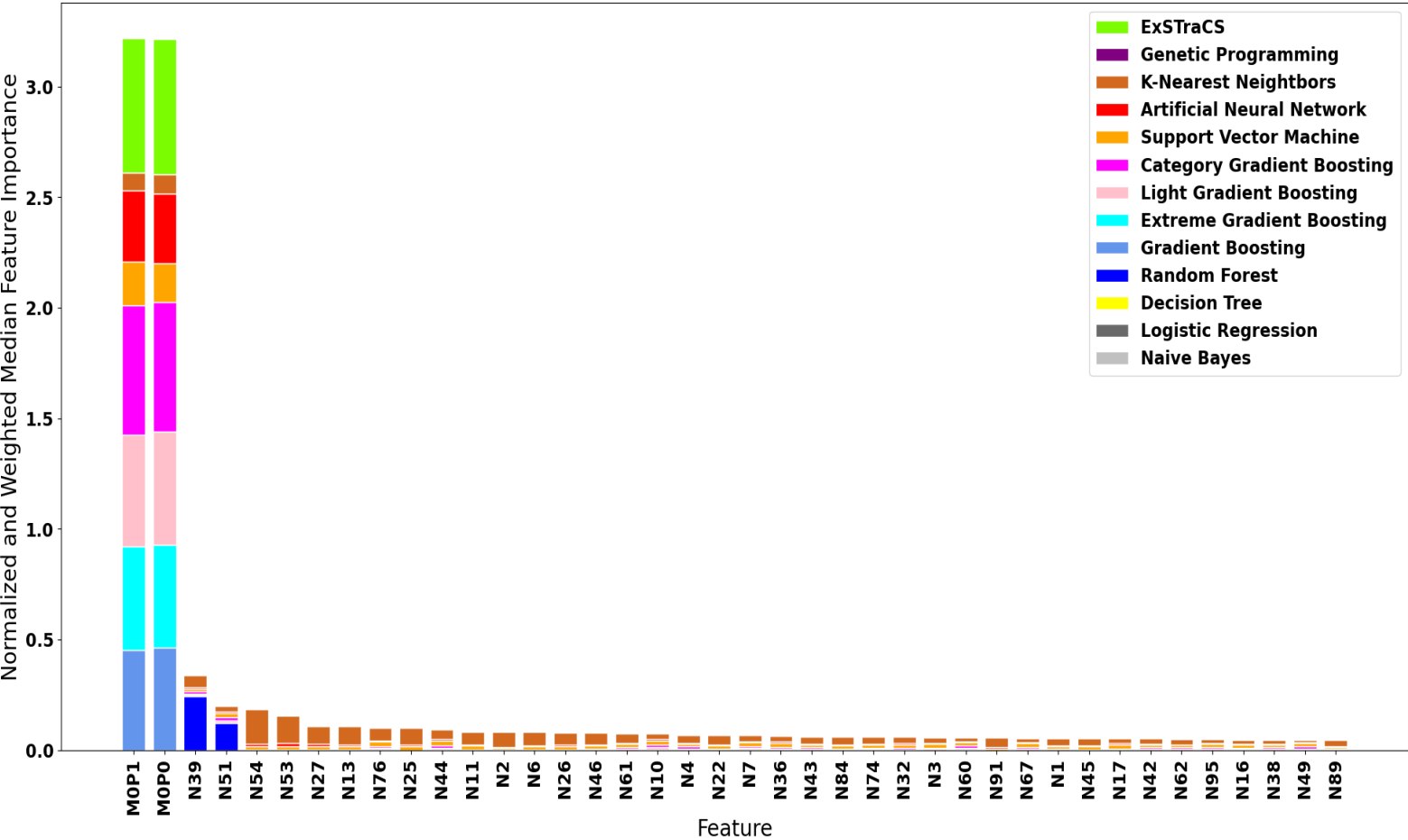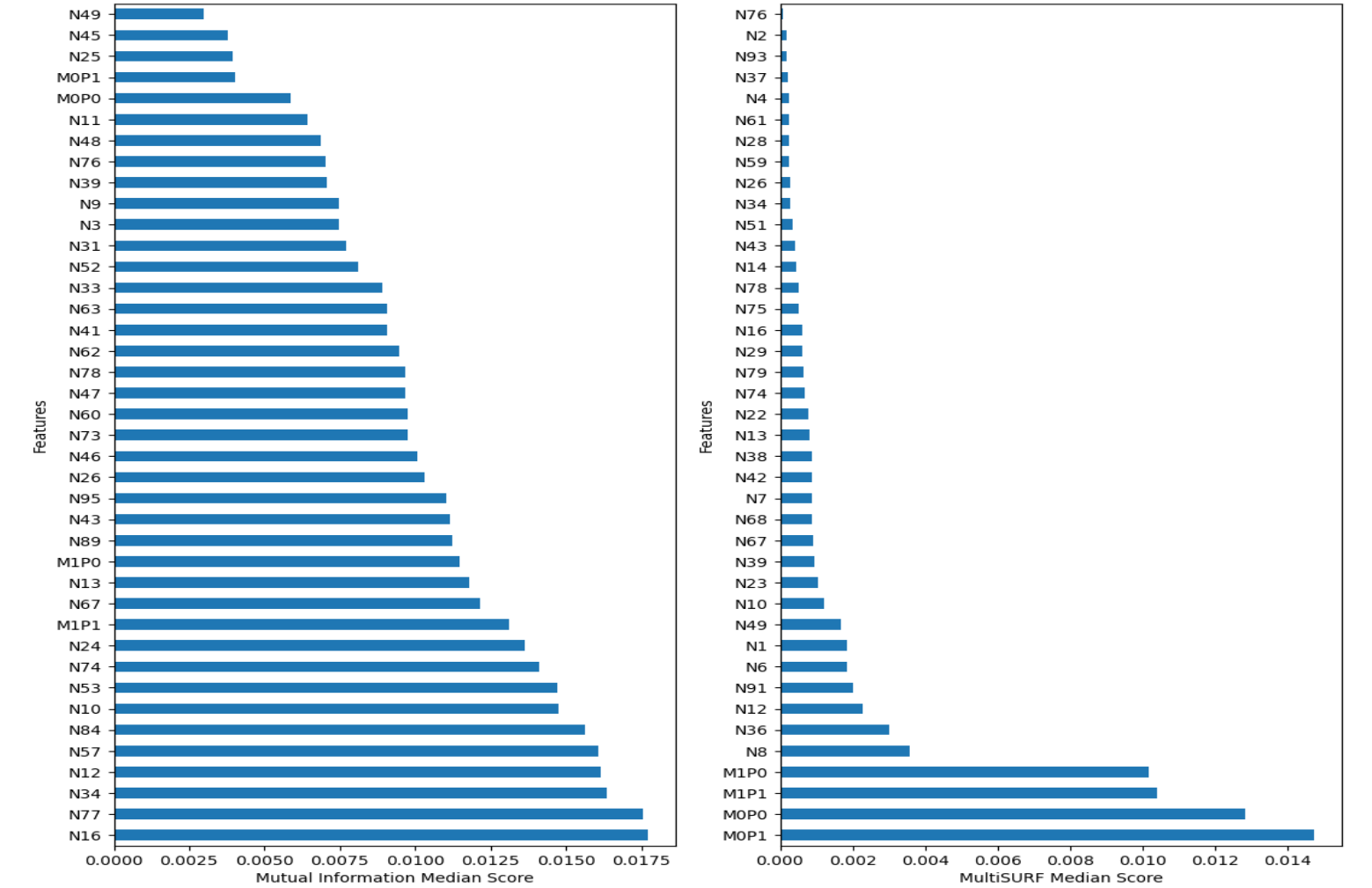