STREAMLINE Training Summary Report: 2022-05-20 14:04:40.239187

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: 1.11064715131114122e-164
M2P0: 1.18667349514186558e-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

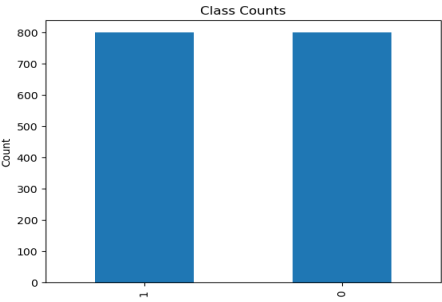| Feature: P-Value |
| --- |
| N8: 0.0022378710091608 |
| N10: 0.00332823354218843 |
| N1: 0.023956678010501 |
| N90: 0.0409688621078246 |
| N16: 0.0475567352335571 |
| N36: 0.0483697186665984 |
| N45: 0.05076830219 |
| N49: 0.0536871730192233 |
| N9: 0.0598766172399342 |
| N61: 0.076713380936202 |

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

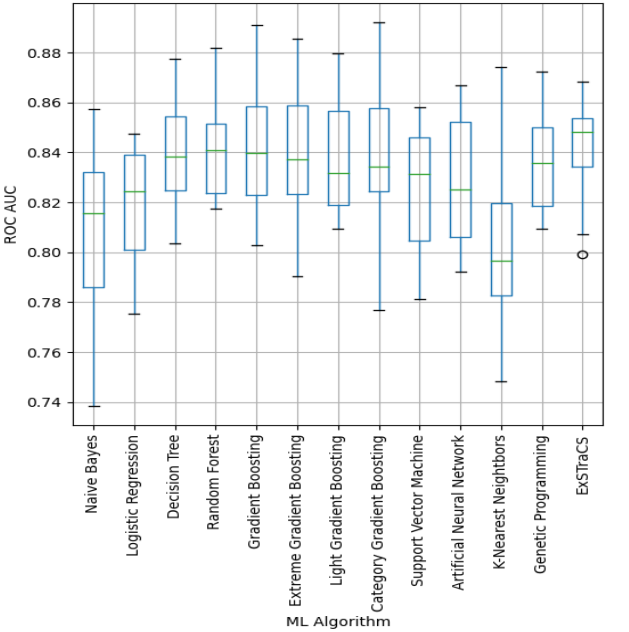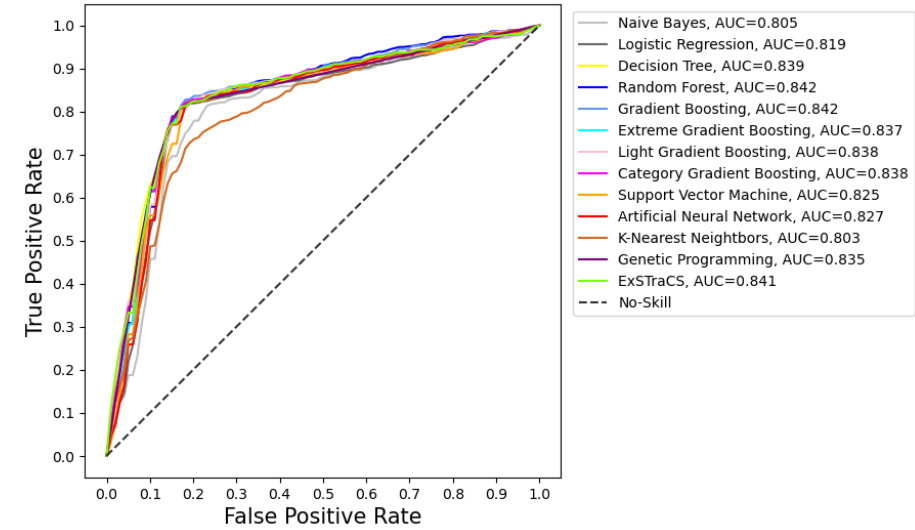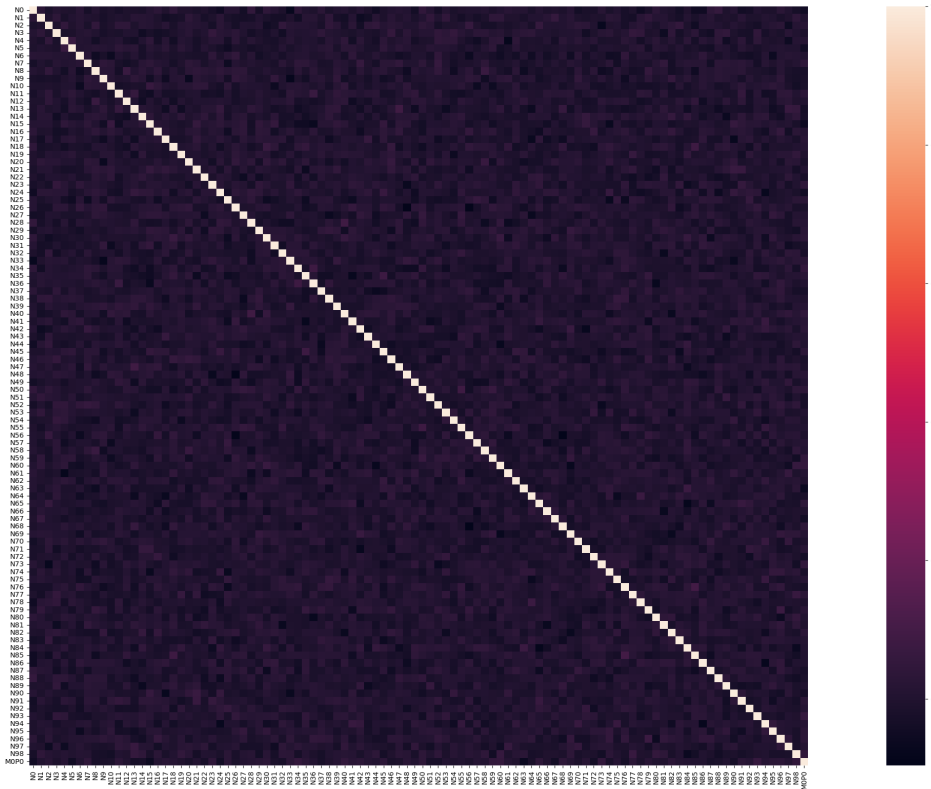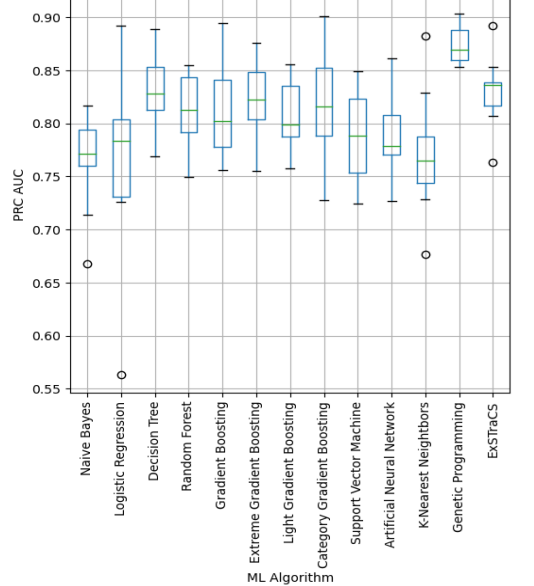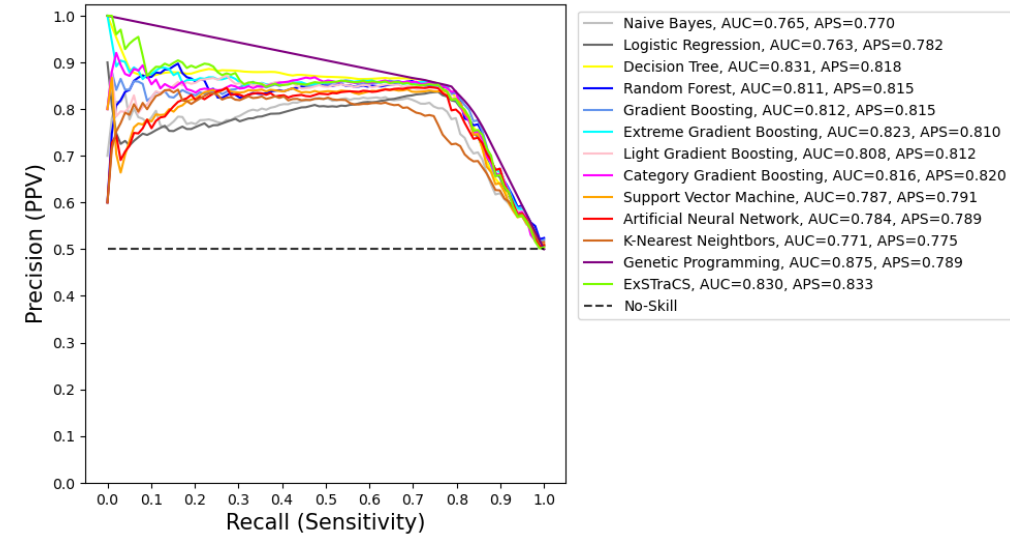# Dataset and Model Prediction Summary: D1 = A_gametes_univariate_L_1_H_0

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):

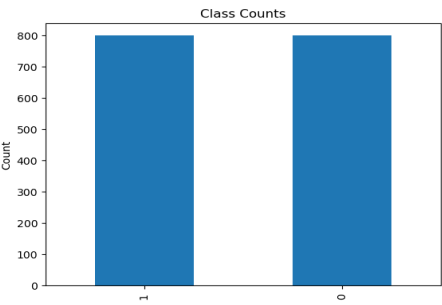| |
|---|
| Best (ROC_AUC): Random Forest = 0.842 |
| Best (Balanced Acc.): ExSTraCS = 0.830 |
| Best (F1 Score): ExSTraCS = 0.827 |
| Best (PRC AUC): Genetic Programming = 0.875 |
| Best (PRC APS): ExSTraCS = 0.833 |

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

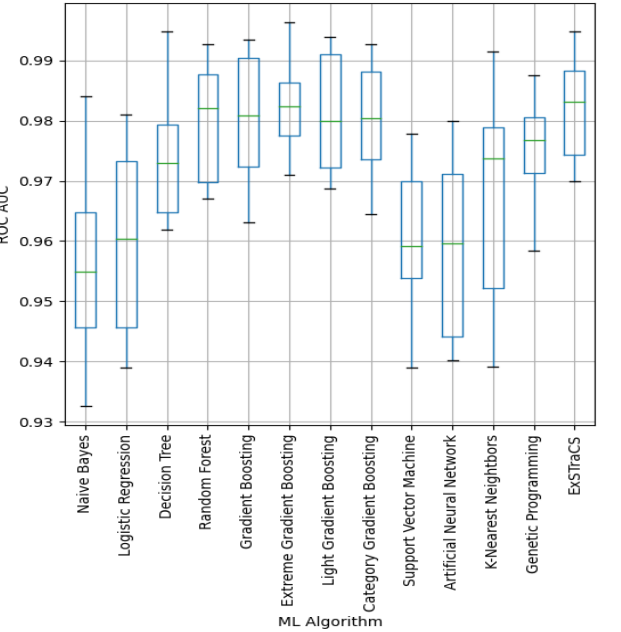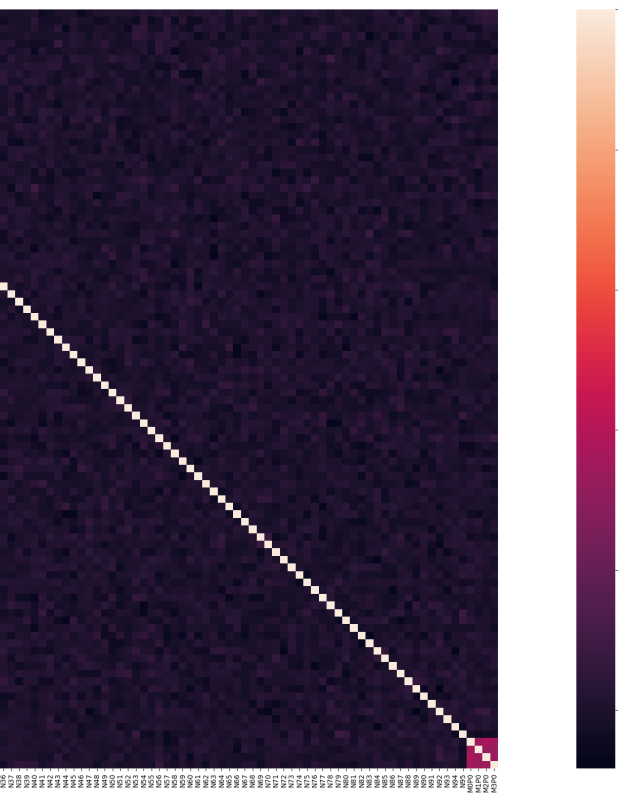*Generated with the URBS-Lab STREAMLINE: (https://github.com/UrbsLab/STREAMLINE)*

