

STREAMLINE Training Summary Report: 2022-06-22 06:10:48.621406

General Pipeline Settings:

Data Path: /content/drive/MyDrive/Colab
Notebooks/Lasso/STREAMLINE-main/MyData
Output Path: /content/drive/MyDrive/Colab
Notebooks/Lasso/STREAMLINE-main/Colab_Output
Experiment Name: vbm_report_1
Class Label: Class
Instance Label: InstanceID
Ignored Features: []
Specified Categorical Features: []
CV Partitions: 3
Partition Method: S
Match Label: None
Categorical Cutoff: 10
Statistical Significance Cutoff: 0.05
Export Feature Correlations: True
Export Univariate Plots: True
Random Seed: 42
Run From Jupyter Notebook: True
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: False
Top Features to Display: 40
Export Feature Importance Plot: True
Overwrite CV Datasets: True
ExSTraCS: False
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 ROC Plot: True
Export PRC Plot: True
Export Metric Boxplots: True
Export Feature Importance Boxplots: True
Top Model Features To Display: 40

ML Modeling Algorithms:

Naive Bayes: True
Elastic Net: 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 Neighbors: True
Genetic Programming: False
eLCS: False
XCS: False

LCS Settings (eLCS,XCS,ExSTraCS):

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

Datasets:

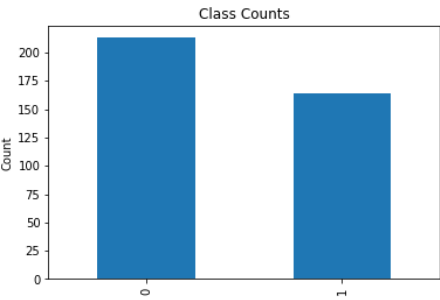
D1 = VBM_with_bi_diagclass

Univariate Analysis of Each Dataset (Top 10 Features for Each)

D1 = VBM_with_bi_diagclass

Feature: P-Value
Precentral_L: 9.118022191884288e-28
Postcentral_R: 2.464228965838577e-25
Postcentral_L: 6.894181237130811e-23
Precentral_R: 1.434752287873083e-20
Putamen_R: 4.407446958391657e-20
Cerebelum_8_L: 1.104077185113905e-18
Cerebelum_7b_L: 4.4667037842088395e-18
Cerebelum_6_R: 5.918799786097711e-18
Parietal_Sup_R: 7.899599654013874e-18
Cerebelum_4_5_R: 1.1909105085962023e-17