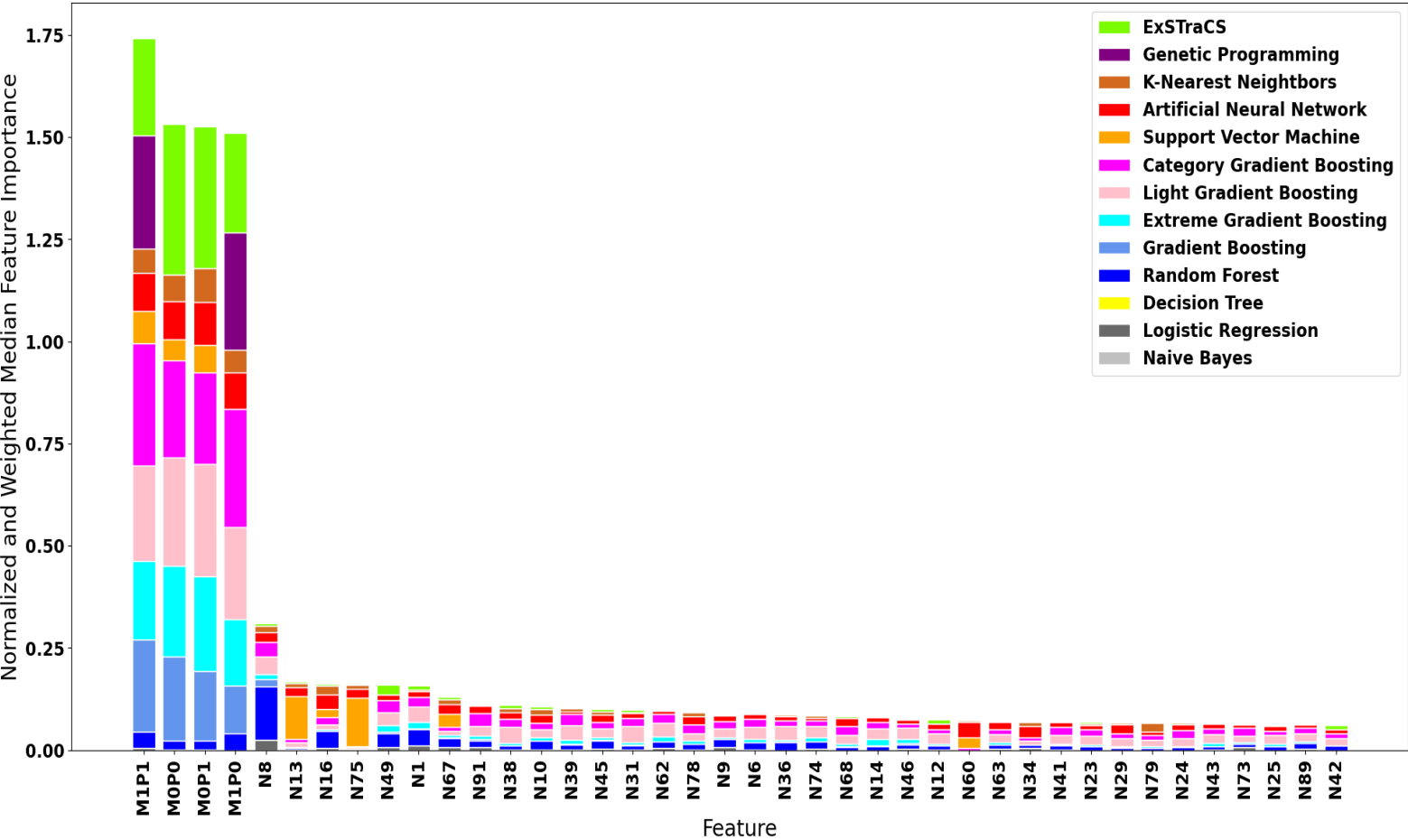# Feature Importance Summary: D3 = C_gametes_uni_4het_L_1_H_0