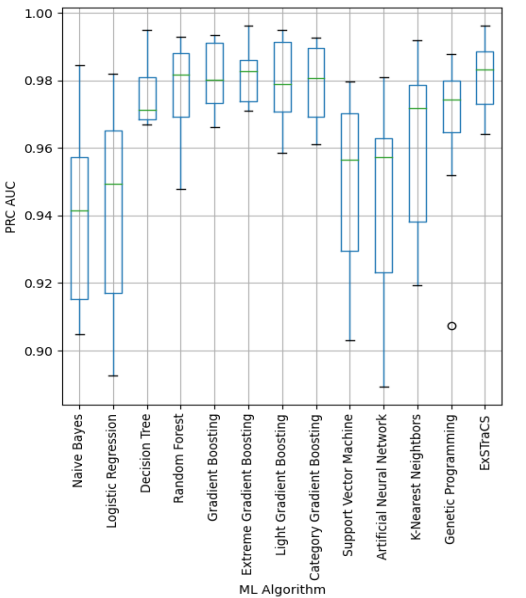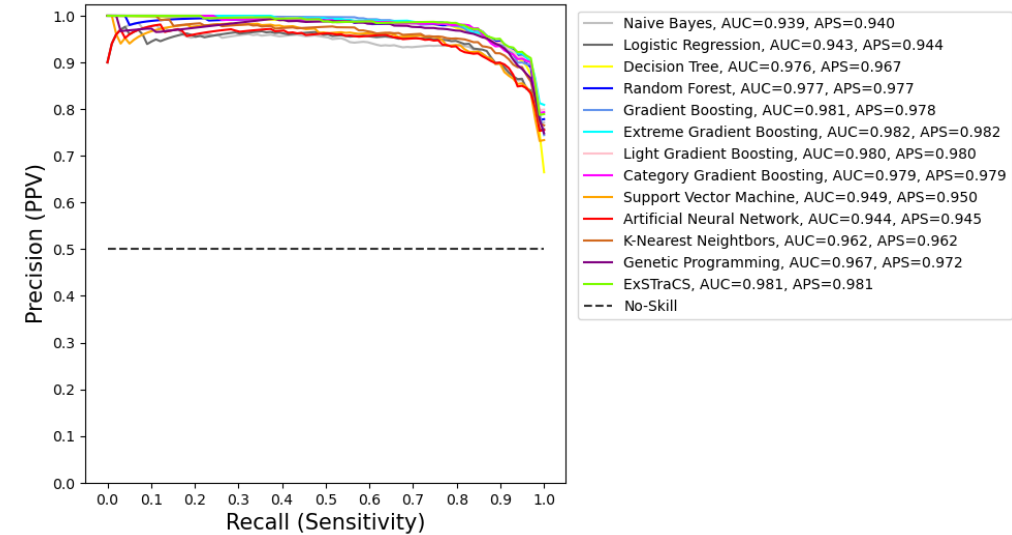# Dataset and Model Prediction Summary: D2 = B_gametes_uni_4add_L_1_H_0

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 = 0.927

Best (F1 Score): Extreme Gradient Boosting = 0.928

Best (PRC AUC): Extreme Gradient Boosting = 0.982
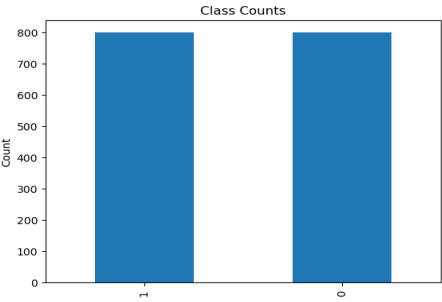
Best (PRC APS): Extreme Gradient Boosting = 0.982

ROC

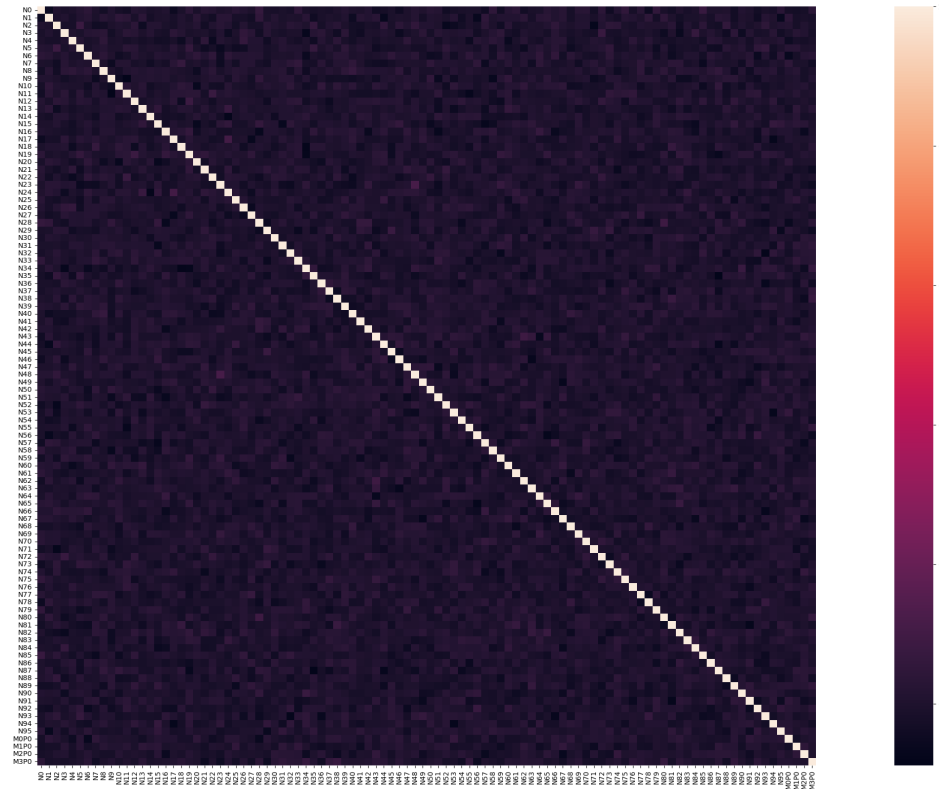- Naive Bayes, AUC=0.956
- Logistic Regression, AUC=0.960
- Decision Tree, AUC=0.974
- Random Forest, AUC=0.980
- Gradient Boosting, AUC=0.980
- Extreme Gradient Boosting, AUC=0.983
- Light Gradient Boosting, AUC=0.981
- Category Gradient Boosting, AUC=0.980
- Support Vector Machine, AUC=0.960
- Artificial Neural Network, AUC=0.959
- K-Nearest Neighbtors, 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=0.979, APS=0.979
- Support Vector Machine, AUC=0.949, APS=0.950
- Artificial Neural Network, AUC=0.944, APS=0.945
- K-Nearest Neighbtors, 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

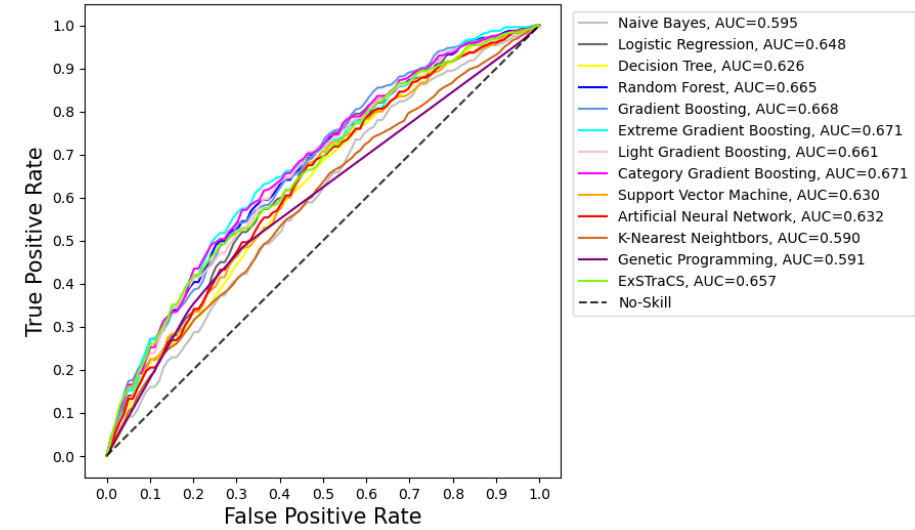**Dataset Counts Summary:**

instances: 1600.0

features: 100.0

categorical_features: 100.0

quantitative_features: 0.0

missing_values: 0.0

missing_percent: 0.0

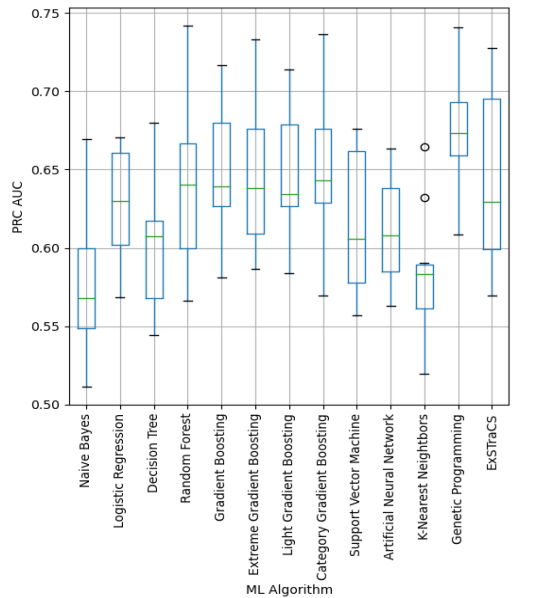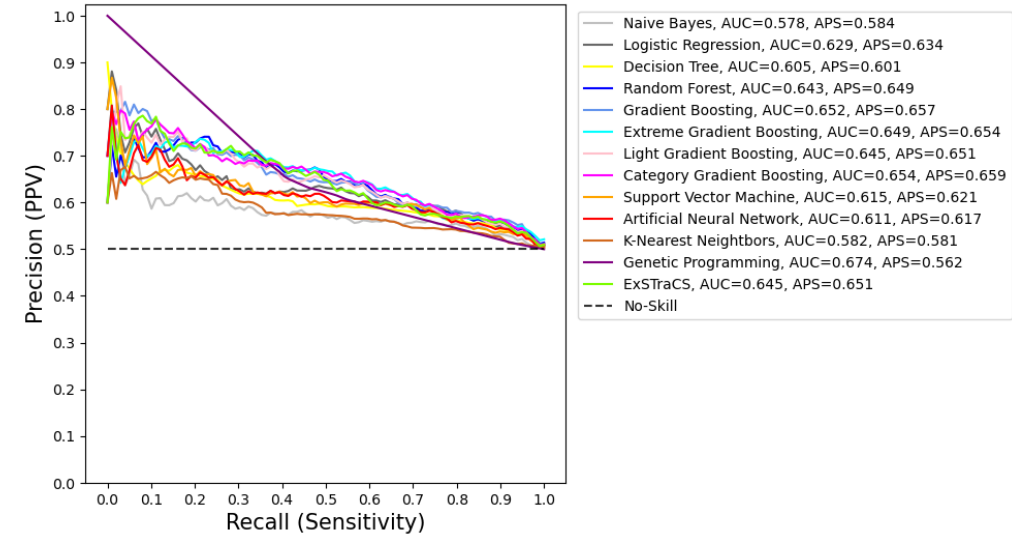**Top ML Algorithm Results (Averaged Over CV Runs):**

Best (ROC_AUC): Extreme Gradient Boosting = 0.671

Best (Balanced Acc.): Category Gradient Boosting = 0.630

Best (F1 Score): Category Gradient Boosting = 0.621

Best (PRC AUC): Genetic Programming = 0.674

Best (PRC APS): Category Gradient Boosting = 0.659

ROC
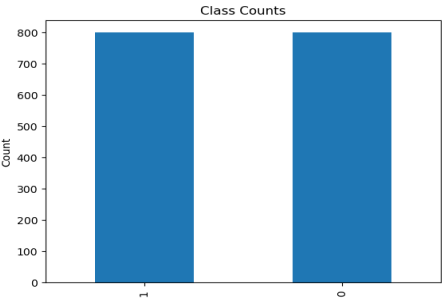

Naive Bayes, AUC=0.595
Logistic Regression, AUC=0.648
Decision Tree, AUC=0.626
Random Forest, AUC=0.665
Gradient Boosting, AUC=0.668
Extreme Gradient Boosting, AUC=0.671
Light Gradient Boosting, AUC=0.661
Category Gradient Boosting, AUC=0.671
Support Vector Machine, AUC=0.630
Artificial Neural Network, 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

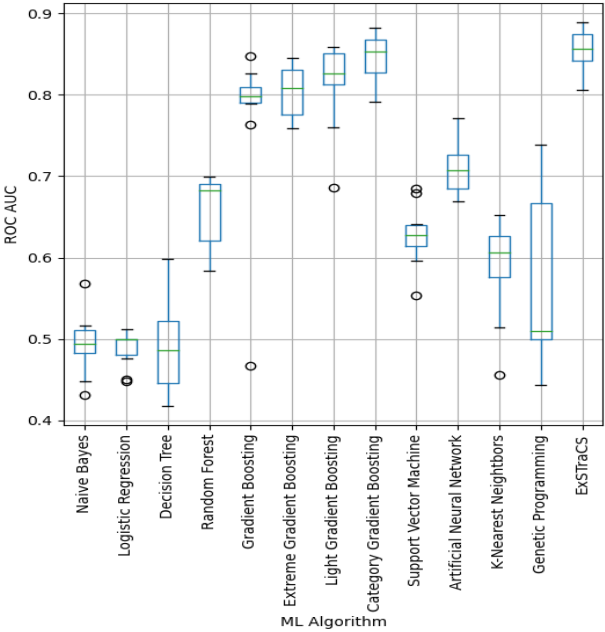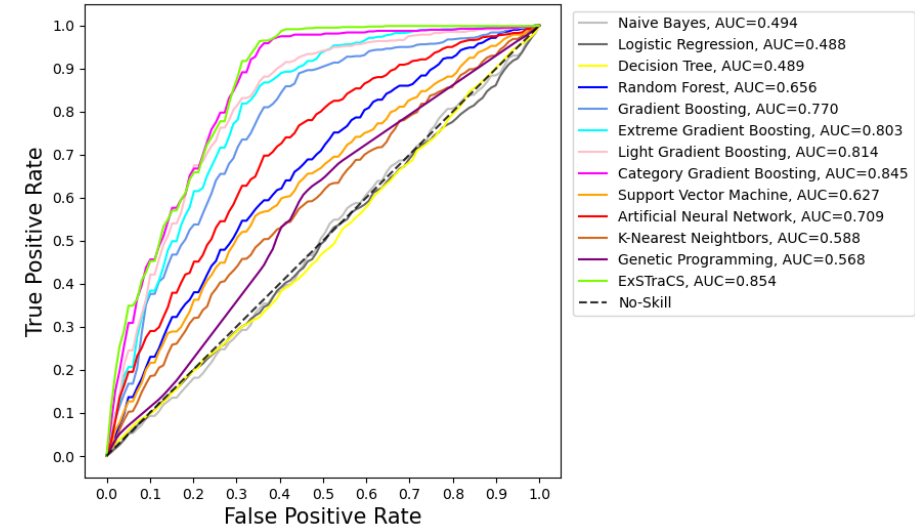*Generated with the URBS-Lab STREAMLINE: (https://github.com/UrbsLab/STREAMLINE)*