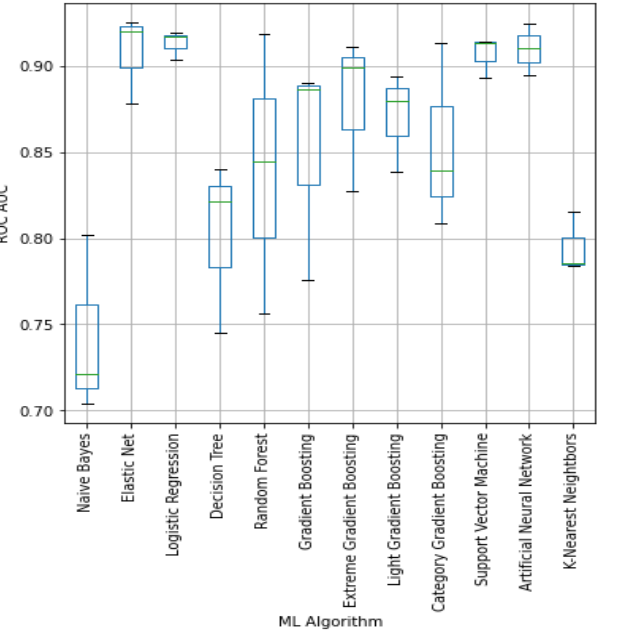
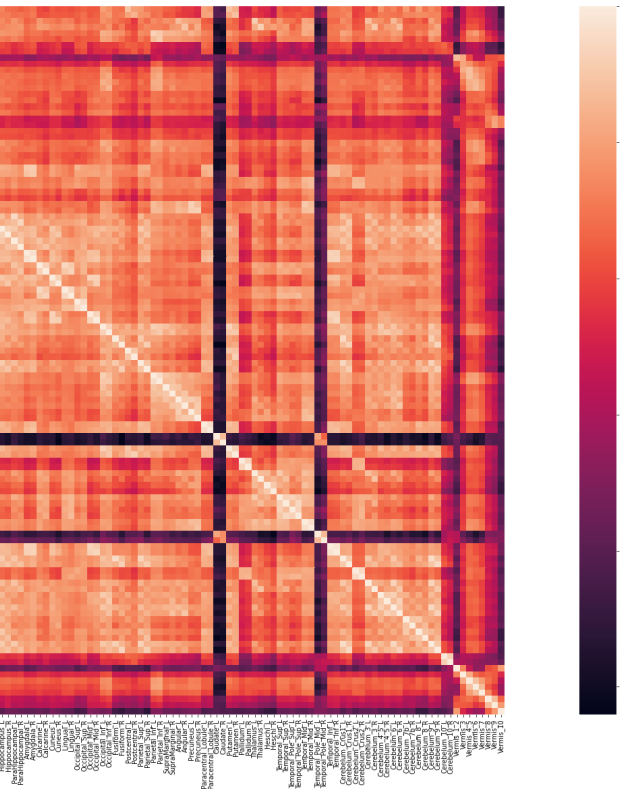
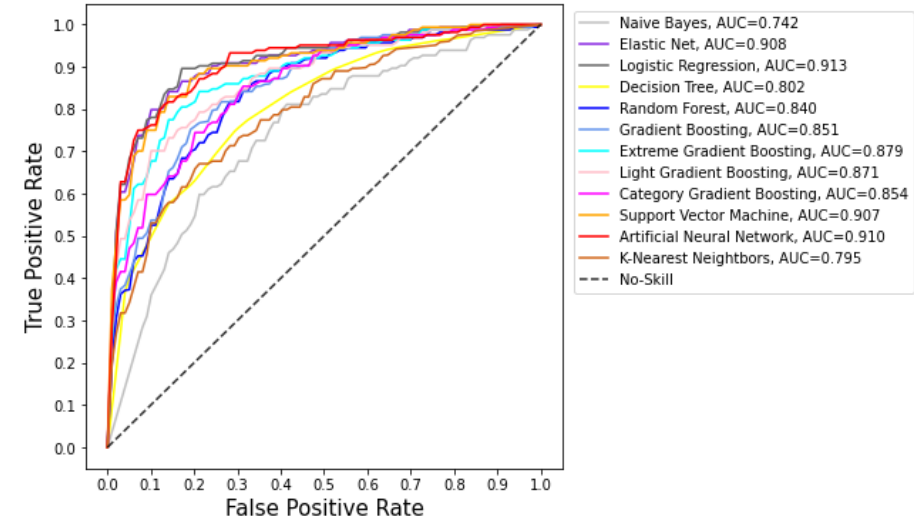
Dataset and Model Prediction Summary: D1 = VBM_with_bi_diagclass