Composite Feature Importance Plot (Normalized and Performance Weighted)

Generated with the URBS-Lab STREAMLINE: (https://github.com/UrbsLab/STREAMLINE)

Feature Importance Summary: D4 = D_gametes_2way_epistasis_L_2_H_0

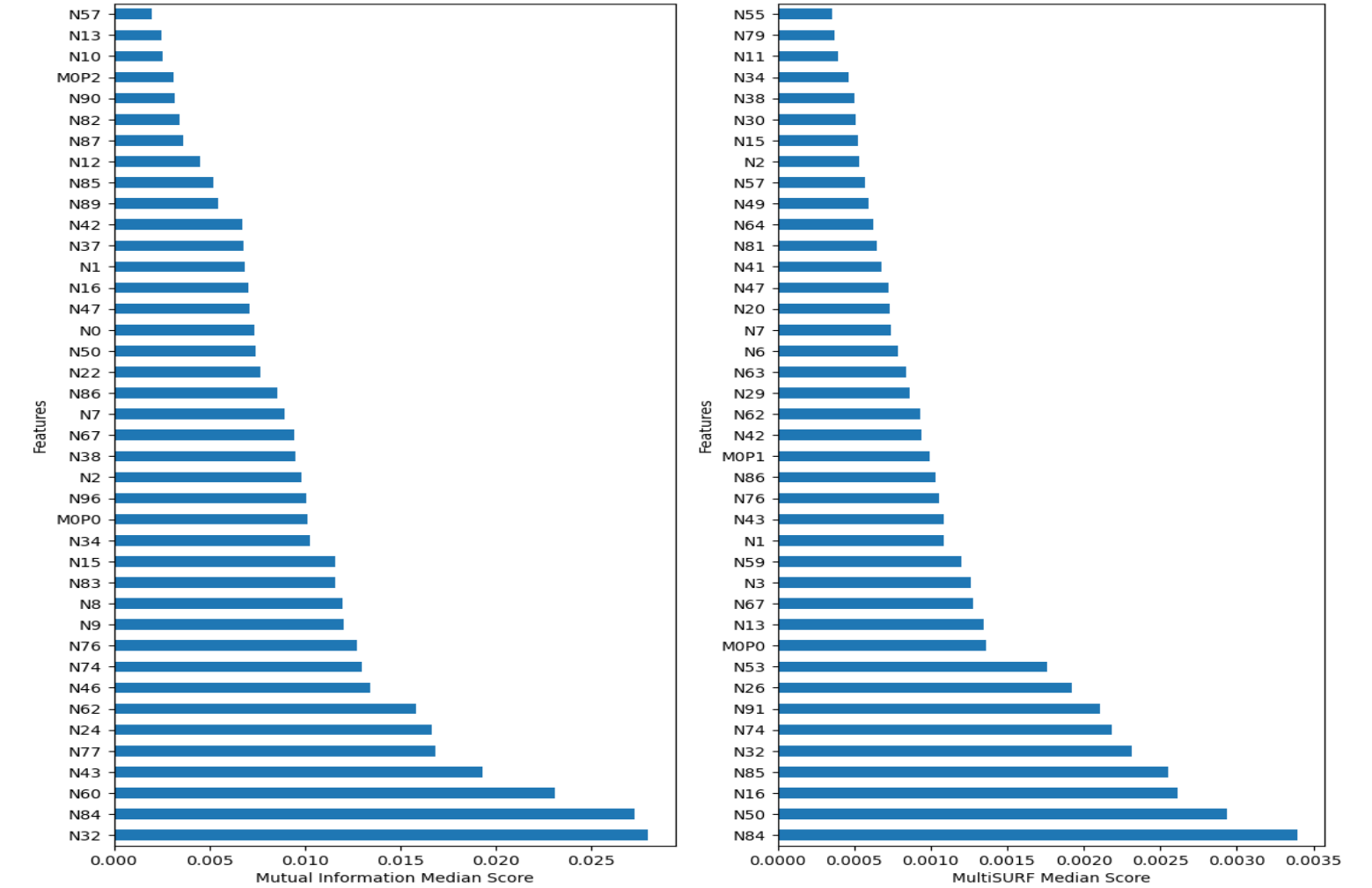Composite Feature Importance Plot (Normalized and Performance Weighted)

