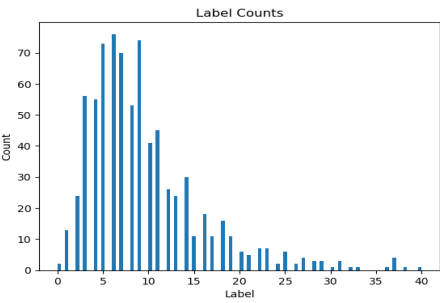


STREAMLINE Training Summary Report: 2023-06-29 20:30:03.430799

General Pipeline Settings:	ML Modeling Algorithms:
Data Path: /Users/yanbo/Dropbox/STREAMLINE-Regression_AMIA/Measurements/ADAS11 Output Path: /Users/yanbo/Dropbox/STREAMLINE-Regression_AMIA/Colab_Output Experiment Name: Demo_Experiment Class Label: Cognition_Score Instance Label: Class Ignored Features: None Specified Categorical Features: None CV Partitions: 5 Partition Method: R 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: False Top Features to Display: 40 Export Feature Importance Plot: True Overwrite CV Datasets: True Primary Metric: explained_variance Uniform Feature Importance Estimation (Models): True Hyperparameter Sweep Number of Trials: 50 Hyperparameter Timeout: 900 Export Hyperparameter Sweep Plots: True Export Metric Boxplots: True Export Feature Importance Boxplots: True Metric Weighting Composite FI Plots: explained_variance Top Model Features To Display: 40	Linear Regression: True Elastic Net: True Group Lasso: False RF Regressor: False AdaBoost: False GradBoost: False SVR: True L21Reg: True L21GMMReg: False L21DGMMReg: False
	Datasets:
	D1 = av45_ADAS11 D2 = fdg_ADAS11 D3 = vbm_ADAS11