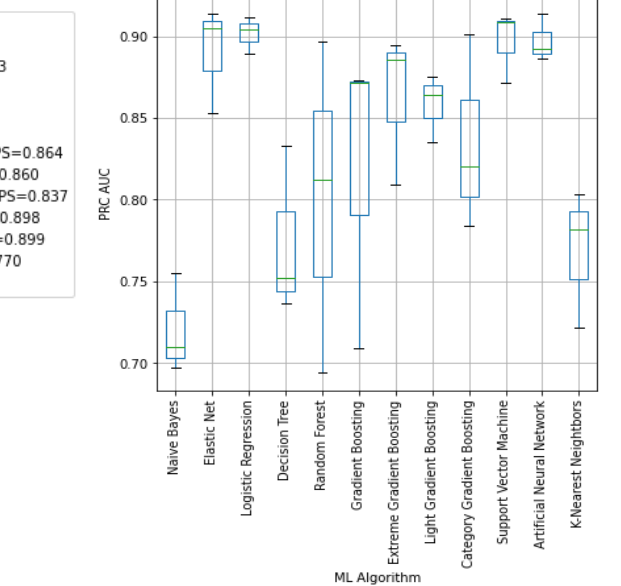
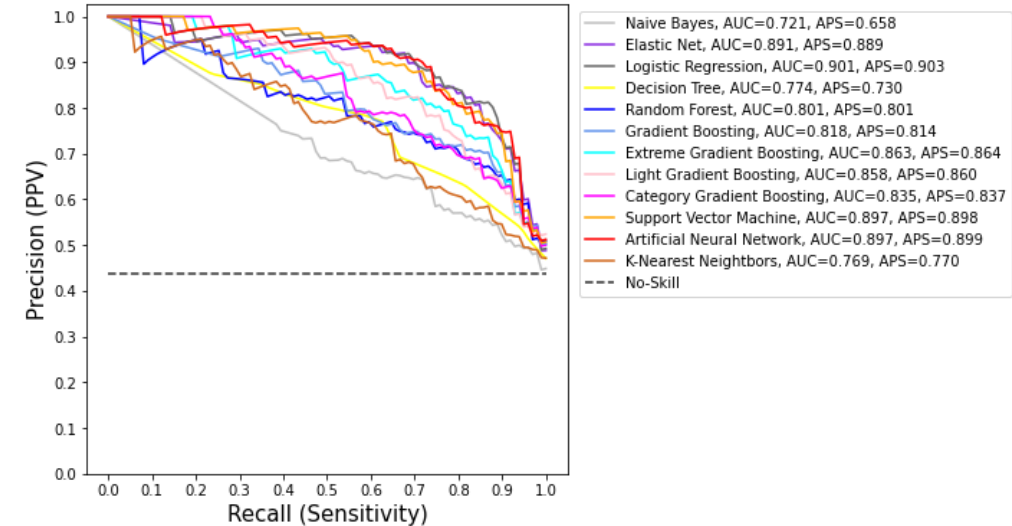
Dataset Counts Summary:	
instances:	377.0
features:	116.0
categorical_features:	0.0
quantitative_features:	116.0
missing_values:	0.0
missing_percent:	0.0

Top ML Algorithm Results (Averaged Over CV Runs):	
Best (ROC_AUC):	Logistic Regression = 0.913
Best (Balanced Acc.):	Logistic Regression = 0.851
Best (F1 Score):	Logistic Regression = 0.828
Best (PRC AUC):	Logistic Regression = 0.901
Best (PRC APS):	Logistic Regression = 0.903