Dataset and Model Prediction Summary: D4 = D_gametes_2way_epistasis_L_2_H_0

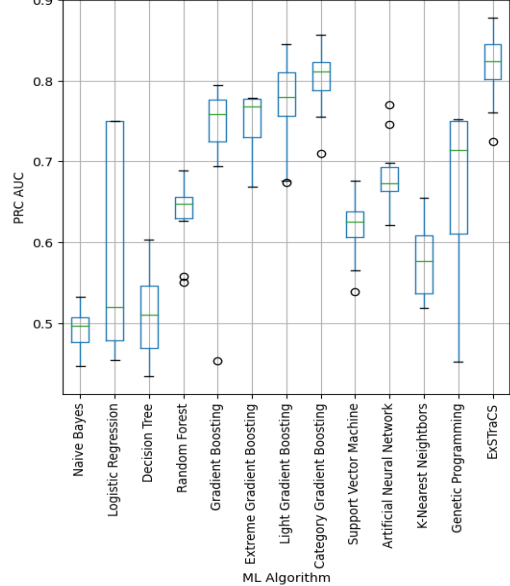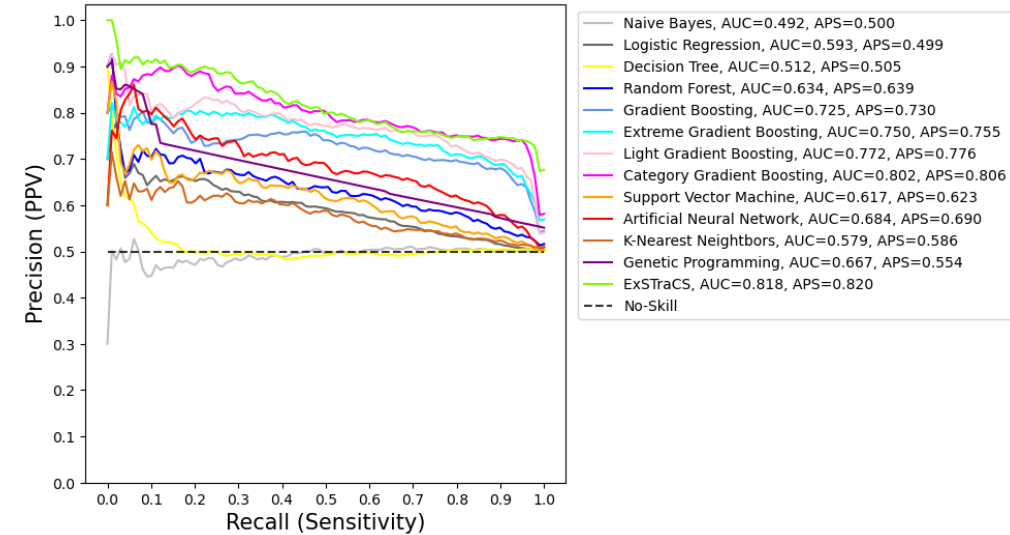### Class Counts

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):

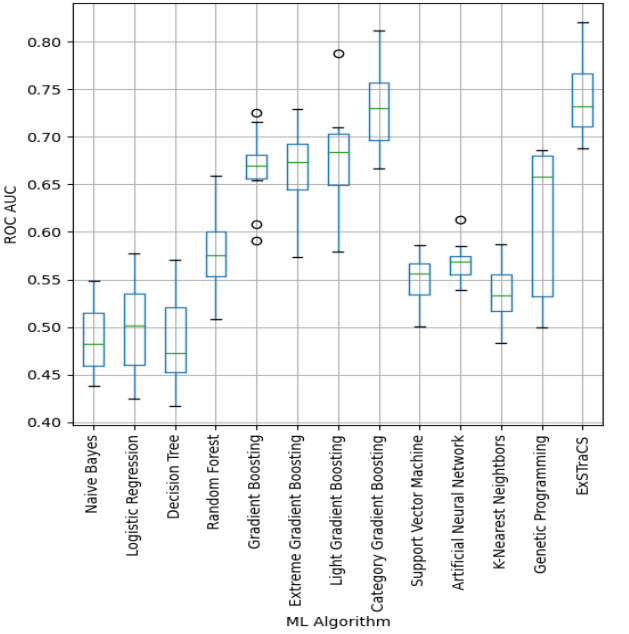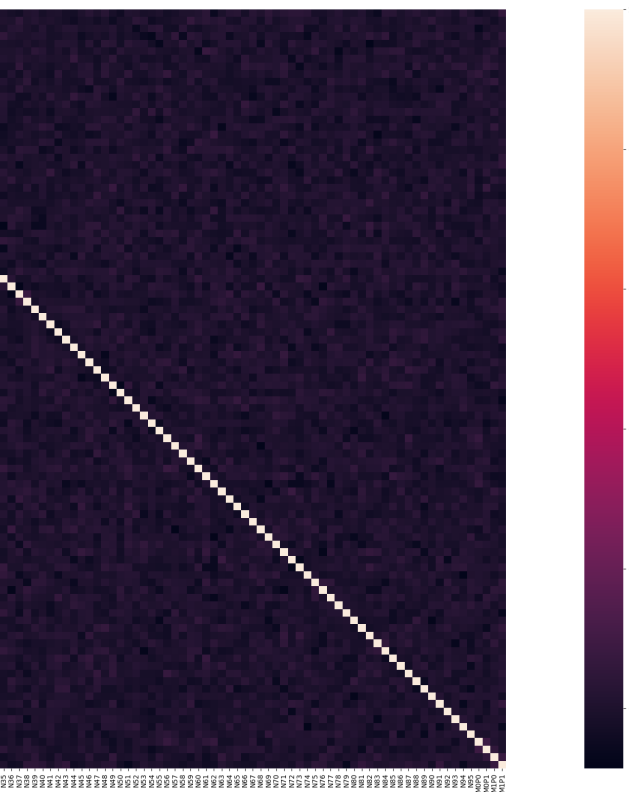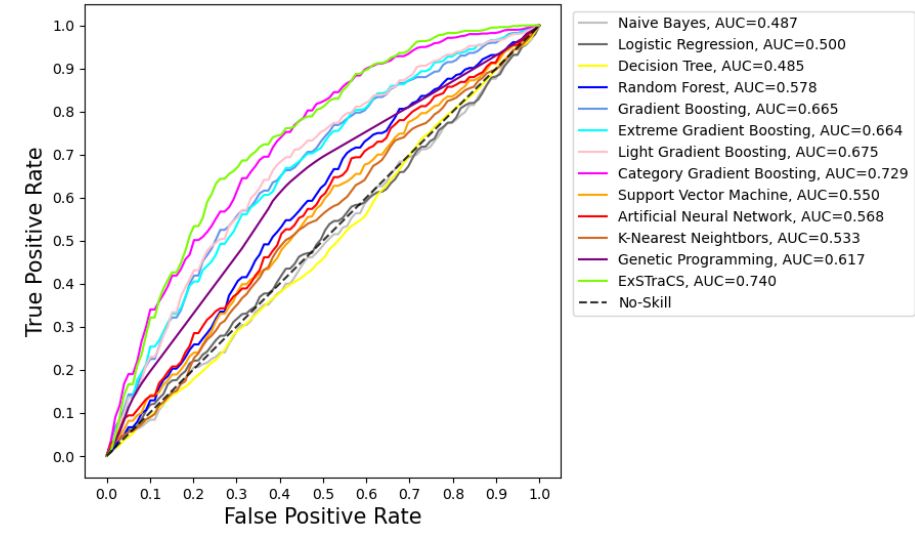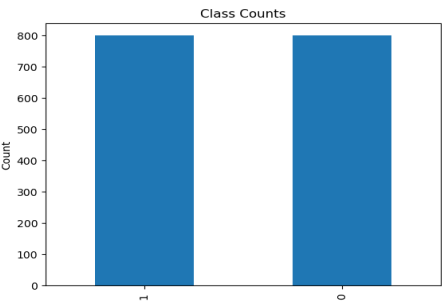| |
|---|
| Best (ROC_AUC): ExSTraCS = 0.854 |
| Best (Balanced Acc.): ExSTraCS = 0.802 |
| Best (F1 Score): ExSTraCS = 0.825 |
| Best (PRC AUC): ExSTraCS = 0.818 |
| Best (PRC APS): ExSTraCS = 0.820 |

ROC

Naive Bayes, AUC=0.494
Logistic Regression, AUC=0.488
Decision Tree, AUC=0.489
Random Forest, AUC=0.656
Gradient Boosting, AUC=0.770
Extreme Gradient Boosting, AUC=0.803
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=0.634, APS=0.639
Gradient Boosting, AUC=0.725, APS=0.730
Extreme Gradient Boosting, AUC=0.750, APS=0.755
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

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

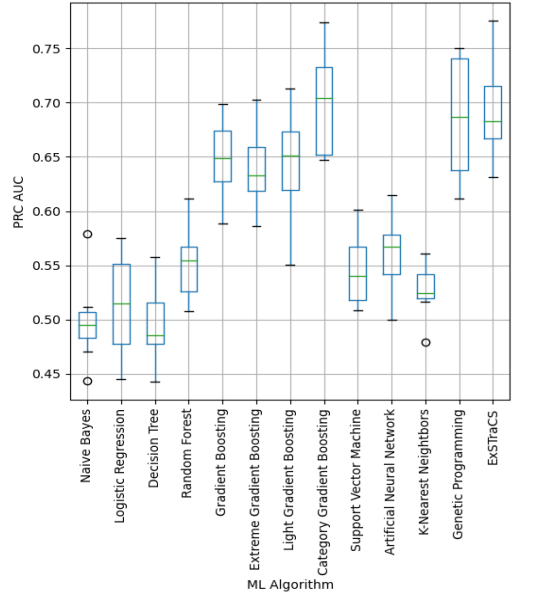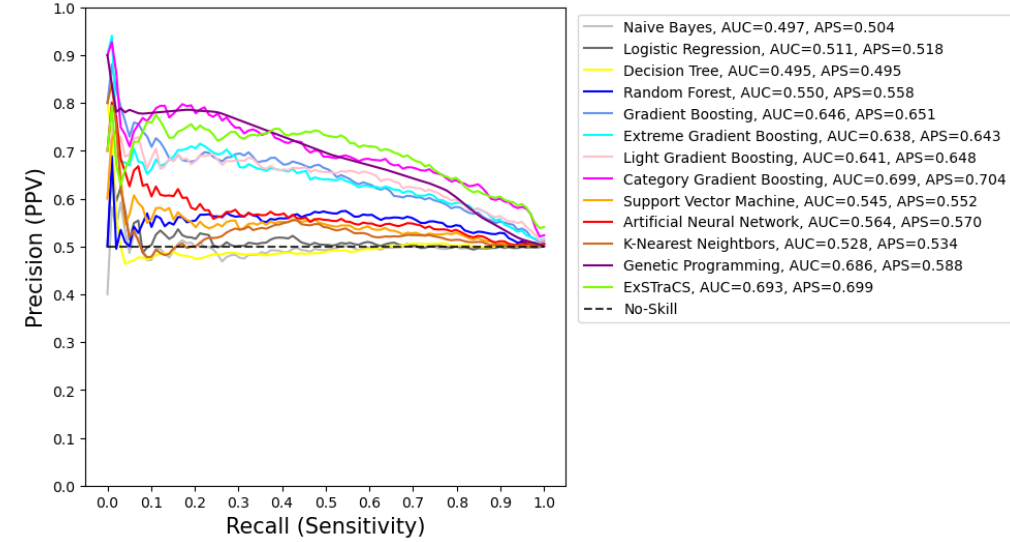# Dataset and Model Prediction Summary: D5 = E_gametes_2way_epi_2het_L_2_H_0

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):