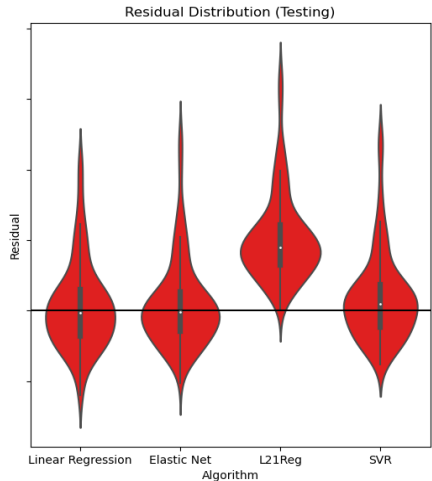
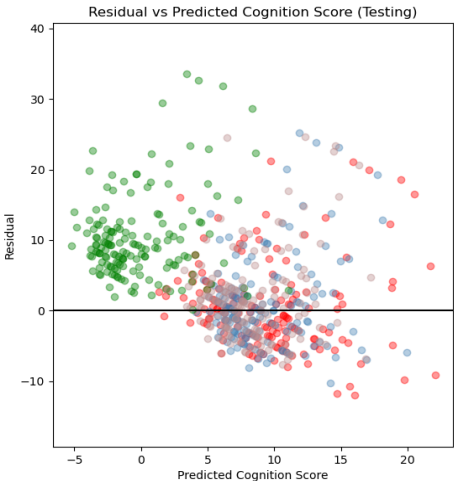
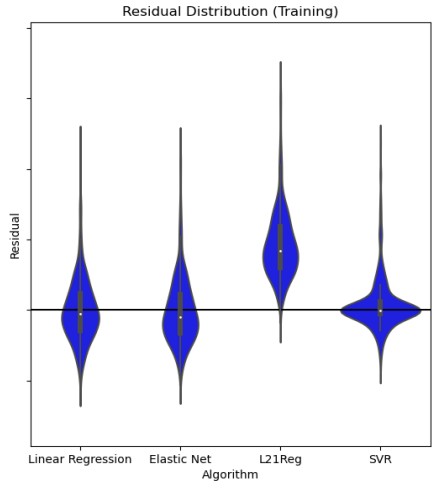
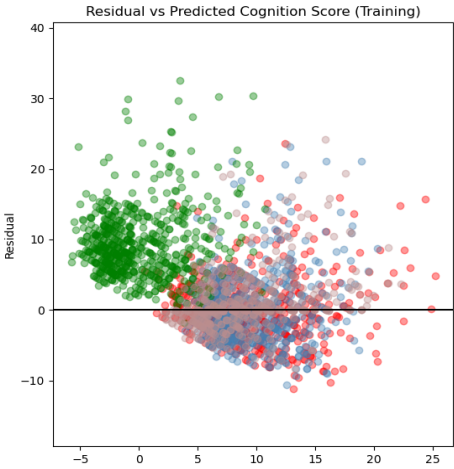
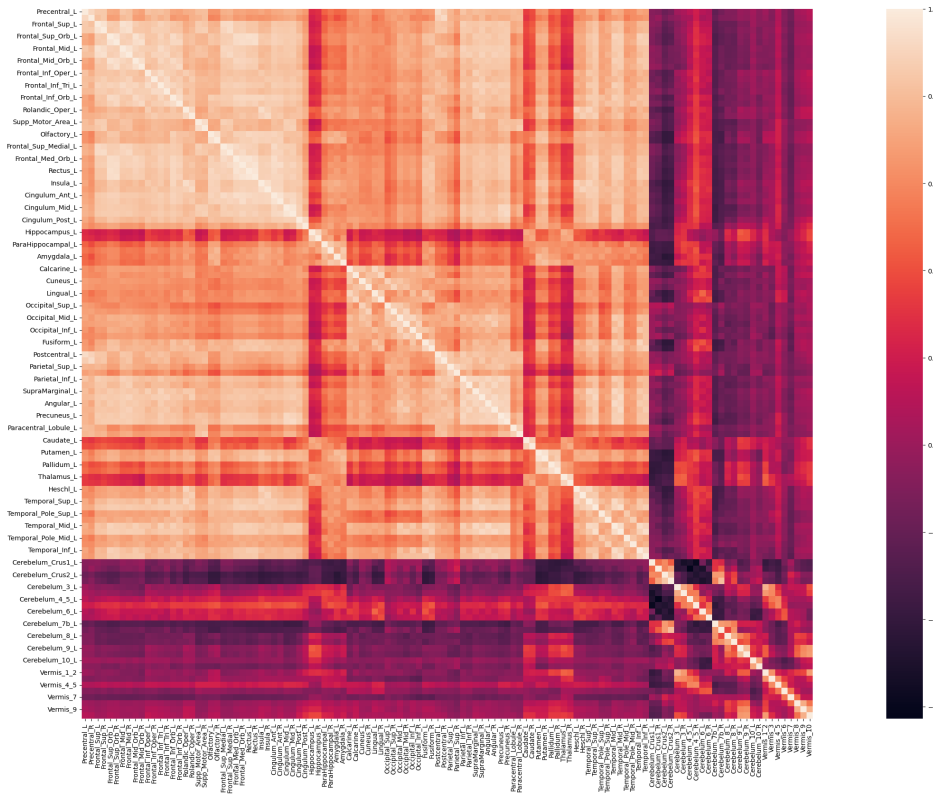
Univariate Analysis of Each Dataset (Top 10 Features for Each)
D1 = av45_ADAS11
Feature: P-Value
Cuneus_R: 0.019047619047619
Heschl_R: 0.1142857142857143
Amygdala_L: 0.1714285714285714
Amygdala_R: 0.1714285714285714
Occipital_Inf_L: 0.1714285714285714
Angular_R: 0.1714285714285714
Paracentral_Lobule_R: 0.1714285714285714
Caudate_L: 0.1714285714285714
Heschl_L: 0.1714285714285714
Temporal_Sup_L: 0.1714285714285714
D2 = fdg_ADAS11
Feature: P-Value
Occipital_Inf_L: 0.0761904761904762
Fusiform_L: 0.0761904761904762
Occipital_Mid_L: 0.1142857142857143
Cuneus_L: 0.1714285714285714
Cuneus_R: 0.1714285714285714
Temporal_Mid_R: 0.1714285714285714
Vermis_1_2: 0.1714285714285714
Vermis_10: 0.1714285714285714
Occipital_Sup_L: 0.2285714285714285
Heschl_L: 0.2285714285714285
D3 = vbm_ADAS11
Feature: P-Value
SupraMarginal_R: 0.0380952380952381
Frontal_Inf_Oper_L: 0.0761904761904762
Angular_R: 0.0761904761904762
Putamen_L: 0.0761904761904762
Heschl_R: 0.0761904761904762
Temporal_Sup_R: 0.0761904761904762
Temporal_Mid_R: 0.0761904761904762
Cerebelum_7b_R: 0.0761904761904762
Frontal_Inf_Tri_L: 0.1142857142857143
Frontal_Inf_Orb_R: 0.1142857142857143