# Feature Importance Summary: D5 = E_gametes_2way_epi_2het_L_2_H_0

Composite Feature Importance Plot (Normalized and Performance Weighted)

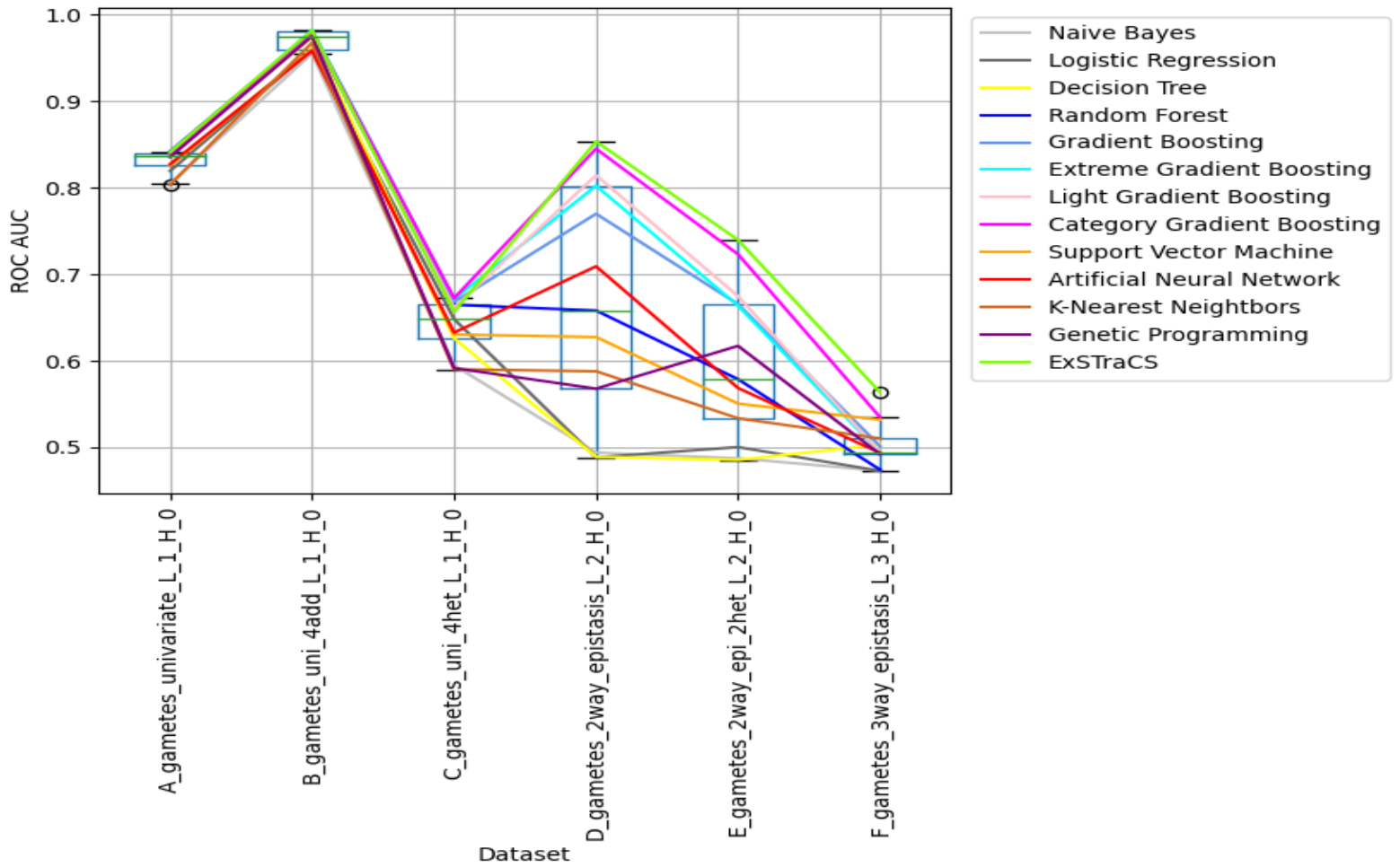*Generated with the URBS-Lab STREAMLINE: (https://github.com/UrbsLab/STREAMLINE)*