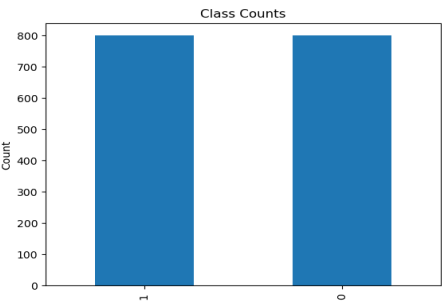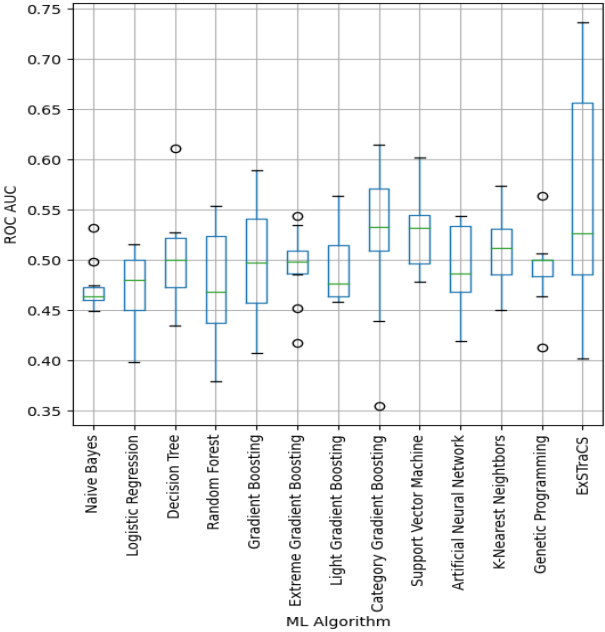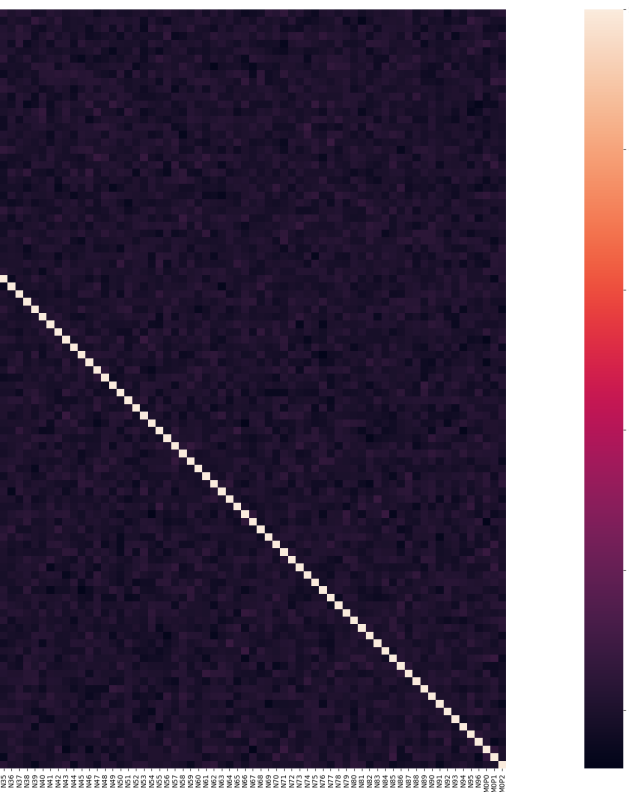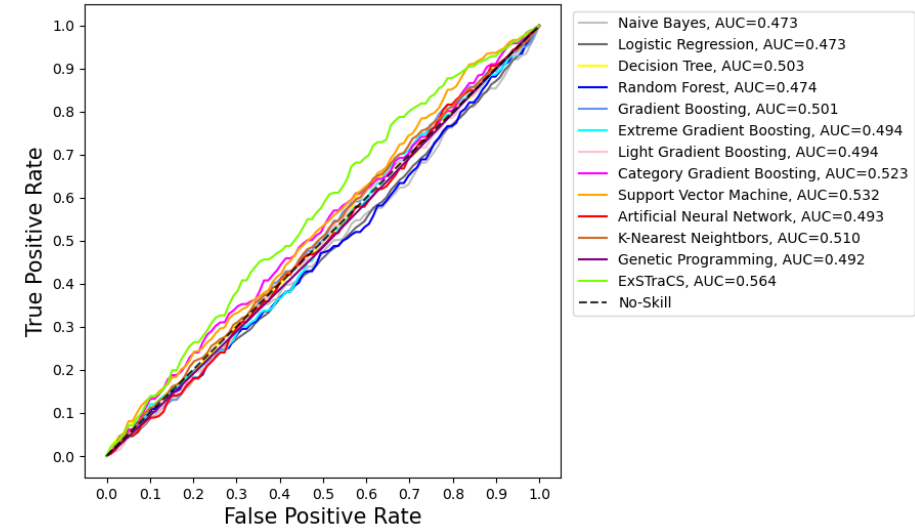| |
|---|
| Best (ROC_AUC): ExSTraCS = 0.740 |
| Best (Balanced Acc.): ExSTraCS = 0.688 |
| Best (F1 Score): ExSTraCS = 0.696 |
| Best (PRC AUC): Category Gradient Boosting = 0.699 |
| Best (PRC APS): Category Gradient Boosting = 0.704 |

ROC

- Naive Bayes, AUC=0.487
- Logistic Regression, AUC=0.500
- Decision Tree, AUC=0.485
- Random Forest, AUC=0.578
- Gradient Boosting, AUC=0.665
- Extreme Gradient Boosting, AUC=0.664
- Light Gradient Boosting, AUC=0.675
- Category Gradient Boosting, AUC=0.729
- 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=0.699, APS=0.704
- 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)

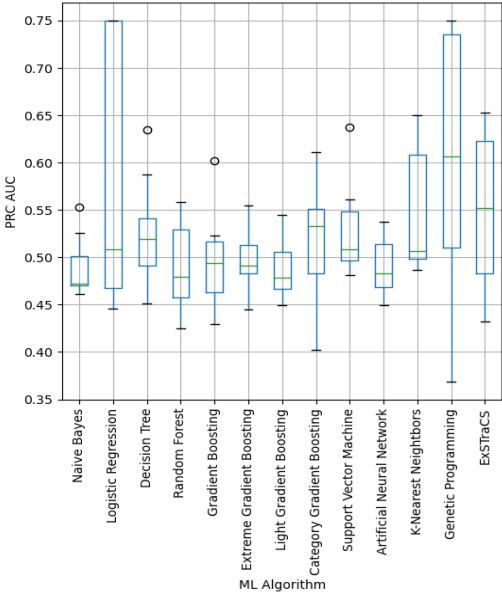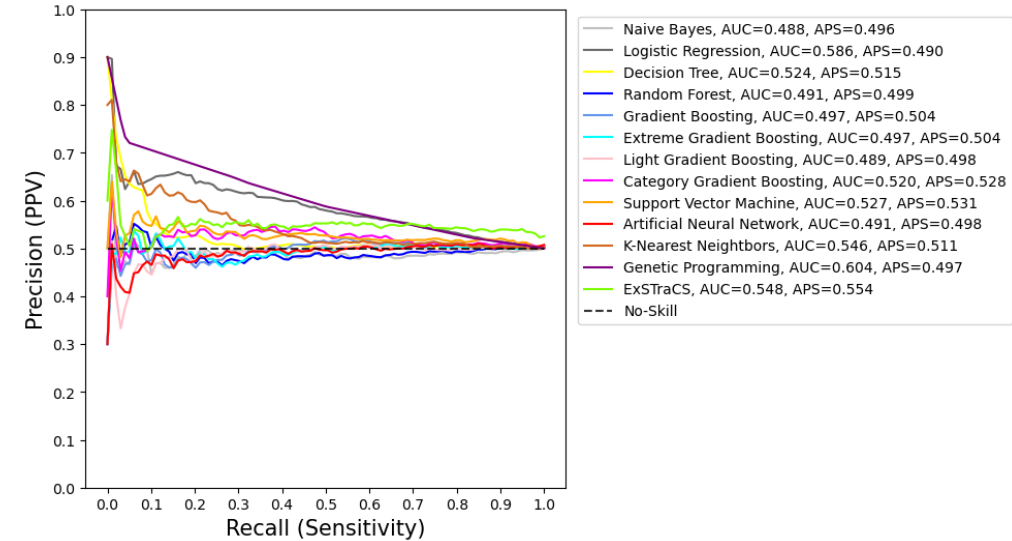# Dataset and Model Prediction Summary: D6 = F_gametes_3way_epistasis_L_3_H_0

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

Legend:
- 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=0.523
- 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

Legend:
- 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=0.520, APS=0.528
- 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

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	0.928	0.928	0.928	0.934	0.921	0.923	74.7	73.7	6.3	5.3	0.933	18.387	0.072	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	0.917	0.917	0.918	0.925	0.909	0.912	74.0	72.7	7.3	6.0	0.925	12.416	0.082	0.98	0.979	0.979
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	0.63	0.63	0.621	0.609	0.651	0.637	48.7	52.1	27.9	31.3	0.626	1.788	0.602	0.671	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

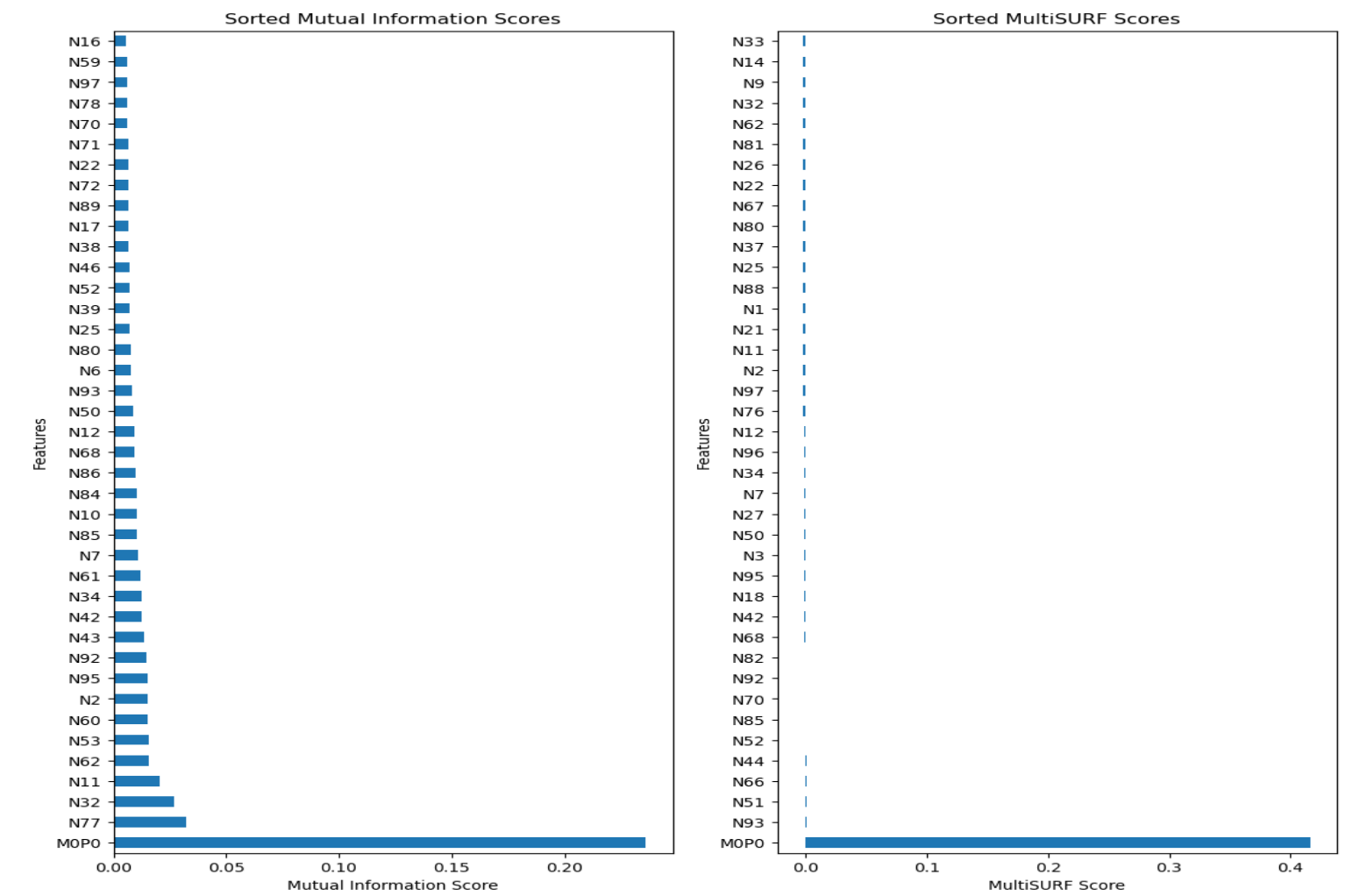| ML Algorithm	Balanced Accuracy	Accuracy	F1 Score	Sensitivity (Recall)	Specificity	Precision (PPV)	TP	TN	FP	FN	NPV	LR+	LR-	ROC AUC	PRC AUC	PRC APS
Naive Bayes	0.493	0.493	0.495	0.499	0.488	0.493	39.9	39.0	41.0	40.1	0.493	0.979	1.036	0.494	0.492	0.5
Logistic Regression	0.494	0.494	0.253	0.252	0.735	0.287	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	0.611	0.611	0.607	0.601	0.621	0.613	48.1	49.7	30.3	31.9	0.61	1.616	0.647	0.656	0.634	0.639
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	0.734	0.734	0.738	0.751	0.716	0.726	60.1	57.3	22.7	19.9	0.744	2.686	0.348	0.803	0.75	0.755
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	0.803	0.874	0.698	0.743	69.9	55.8	24.2	10.1	0.848	2.93	0.181	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