Dataset and Model Prediction Summary: D1 = av45_ADAS11

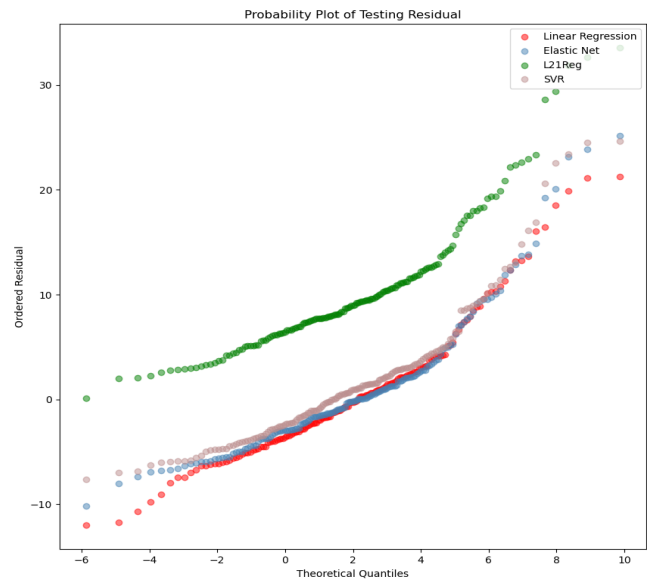
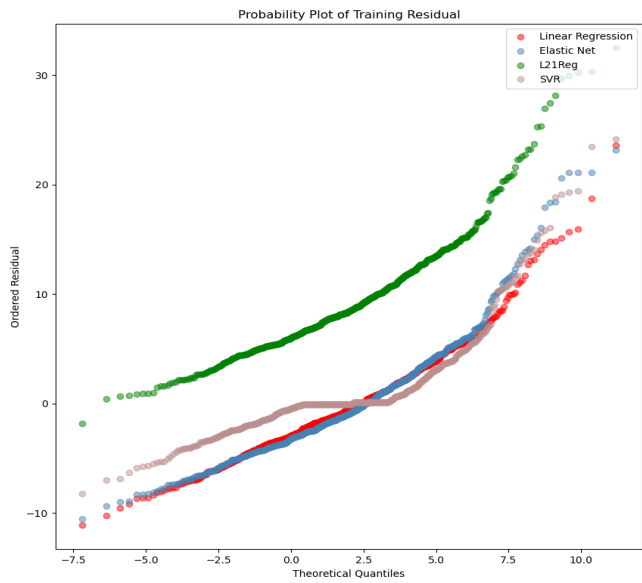
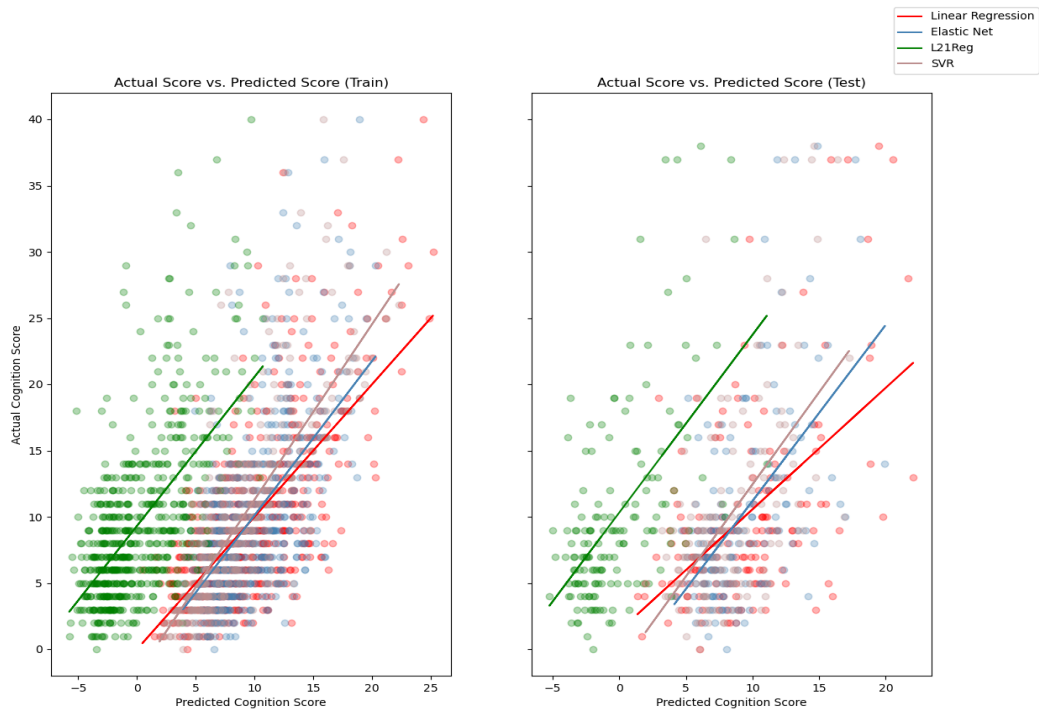