Feature Importance Summary: D6 = F_gametes_3way_epistasis_L_3_H_0

Composite Feature Importance Plot (Normalized and Performance Weighted)

Generated with the URBS-Lab STREAMLINE: (https://github.com/UrbsLab/STREAMLINE)
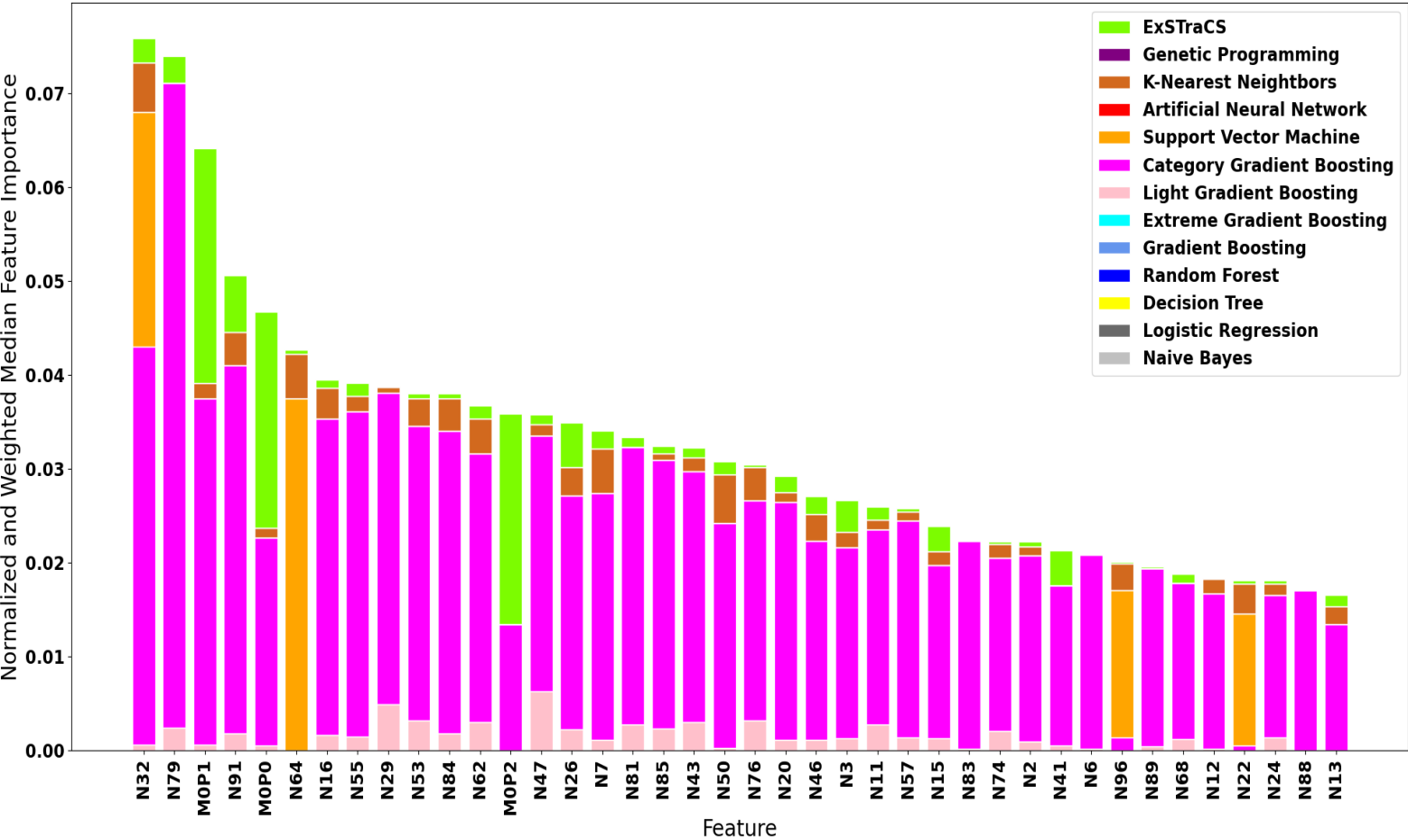
Compare ML Performance Across Datasets
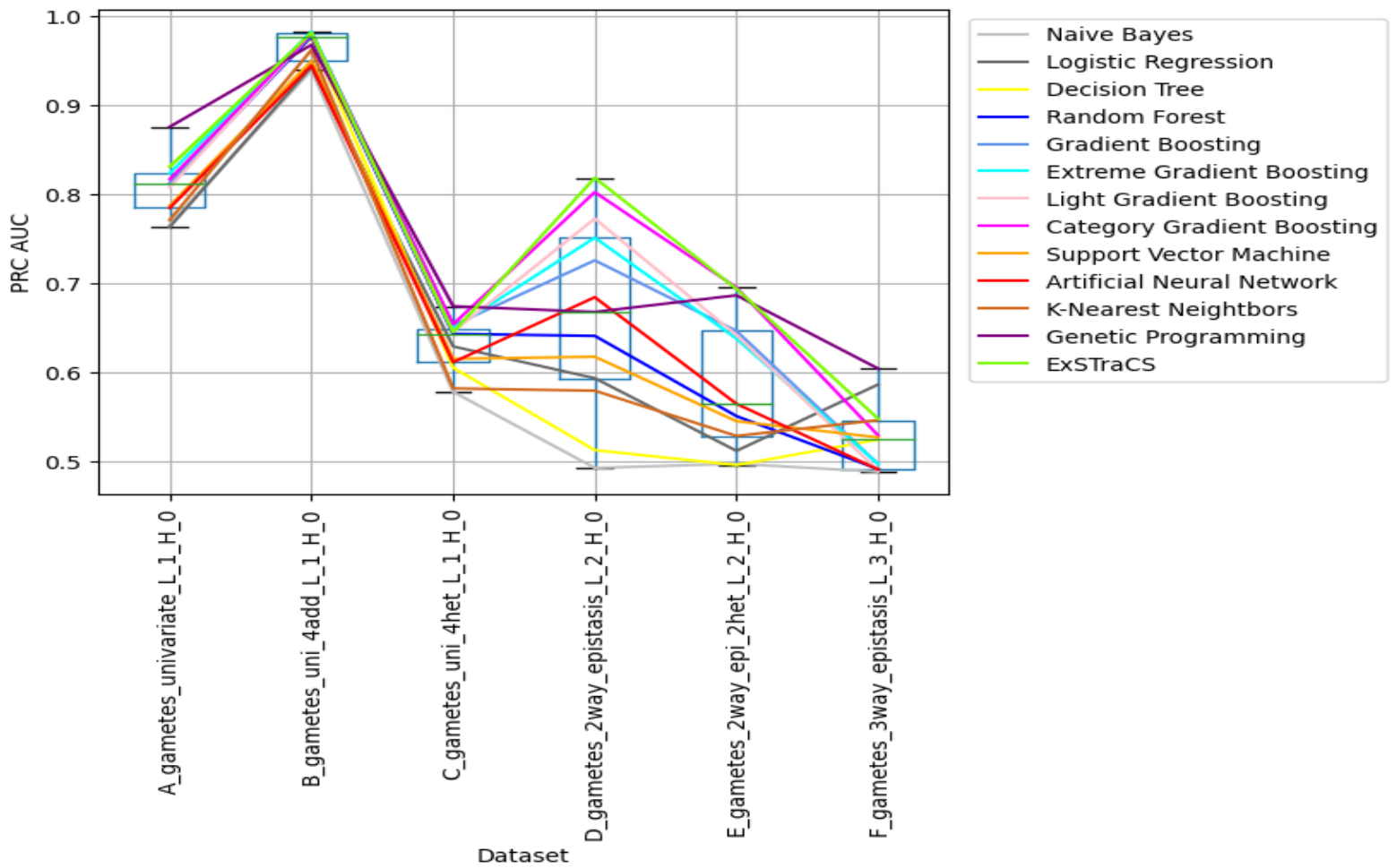
Using Best Performing Algorithms (Kruskall Wallis Compare Datasets)