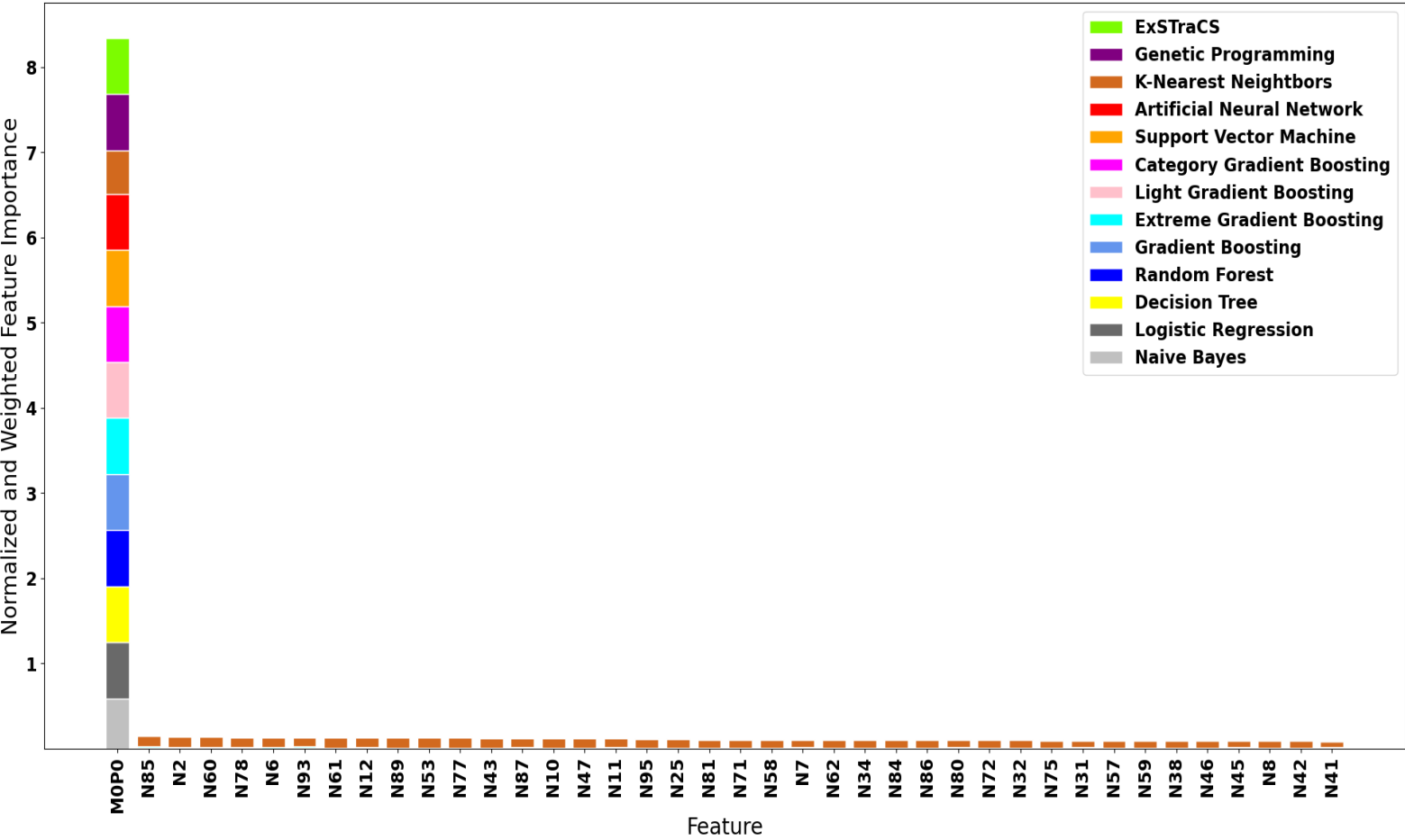| ML Algorithm	Balanced Accuracy	Accuracy	F1 Score	Sensitivity (Recall)	Specificity	Precision (PPV)	TP	TN	FP	FN	NPV	LR+	LR-	ROC AUC	PRC AUC	PRC APS
Naive Bayes	0.491	0.491	0.493	0.5	0.482	0.49	40.0	38.6	41.4	40.0	0.493	0.969	1.039	0.487	0.497	0.504
Logistic Regression	0.507	0.507	0.494	0.489	0.525	0.506	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	0.564	0.57	0.551	0.559	45.6	44.1	35.9	34.4	0.563	1.286	0.785	0.578	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	0.667	0.667	0.668	0.67	0.664	0.668	53.6	53.1	26.9	26.4	0.667	2.096	0.505	0.729	0.699	0.704
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

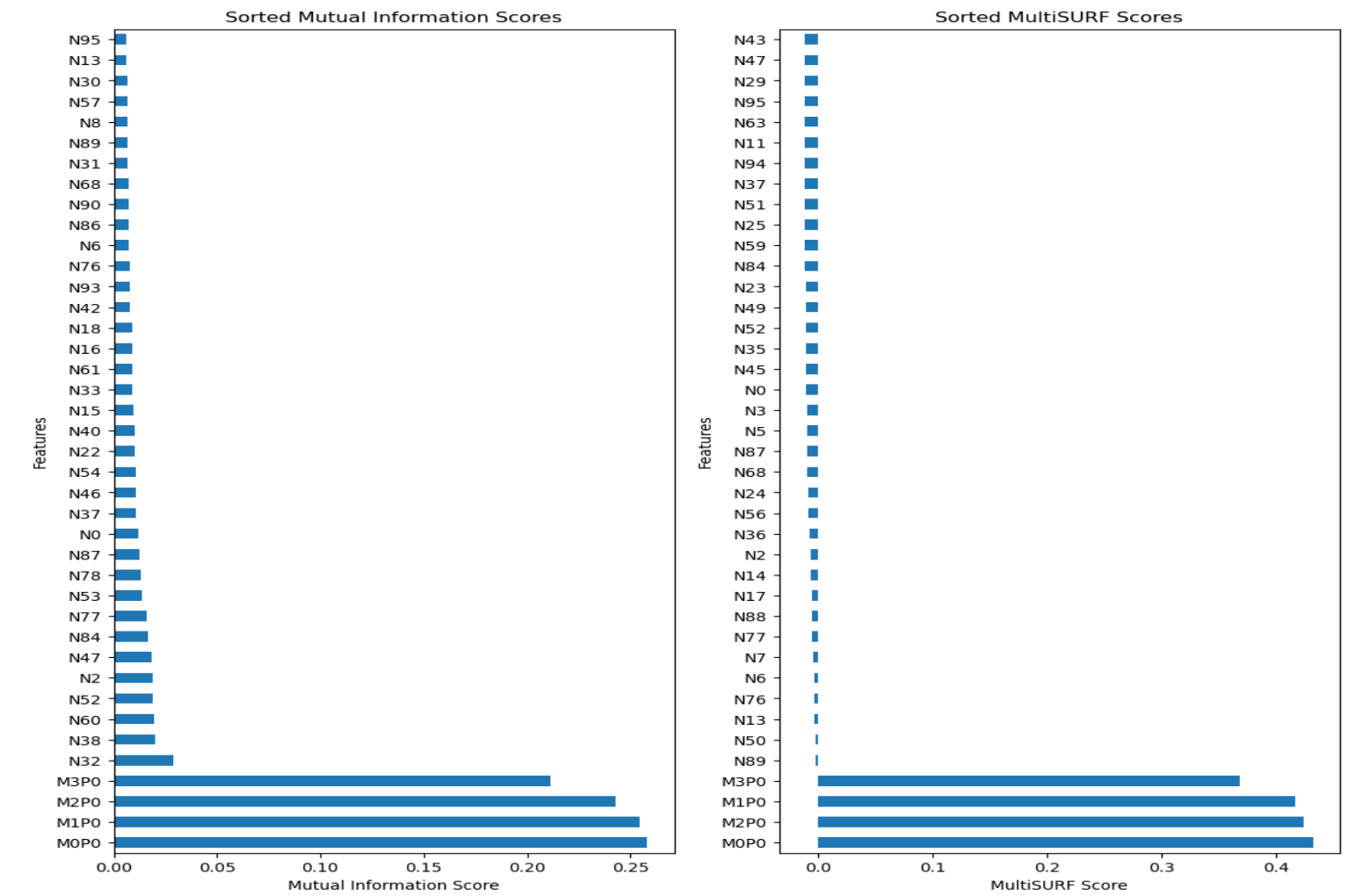| ML Algorithm	Balanced Accuracy	Accuracy	F1 Score	Sensitivity (Recall)	Specificity	Precision (PPV)	TP	TN	FP	FN	NPV	LR+	LR-	ROC AUC	PRC AUC	PRC APS
Naive Bayes	0.478	0.478	0.467	0.461	0.495	0.476	36.9	39.6	40.4	43.1	0.48	0.911	1.087	0.473	0.488	0.496
Logistic Regression	0.483	0.483	0.26	0.25	0.716	0.28	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	0.518	0.518	0.517	0.518	0.519	0.517	41.4	41.5	38.5	38.6	0.519	1.104	0.954	0.523	0.52	0.528
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

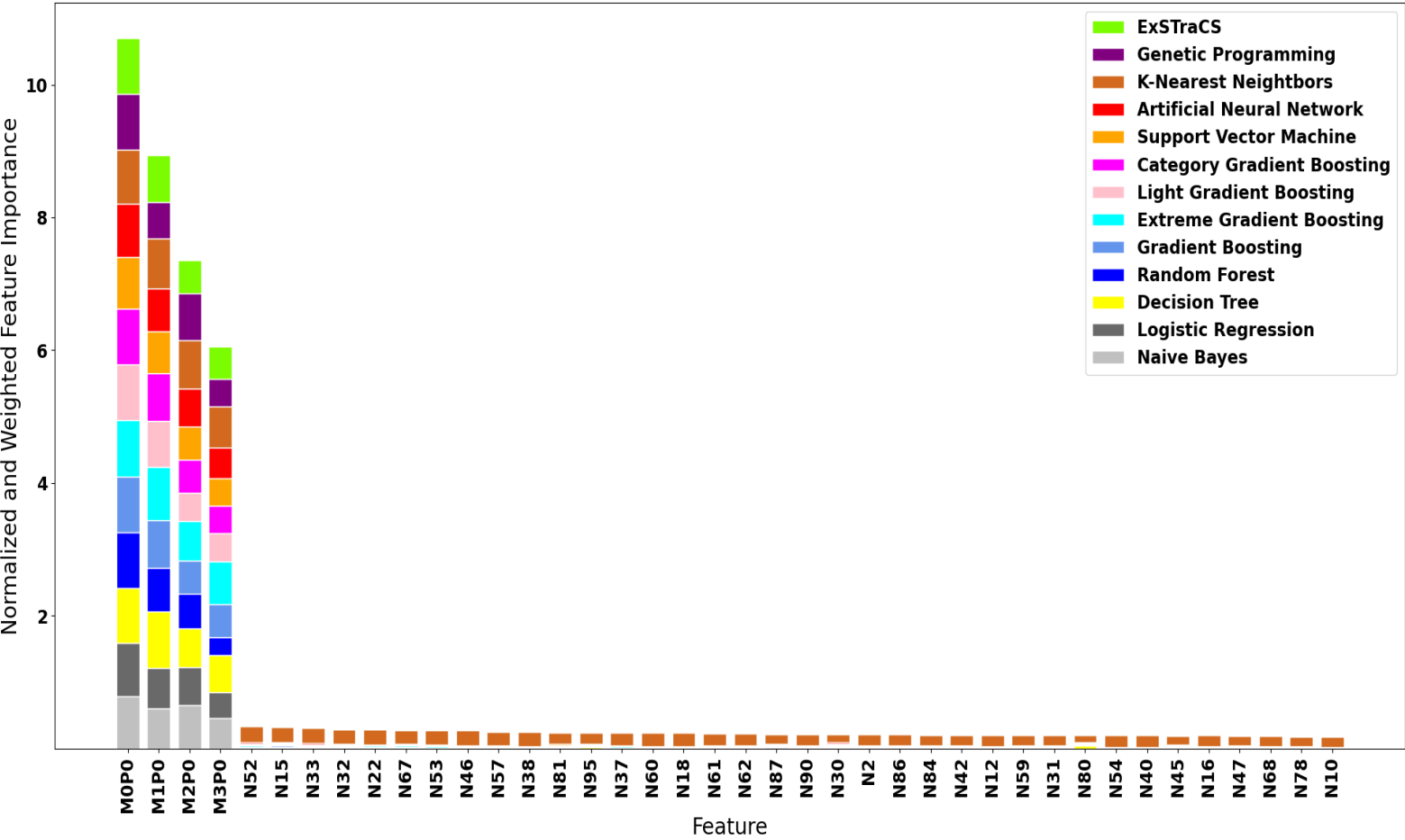# Feature Importance Summary: D1 = A_gametes_univariate_L_1_H_0

Sorted Mutual Information Scores

Sorted MultiSURF Scores

Composite Feature Importance Plot (Normalized and Performance Weighted)

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

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

Feature Importance Summary: D2 = B_gametes_uni_4add_L_1_H_0

Sorted Mutual Information Scores

Sorted MultiSURF Scores

Composite Feature Importance Plot (Normalized and Performance Weighted)