ROC



PRC

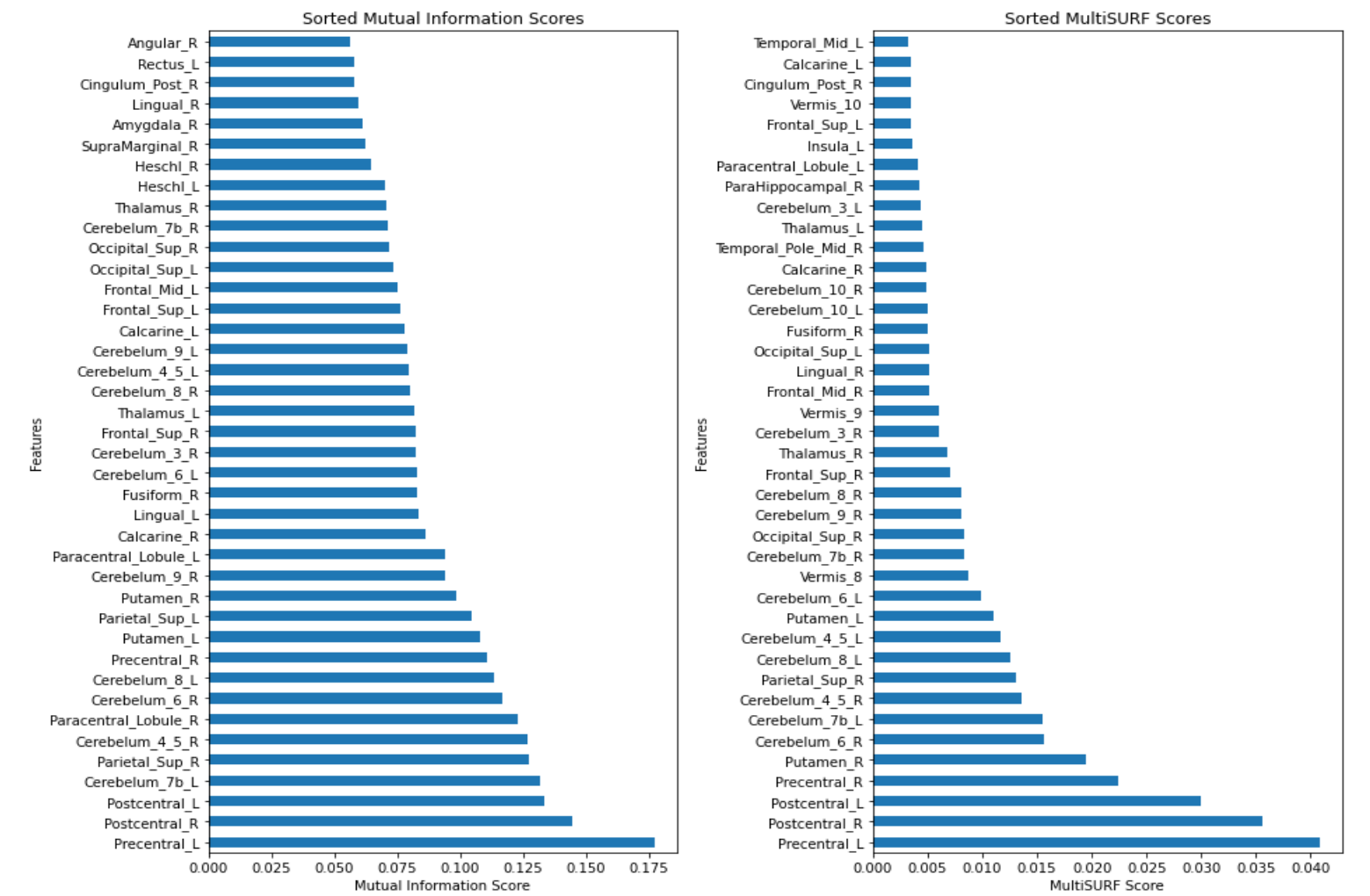


Average Model Prediction Statistics (Rounded to 3 Decimal Points)