Datasets:

D1 = A_gametes_univariate_L_1_H_0

D2 = B_gametes_uni_4add_L_1_H_0

D3 = C_gametes_uni_4het_L_1_H_0

D4 = D_gametes_2way_epistasis_L_2_H_0

D5 = E_gametes_2way_epi_2het_L_2_H_0

D6 = F_gametes_3way_epistasis_L_3_H_0

index	P-Value	Best_Alg_D1	Median_D1	Best_Alg_D2	Median_D2	Best_Alg_D3	Median_D3
Balanced Accuracy	0.0	Logistic Regression	0.8344	Extreme Gradient Boosting	0.9219	Random Forest	0.6281
Accuracy	0.0	Logistic Regression	0.8344	Extreme Gradient Boosting	0.9219	Random Forest	0.6281
F1 Score	0.0	Logistic Regression	0.8307	Category Gradient Boosting	0.924	Random Forest	0.634
Sensitivity (Recall)	0.0	Logistic Regression	0.8062	Random Forest	0.9438	Decision Tree	0.6688
Specificity	0.0047	ExSTraCS	0.85	K-Nearest Neightbors	0.9438	K-Nearest Neightbors	0.6938
Precision (PPV)	0.0	ExSTraCS	0.84	K-Nearest Neightbors	0.9391	Category Gradient Boosting	0.6431
TP	0.0	Logistic Regression	64.5	Random Forest	75.5	Decision Tree	53.5
TN	0.0047	ExSTraCS	68.0	K-Nearest Neightbors	75.5	K-Nearest Neightbors	55.5
FP	0.0	Naive Bayes	15.0	Artificial Neural Network	10.5	Decision Tree	36.0
FN	0.0	K-Nearest Neightbors	26.0	Naive Bayes	9.0	K-Nearest Neightbors	45.0
NPV	0.0	Logistic Regression	0.8132	Random Forest	0.9417	Random Forest	0.6359
LR+	0.0	ExSTraCS	5.25	K-Nearest Neightbors	15.675	Category Gradient Boosting	1.8024
LR-	0.0	K-Nearest Neightbors	0.3769	K-Nearest Neightbors	0.1225	K-Nearest Neightbors	0.8518
ROC AUC	0.0	ExSTraCS	0.8482	ExSTraCS	0.983	Extreme Gradient Boosting	0.6724
PRC AUC	0.0	Genetic Programming	0.8698	ExSTraCS	0.9833	Genetic Programming	0.6732
PRC APS	0.0	ExSTraCS	0.8382	ExSTraCS	0.9834	Category Gradient Boosting	0.648

index	P-Value	Best_Alg_D4	Median_D4	Best_Alg_D5	Median_D5	Best_Alg_D6	Median_D6
Balanced Accuracy	0.0	ExSTraCS	0.8062	ExSTraCS	0.6844	Category Gradient Boosting	0.5344
Accuracy	0.0	ExSTraCS	0.8062	ExSTraCS	0.6844	Category Gradient Boosting	0.5344
F1 Score	0.0	ExSTraCS	0.8297	ExSTraCS	0.6886	Support Vector Machine	0.545
Sensitivity (Recall)	0.0	ExSTraCS	0.95	Genetic Programming	0.775	Support Vector Machine	0.5875
Specificity	0.0047	Genetic Programming	0.7625	Category Gradient Boosting	0.6562	Logistic Regression	0.6562
Precision (PPV)	0.0	ExSTraCS	0.7477	ExSTraCS	0.6726	Category Gradient Boosting	0.5332
TP	0.0	ExSTraCS	76.0	Genetic Programming	62.0	Support Vector Machine	47.0
TN	0.0047	Genetic Programming	61.0	Category Gradient Boosting	52.5	Logistic Regression	52.5
FP	0.0	Naive Bayes	40.0	Naive Bayes	43.0	Naive Bayes	41.0
FN	0.0	Logistic Regression	60.5	Decision Tree	46.5	Logistic Regression	56.5
NPV	0.0	ExSTraCS	0.9333	ExSTraCS	0.6912	Category Gradient Boosting	0.5351
LR+	0.0	ExSTraCS	2.9644	ExSTraCS	2.0556	Category Gradient Boosting	1.1434
LR-	0.0	Naive Bayes	1.0754	Decision Tree	1.1481	Naive Bayes	1.085
ROC AUC	0.0	ExSTraCS	0.8567	ExSTraCS	0.732	Category Gradient Boosting	0.5328
PRC AUC	0.0	ExSTraCS	0.8244	Category Gradient Boosting	0.6988	Genetic Programming	0.6061
PRC APS	0.0	ExSTraCS	0.8263	Category Gradient Boosting	0.703	ExSTraCS	0.5593