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

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

Feature Importance Summary: D3 = C_gametes_uni_4het_L_1_H_0

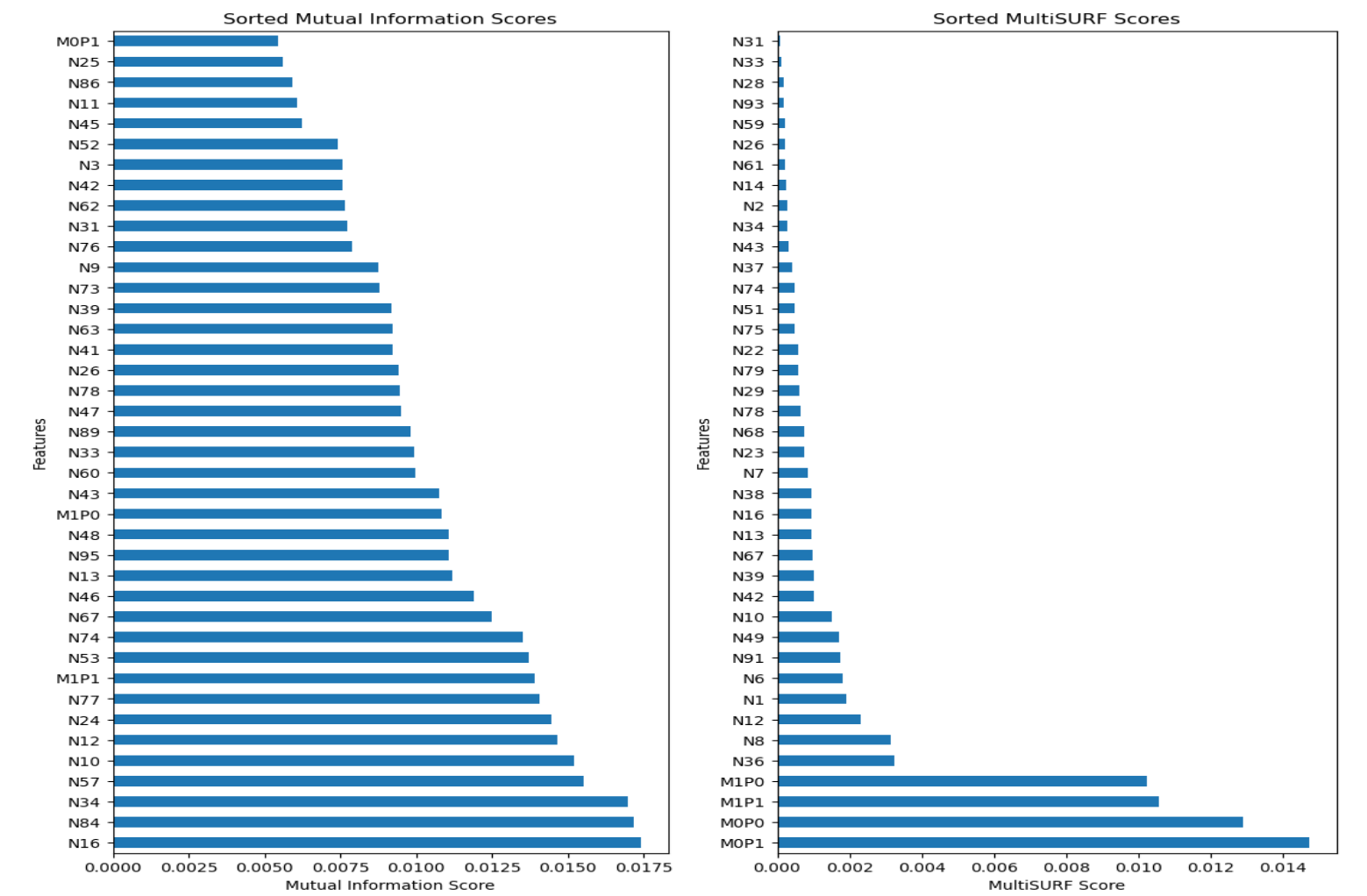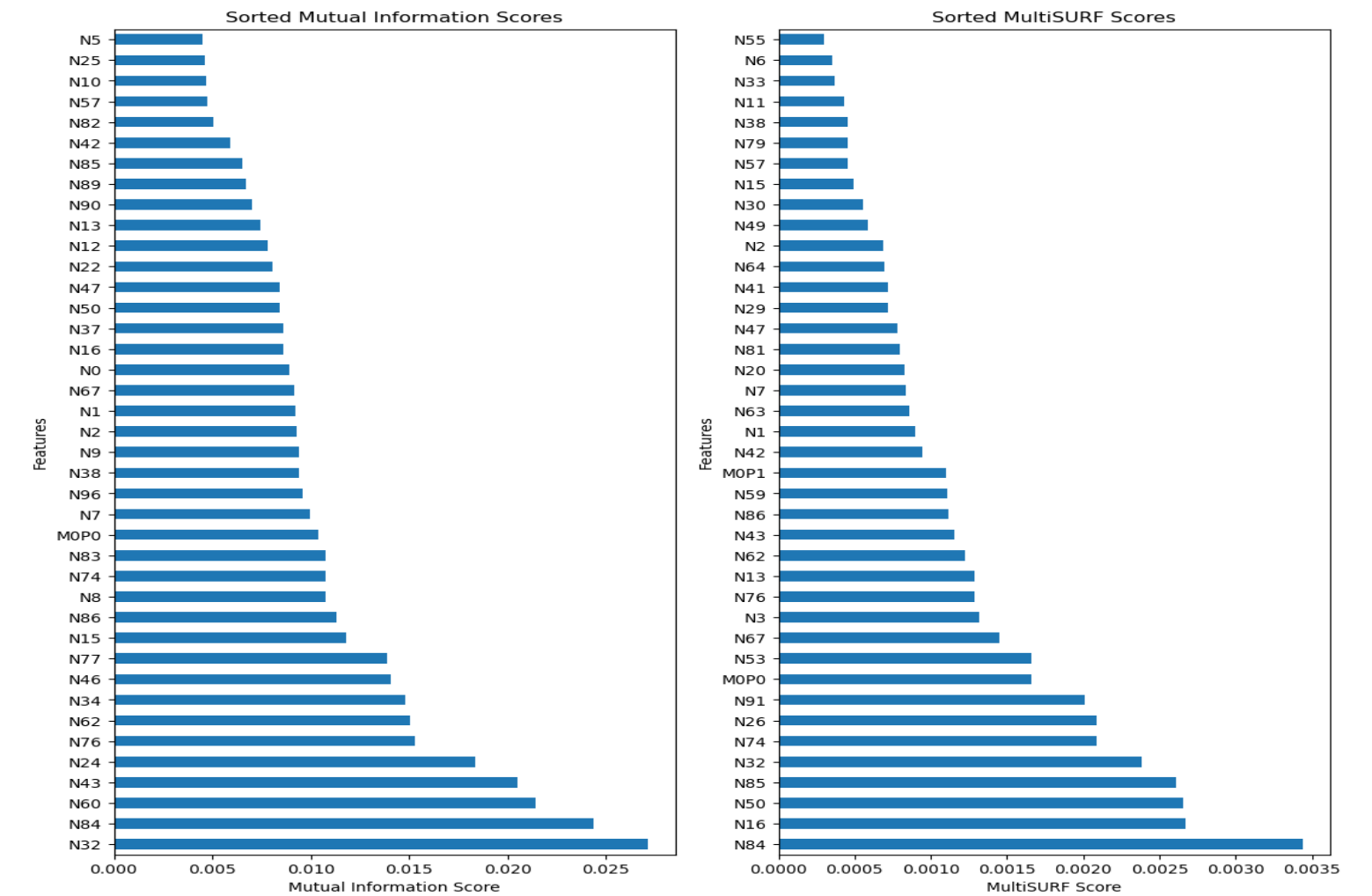## Composite Feature Importance Plot (Normalized and Performance Weighted)

Generated with the URBS-Lab STREAMLINE: (https://github.com/UrbsLab/STREAMLINE)
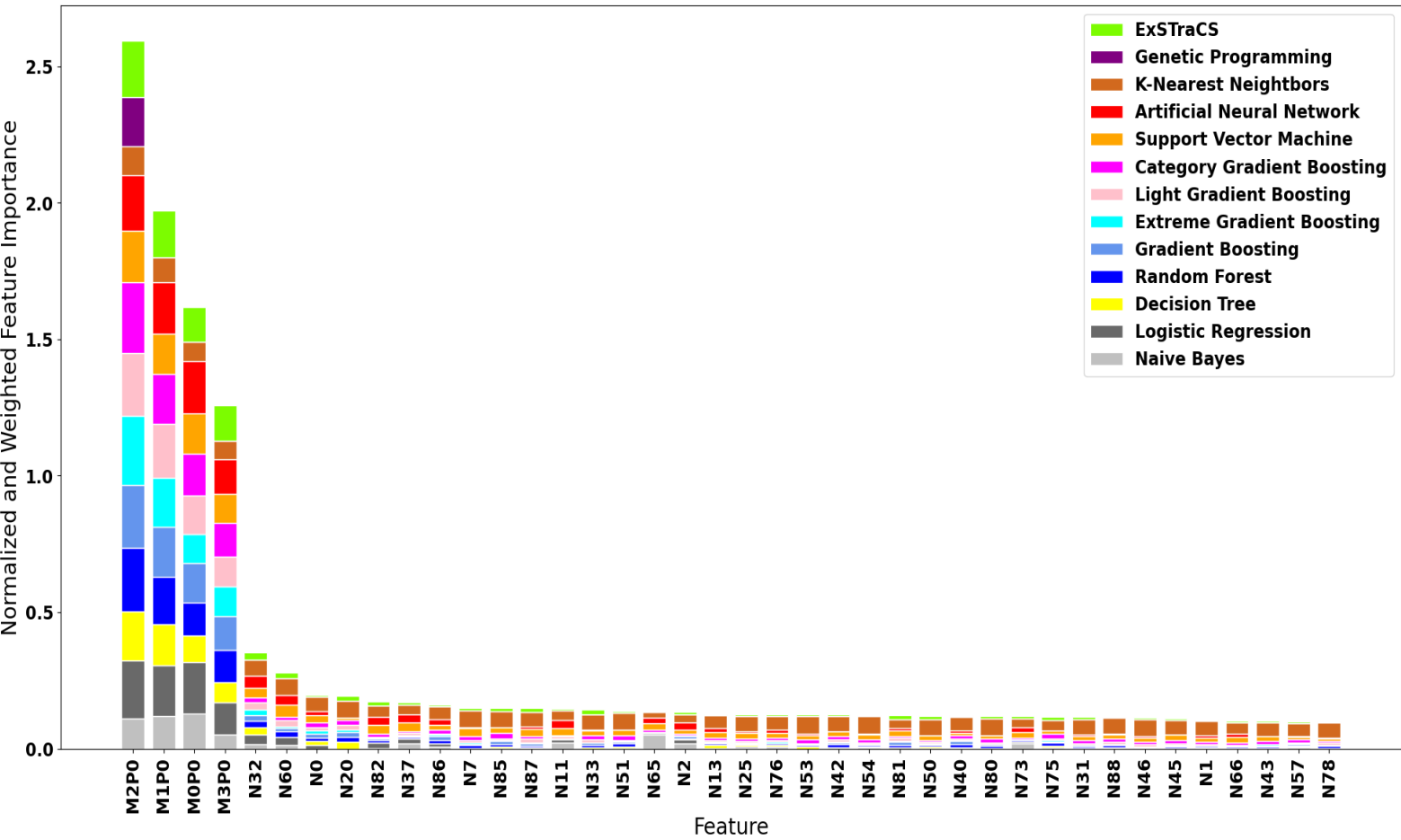
Feature Importance Summary: D4 = D_gametes_2way_epistasis_L_2_H_0

Sorted Mutual Information Scores

Sorted MultiSURF Scores
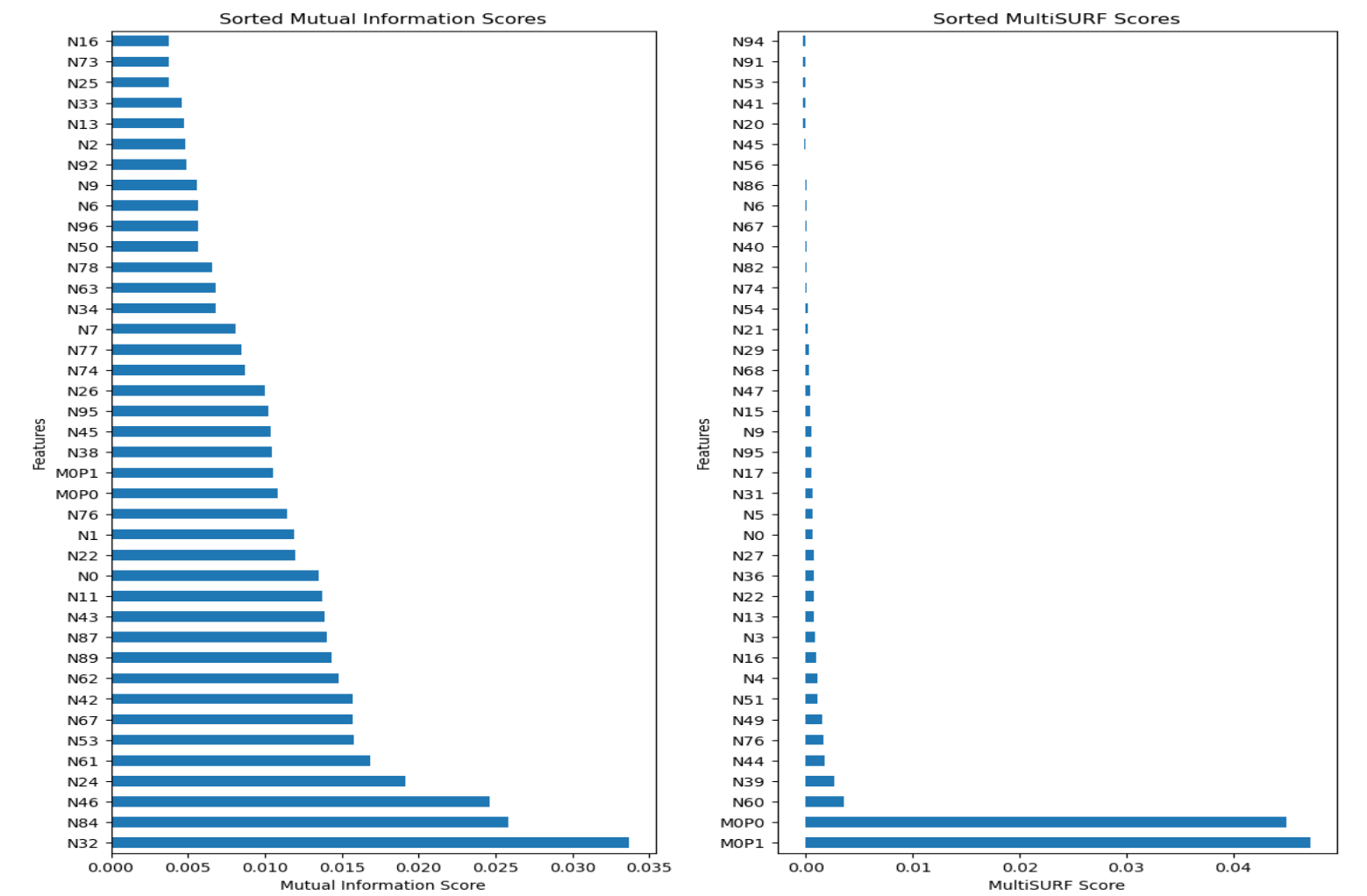
Composite Feature Importance Plot (Normalized and Performance Weighted)
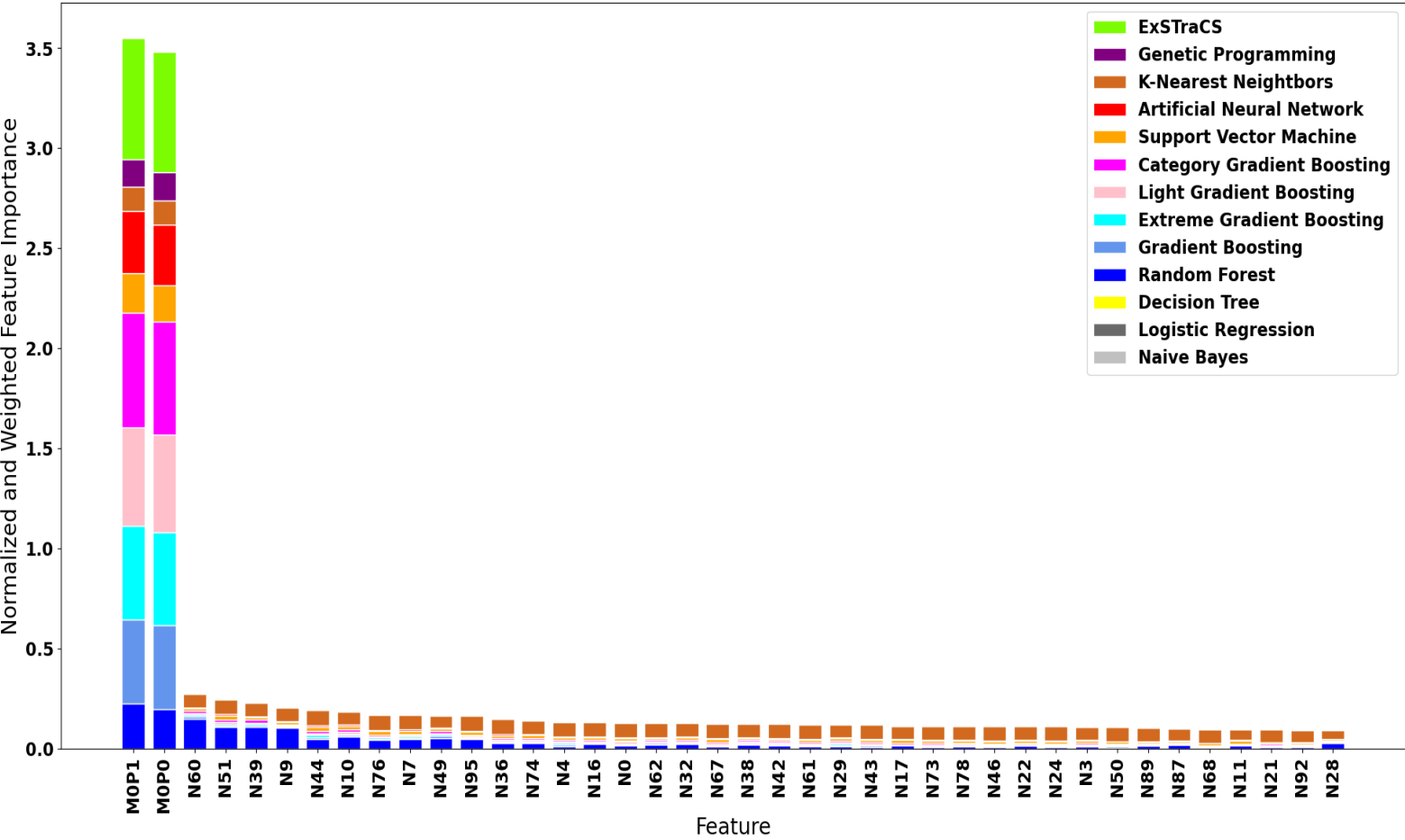
Feature Importance Summary: D5 = E_gametes_2way_epi_2het_L_2_H_0