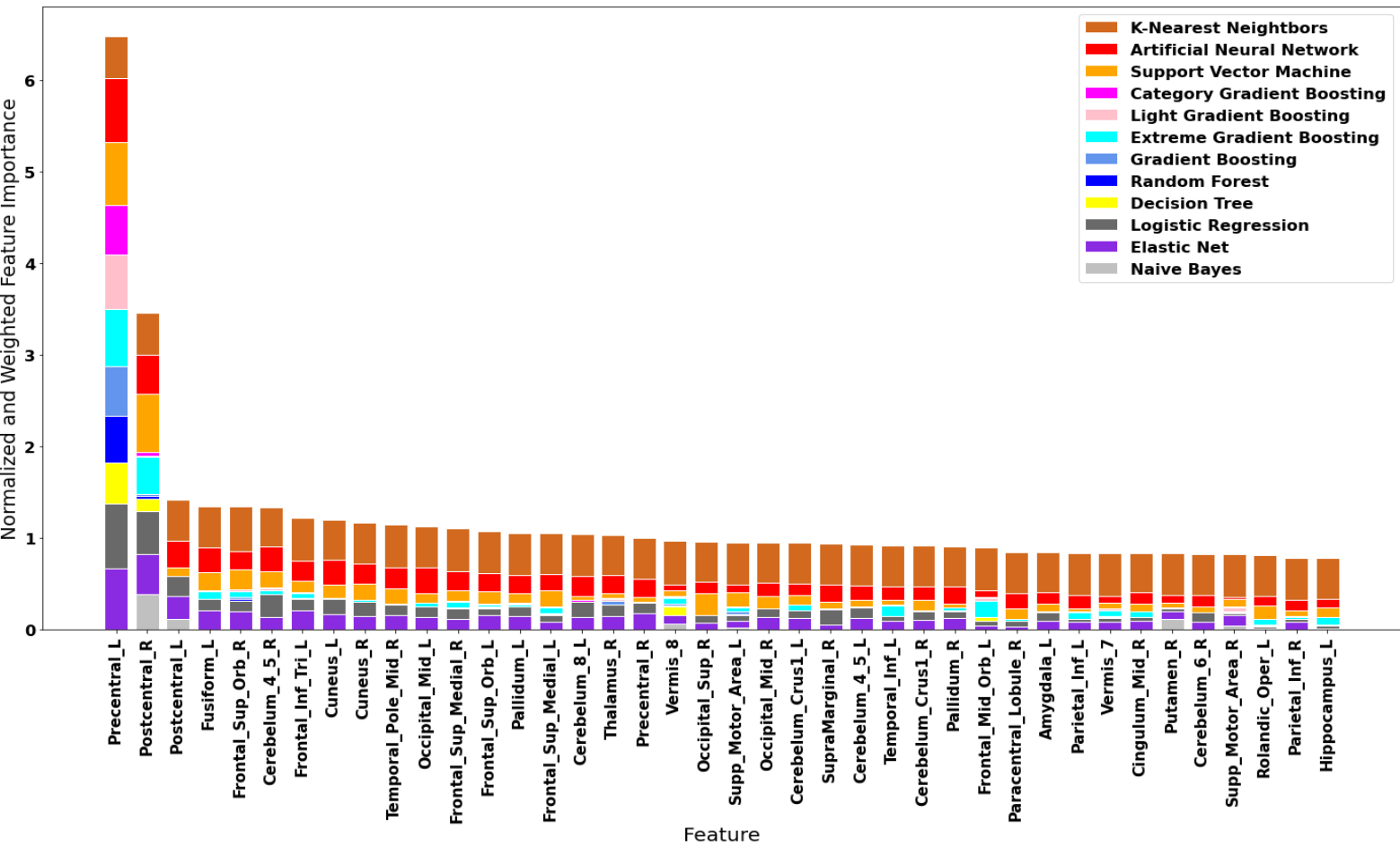
D1 = VBM_with_bi_diagclass

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	0.688	0.692	0.648	0.653	0.723	0.647	35.667	51.333	19.667	19.0	0.731	2.44	0.481	0.742	0.721	0.658
Elastic Net	0.816	0.83	0.784	0.707	0.925	0.879	38.667	65.667	5.333	16.0	0.804	9.484	0.316	0.908	0.891	0.889
Logistic Regression	0.851	0.857	0.828	0.805	0.897	0.858	44.0	63.667	7.333	10.667	0.86	7.906	0.216	0.913	0.901	0.903
Decision Tree	0.713	0.724	0.663	0.622	0.803	0.712	34.0	57.0	14.0	20.667	0.734	3.317	0.471	0.802	0.774	0.73
Random Forest	0.75	0.756	0.716	0.701	0.798	0.733	38.333	56.667	14.333	16.333	0.776	4.297	0.38	0.84	0.801	0.801
Gradient Boosting	0.764	0.772	0.726	0.702	0.826	0.754	38.333	58.667	12.333	16.333	0.785	4.083	0.362	0.851	0.818	0.814
Extreme Gradient Boosting	0.807	0.812	0.781	0.774	0.84	0.792	42.333	59.667	11.333	12.333	0.831	5.341	0.268	0.879	0.863	0.864
Light Gradient Boosting	0.792	0.798	0.763	0.744	0.84	0.785	40.667	59.667	11.333	14.0	0.811	4.966	0.305	0.871	0.858	0.86
Category Gradient Boosting	0.767	0.772	0.735	0.726	0.808	0.747	39.667	57.333	13.667	15.0	0.792	4.096	0.341	0.854	0.835	0.837
Support Vector Machine	0.833	0.844	0.807	0.75	0.915	0.872	41.0	65.0	6.0	13.667	0.826	8.877	0.273	0.907	0.897	0.898
Artificial Neural Network	0.84	0.846	0.817	0.792	0.887	0.846	43.333	63.0	8.0	11.333	0.849	7.653	0.234	0.91	0.897	0.899
K-Nearest Neighbors	0.727	0.745	0.668	0.591	0.864	0.772	32.333	61.333	9.667	22.333	0.734	4.512	0.473	0.795	0.769	0.77

Feature Importance Summary: D1 = VBM_with_bi_diagclass



Composite Feature Importance Plot (Normalized and Performance Weighted)



Using Best Performing Algorithms (Kruskall Wallis Compare Datasets)

Datasets:
D1 = VBM_with_bi_diagclass

Pipeline Runtime Summary

VBM_with_bi_diagclass	
Pipeline Component	Time (sec)
Exploratory Analysis	34.24
Preprocessing	0.12
Mutual Information	1.06
MultiSURF	18.54
Feature Selection	2.27
Naive Bayes	4.0
Elastic Net	21.83
Logistic Regression	97.06
Decision Tree	33.71
Random Forest	2117.9
Gradient Boosting	2729.34
Extreme Gradient Boosting	459.68
Light Gradient Boosting	184.27
Category Gradient Boosting	2895.15
Support Vector Machine	65.87
Artificial Neural Network	665.92
K-Nearest Neighbors	769.28
Stats Summary	60.87