Pipeline Runtime Summary

A_gametes_univariate_L_1_H_0

Pipeline Component	Time (sec)
Exploratory Analysis	3.56
Preprocessing	0.14
Mutual Information	4.82
MultiSURF	741.99
Feature Selection	1.37
Naive Bayes	10.49
Logistic Regression	71.08
Decision Tree	70.74
Random Forest	4599.85
Gradient Boosting	8725.28
Extreme Gradient Boosting	9387.48
Light Gradient Boosting	515.69
Category Gradient Boosting	10465.59
Support Vector Machine	150590.89
Artificial Neural Network	4456.82
K-Nearest Neightbors	1032.39
Genetic Programming	13490.22
ExSTraCS	22008.79
Stats Summary	23.19

B_gametes_uni_4add_L_1_H_0

Pipeline Component	Time (sec)
Exploratory Analysis	3.53
Preprocessing	0.13
Mutual Information	4.98
MultiSURF	753.99
Feature Selection	1.38
Naive Bayes	9.37
Logistic Regression	75.28
Decision Tree	69.73
Random Forest	5310.54
Gradient Boosting	7656.26
Extreme Gradient Boosting	8921.79
Light Gradient Boosting	764.84
Category Gradient Boosting	10330.49
Support Vector Machine	8812.14
Artificial Neural Network	3113.32
K-Nearest Neightbors	936.97
Genetic Programming	12827.64
ExSTraCS	18178.19
Stats Summary	22.8

C_gametes_uni_4het_L_1_H_0

Pipeline Component	Time (sec)
Exploratory Analysis	3.59
Preprocessing	0.13
Mutual Information	4.83
MultiSURF	750.97
Feature Selection	1.44
Naive Bayes	11.66
Logistic Regression	86.2
Decision Tree	72.05
Random Forest	4926.09
Gradient Boosting	6673.99
Extreme Gradient Boosting	8106.25
Light Gradient Boosting	478.65
Category Gradient Boosting	10051.79
Support Vector Machine	16497.09
Artificial Neural Network	4220.77
K-Nearest Neightbors	1173.02
Genetic Programming	12738.81
ExSTraCS	23637.57
Stats Summary	23.11

D_gametes_2way_epistasis_L_2_H_0

Pipeline Component	Time (sec)
Exploratory Analysis	3.57
Preprocessing	0.13
Mutual Information	4.92
MultiSURF	744.19
Feature Selection	1.48
Naive Bayes	12.4
Logistic Regression	81.62
Decision Tree	74.78
Random Forest	9801.74
Gradient Boosting	9244.13
Extreme Gradient Boosting	9321.31
Light Gradient Boosting	1767.01
Category Gradient Boosting	10382.96
Support Vector Machine	10358.17
Artificial Neural Network	4985.21
K-Nearest Neightbors	1199.77
Genetic Programming	12804.02
ExSTraCS	20575.81
Stats Summary	22.73

E_gametes_2way_epi_2het_L_2_H_0

Pipeline Component	Time (sec)
Exploratory Analysis	3.53
Preprocessing	0.13
Mutual Information	4.9
MultiSURF	755.69
Feature Selection	1.56
Naive Bayes	14.19
Logistic Regression	102.99
Decision Tree	84.57
Random Forest	9643.53
Gradient Boosting	9136.99
Extreme Gradient Boosting	9287.18
Light Gradient Boosting	1979.71
Category Gradient Boosting	10367.47
Support Vector Machine	13328.72
Artificial Neural Network	5208.84
K-Nearest Neightbors	1540.68
Genetic Programming	13292.4
ExSTraCS	23669.42
Stats Summary	23.83

F_gametes_3way_epistasis_L_3_H_0

Pipeline Component	Time (sec)
Exploratory Analysis	3.5
Preprocessing	0.13
Mutual Information	4.83
MultiSURF	746.31
Feature Selection	1.5
Naive Bayes	13.91
Logistic Regression	94.23
Decision Tree	77.73
Random Forest	5913.89
Gradient Boosting	9364.37
Extreme Gradient Boosting	8380.31
Light Gradient Boosting	2402.52
Category Gradient Boosting	10220.63
Support Vector Machine	12244.93
Artificial Neural Network	4178.95
K-Nearest Neightbors	1102.72
Genetic Programming	12390.55
ExSTraCS	25087.16
Stats Summary	23.65