## Sorted Mutual Information Scores

Sorted MultiSURF Scores

Composite Feature Importance Plot (Normalized and Performance Weighted)

Legend:
- ExSTraCS
- Genetic Programming
- K-Nearest Neightbors
- Artificial Neural Network
- Support Vector Machine
- Category Gradient Boosting
- Light Gradient Boosting
- Extreme Gradient Boosting
- Gradient Boosting
- Random Forest
- Decision Tree
- Logistic Regression
- Naive Bayes
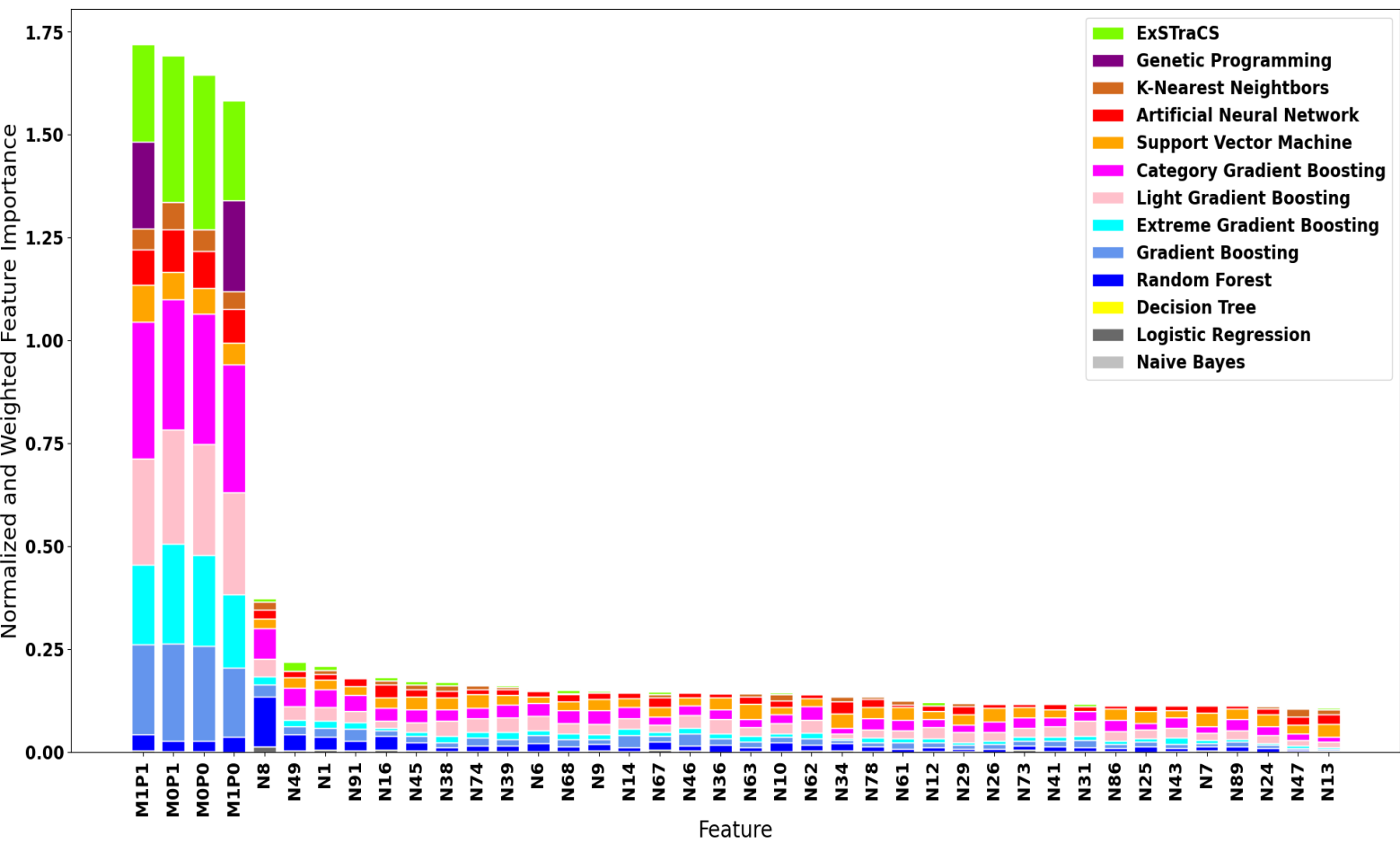
Feature Importance Summary: D6 = F_gametes_3way_epistasis_L_3_H_0

Sorted Mutual Information Scores

Sorted MultiSURF Scores

Composite Feature Importance Plot (Normalized and Performance Weighted)
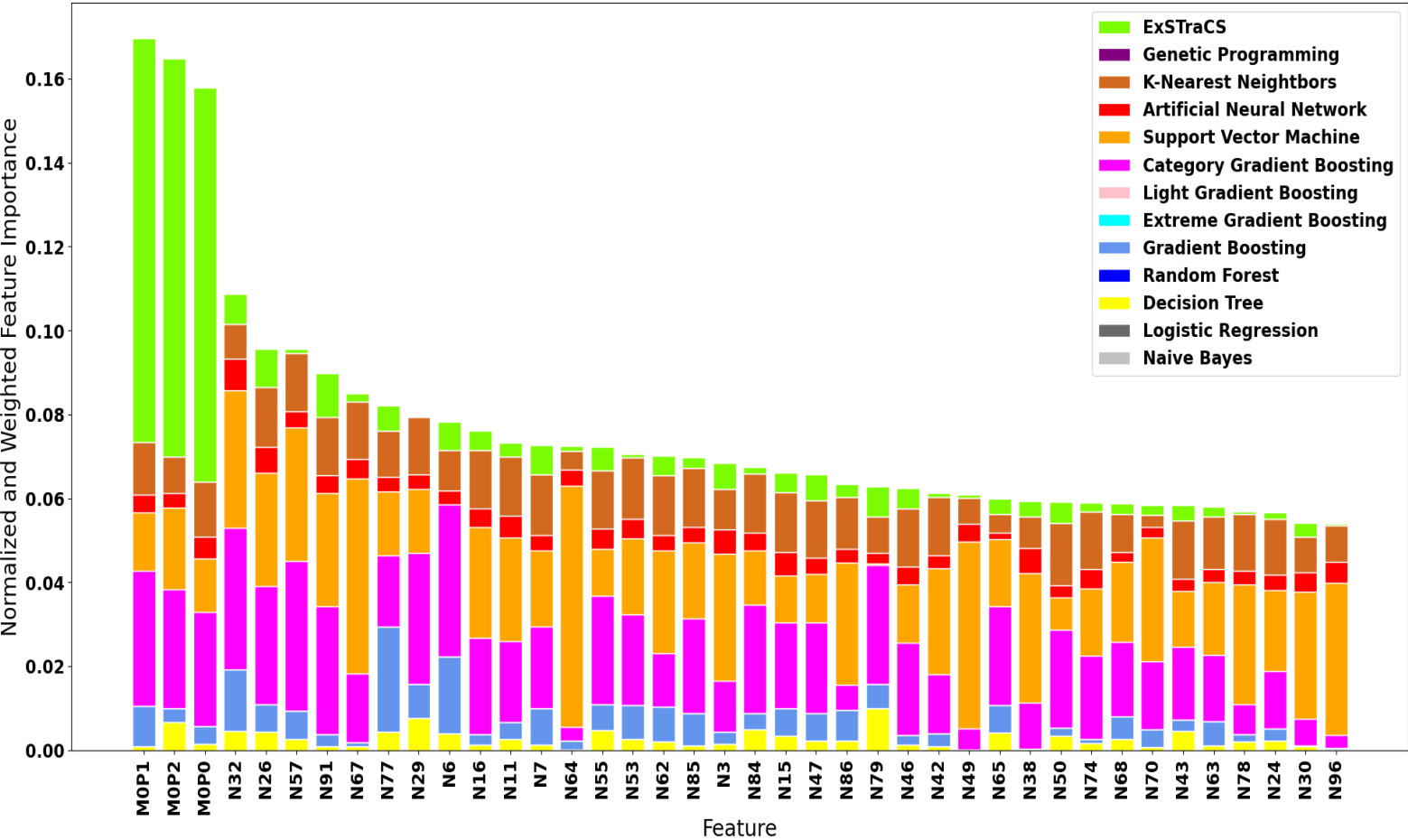
Compare ML Performance Across Datasets
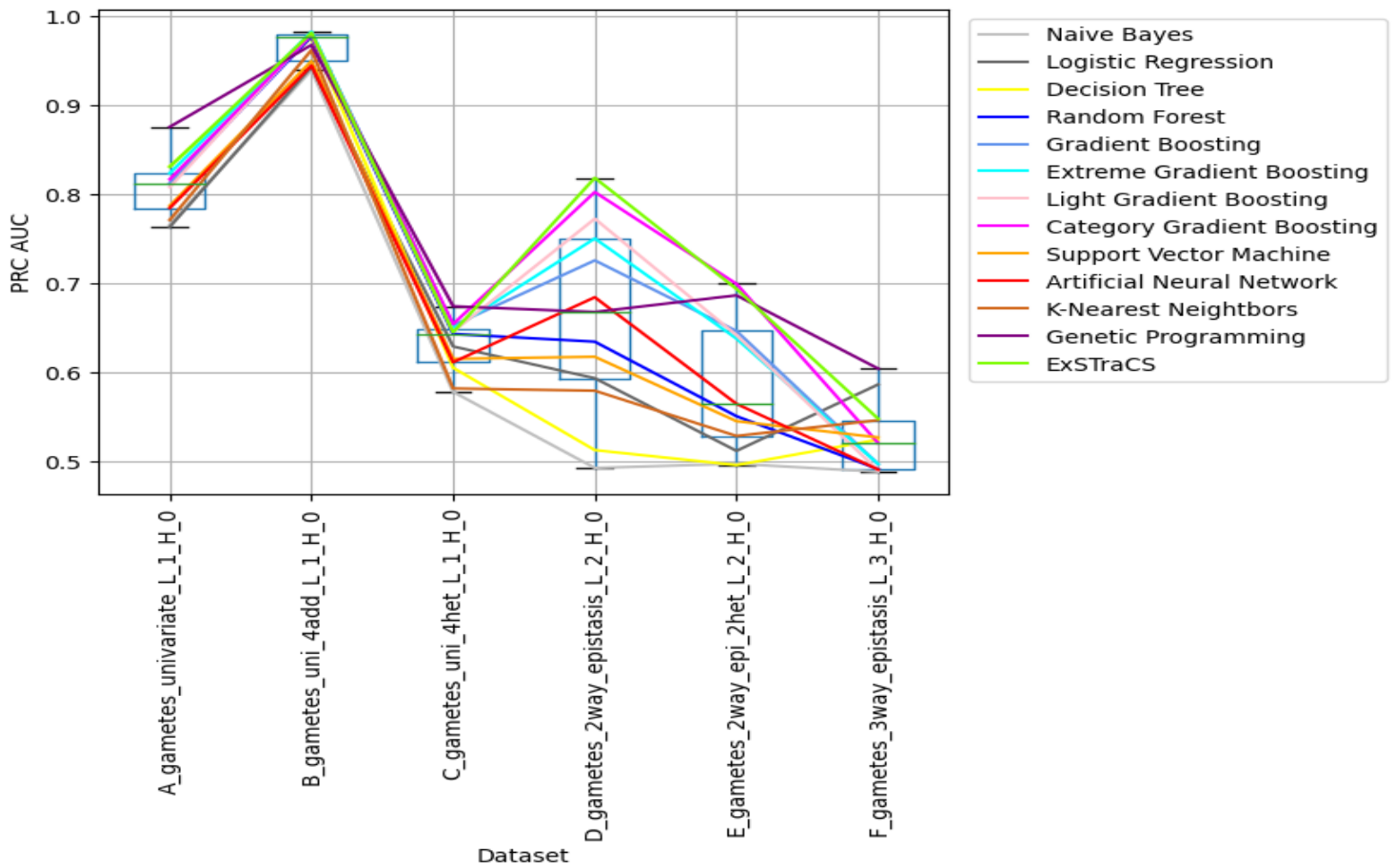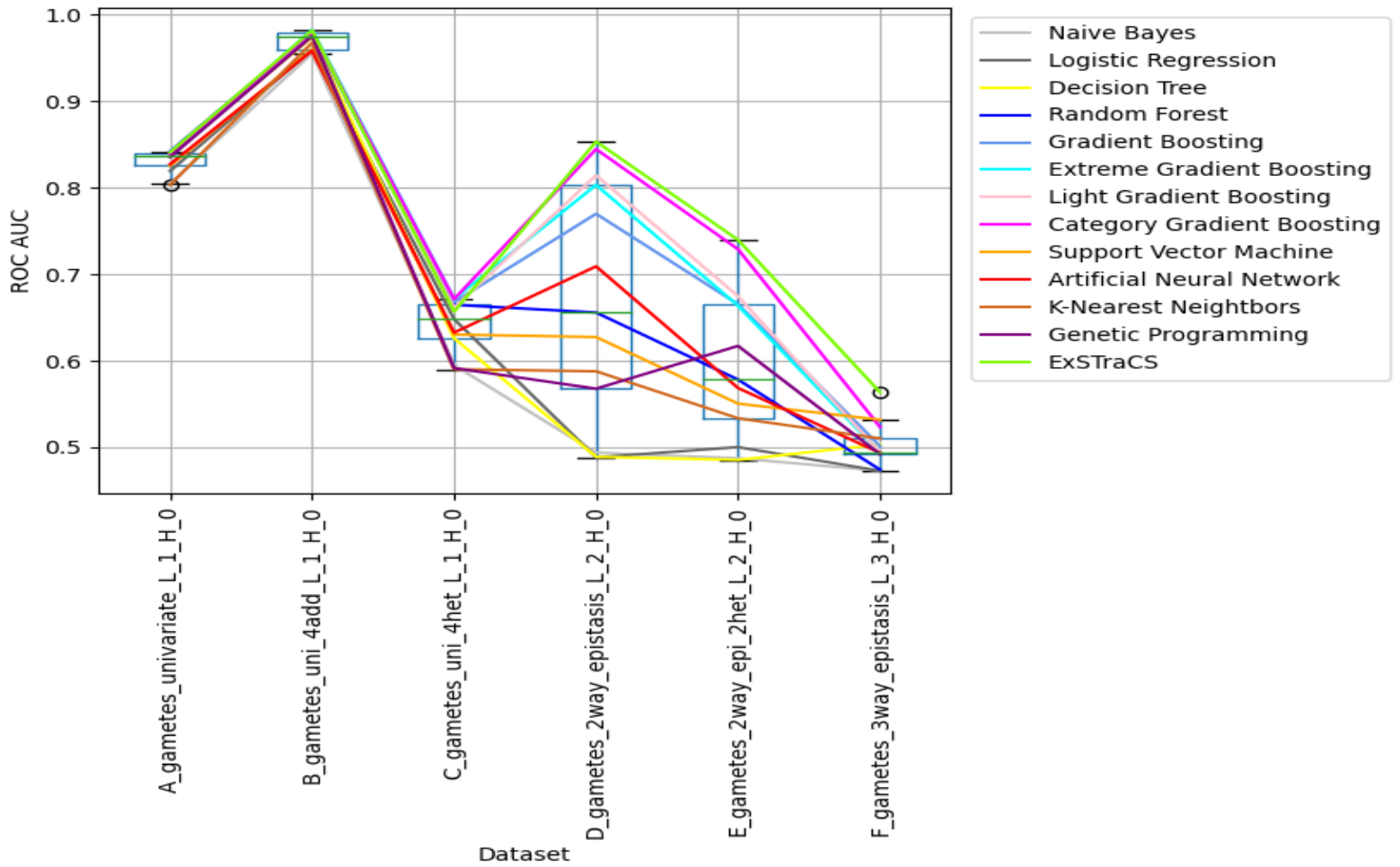
Using Best Performing Algorithms (Kruskall Wallis Compare Datasets)

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	Mean_D1	Best_Alg_D2	Mean_D2	Best_Alg_D3	Mean_D3
Balanced Accuracy	0.0	ExSTraCS	0.83	Extreme Gradient Boosting	0.9275	Category Gradient Boosting	0.63
Accuracy	0.0	ExSTraCS	0.83	Extreme Gradient Boosting	0.9275	Category Gradient Boosting	0.63
F1 Score	0.0	ExSTraCS	0.8266	Extreme Gradient Boosting	0.9281	Category Gradient Boosting	0.6207
Sensitivity (Recall)	0.0	Logistic Regression	0.8112	ExSTraCS	0.9388	Decision Tree	0.6325
Specificity	0.0057	ExSTraCS	0.8488	K-Nearest Neightbors	0.93	Genetic Programming	0.7188
Precision (PPV)	0.0	ExSTraCS	0.8435	K-Nearest Neightbors	0.9295	Category Gradient Boosting	0.6375
TP	0.0	Logistic Regression	64.9	ExSTraCS	75.1	Decision Tree	50.6
TN	0.0057	ExSTraCS	67.9	K-Nearest Neightbors	74.4	Genetic Programming	57.5
FP	0.0	Naive Bayes	15.2	Artificial Neural Network	9.5	Decision Tree	36.2
FN	0.0	K-Nearest Neightbors	25.8	K-Nearest Neightbors	8.9	K-Nearest Neightbors	44.7
NPV	0.0	ExSTraCS	0.8188	ExSTraCS	0.9383	Category Gradient Boosting	0.6261
LR+	0.0	ExSTraCS	5.5415	K-Nearest Neightbors	20.6883	Category Gradient Boosting	1.7879
LR-	0.0	K-Nearest Neightbors	0.3843	K-Nearest Neightbors	0.1191	K-Nearest Neightbors	0.8403
ROC AUC	0.0	Random Forest	0.842	Extreme Gradient Boosting	0.9826	Extreme Gradient Boosting	0.6715
PRC AUC	0.0	Genetic Programming	0.8753	Extreme Gradient Boosting	0.9822	Genetic Programming	0.6738
PRC APS	0.0	ExSTraCS	0.8326	Extreme Gradient Boosting	0.9821	Category Gradient Boosting	0.6587

index	P-Value	Best_Alg_D4	Mean_D4	Best_Alg_D5	Mean_D5	Best_Alg_D6	Mean_D6
Balanced Accuracy	0.0	ExSTraCS	0.8025	ExSTraCS	0.6881	ExSTraCS	0.5481
Accuracy	0.0	ExSTraCS	0.8025	ExSTraCS	0.6881	ExSTraCS	0.5481
F1 Score	0.0	ExSTraCS	0.825	ExSTraCS	0.6956	ExSTraCS	0.56
Sensitivity (Recall)	0.0	ExSTraCS	0.9325	Genetic Programming	0.7262	ExSTraCS	0.5888
Specificity	0.0057	Logistic Regression	0.735	Category Gradient Boosting	0.6637	Logistic Regression	0.7163
Precision (PPV)	0.0	Category Gradient Boosting	0.7434	ExSTraCS	0.6802	ExSTraCS	0.5385
TP	0.0	ExSTraCS	74.6	Genetic Programming	58.1	ExSTraCS	47.1
TN	0.0057	Logistic Regression	58.8	Category Gradient Boosting	53.1	Logistic Regression	57.3
FP	0.0	Naive Bayes	41.0	Naive Bayes	41.4	Support Vector Machine	41.6
FN	0.0	Logistic Regression	59.8	Decision Tree	44.9	Logistic Regression	60.0
NPV	0.0	ExSTraCS	0.9126	ExSTraCS	0.6976	ExSTraCS	0.5633
LR+	0.0	Category Gradient Boosting	2.93	ExSTraCS	2.1915	ExSTraCS	1.2076
LR-	0.0	Decision Tree	1.0544	Decision Tree	1.098	Random Forest	1.1211
ROC AUC	0.0	ExSTraCS	0.8535	ExSTraCS	0.7397	ExSTraCS	0.5635
PRC AUC	0.0	ExSTraCS	0.818	Category Gradient Boosting	0.699	Genetic Programming	0.6038
PRC APS	0.0	ExSTraCS	0.8204	Category Gradient Boosting	0.7036	ExSTraCS	0.5544

Pipeline Runtime Summary

A_gametes_univariate_L_1_H_0

Pipeline Component	Time (sec)
Exploratory Analysis	3.55
Preprocessing	0.13
Mutual Information	5.12
MultiSURF	743.69
Feature Selection	1.92
Naive Bayes	11.24
Logistic Regression	70.33
Decision Tree	69.24
Random Forest	4586.72
Gradient Boosting	8664.31
Extreme Gradient Boosting	9385.29
Light Gradient Boosting	522.38
Category Gradient Boosting	10480.3
Support Vector Machine	149433.96
Artificial Neural Network	4279.6
K-Nearest Neightbors	1023.39
Genetic Programming	12727.35
ExSTraCS	21502.83
Stats Summary	24.55

B_gametes_uni_4add_L_1_H_0

Pipeline Component	Time (sec)
Exploratory Analysis	3.85
Preprocessing	0.13
Mutual Information	5.03
MultiSURF	747.69
Feature Selection	2.04
Naive Bayes	10.21
Logistic Regression	79.11
Decision Tree	74.86
Random Forest	5462.24
Gradient Boosting	7783.81
Extreme Gradient Boosting	8987.97
Light Gradient Boosting	819.18
Category Gradient Boosting	10317.79
Support Vector Machine	8972.13
Artificial Neural Network	3309.18
K-Nearest Neightbors	1000.56
Genetic Programming	10403.74
ExSTraCS	18191.46
Stats Summary	26.27

C_gametes_uni_4het_L_1_H_0

Pipeline Component	Time (sec)
Exploratory Analysis	3.88
Preprocessing	0.13
Mutual Information	5.16
MultiSURF	748.85
Feature Selection	2.12
Naive Bayes	11.63
Logistic Regression	82.51
Decision Tree	70.41
Random Forest	4896.16
Gradient Boosting	6639.7
Extreme Gradient Boosting	8121.46
Light Gradient Boosting	449.54
Category Gradient Boosting	10225.78
Support Vector Machine	16696.0
Artificial Neural Network	4208.08
K-Nearest Neightbors	1147.51
Genetic Programming	13415.76
ExSTraCS	24315.9
Stats Summary	24.96

D_gametes_2way_epistasis_L_2_H_0

Pipeline Component	Time (sec)
Exploratory Analysis	3.58
Preprocessing	0.13
Mutual Information	5.2
MultiSURF	748.77
Feature Selection	1.78
Naive Bayes	13.15
Logistic Regression	85.03
Decision Tree	76.25
Random Forest	9798.89
Gradient Boosting	9205.88
Extreme Gradient Boosting	9299.29
Light Gradient Boosting	1816.47
Category Gradient Boosting	10354.57
Support Vector Machine	10315.38
Artificial Neural Network	5106.34
K-Nearest Neightbors	1234.65
Genetic Programming	12639.7
ExSTraCS	20601.63
Stats Summary	24.47

E_gametes_2way_epi_2het_L_2_H_0

Pipeline Component	Time (sec)
Exploratory Analysis	3.47
Preprocessing	0.13
Mutual Information	5.01
MultiSURF	746.72
Feature Selection	1.81
Naive Bayes	13.55
Logistic Regression	98.23
Decision Tree	79.21
Random Forest	9646.6
Gradient Boosting	9160.14
Extreme Gradient Boosting	9290.73
Light Gradient Boosting	1885.75
Category Gradient Boosting	10379.98
Support Vector Machine	12577.66
Artificial Neural Network	5092.9
K-Nearest Neightbors	1490.95
Genetic Programming	13570.32
ExSTraCS	24006.04
Stats Summary	24.92

F_gametes_3way_epistasis_L_3_H_0

Pipeline Component	Time (sec)
Exploratory Analysis	3.58
Preprocessing	0.16
Mutual Information	5.29
MultiSURF	743.13
Feature Selection	1.84
Naive Bayes	14.76
Logistic Regression	97.95
Decision Tree	80.13
Random Forest	5842.91
Gradient Boosting	9376.3
Extreme Gradient Boosting	8350.03
Light Gradient Boosting	2398.29
Category Gradient Boosting	10132.32
Support Vector Machine	12143.42
Artificial Neural Network	4131.81
K-Nearest Neightbors	1132.58
Genetic Programming	12133.17
ExSTraCS	25100.5
Stats Summary	24.0