Dataset Counts Summary:
instances: 787.0
features: 116.0
categorical_features: 0.0
quantitative_features: 116.0
missing_values: 0.0
missing_percent: 0.0

Top ML Algorithm Results (Averaged Over CV Runs):
Best (Max Error): Linear Regression = 21.428
Best (Mean Absolute Error): SVR = 3.948
Best (Mean Squared Error): Elastic Net = 29.396
Best (Explained Variance): Elastic Net = 0.288
Best (Median Absolute Error): SVR = 2.937
Best (Pearson Correlation): Linear Regression = 0.491

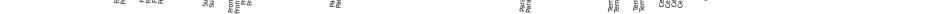


Dataset and Model Prediction Summary: D1 = av45_ADAS11

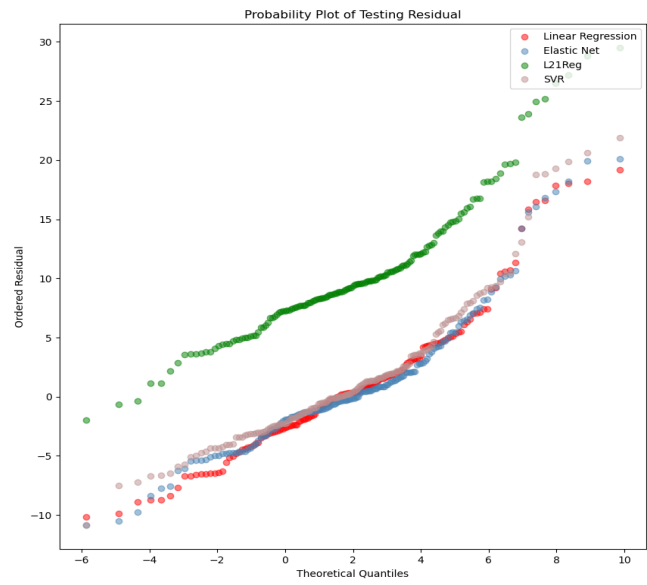
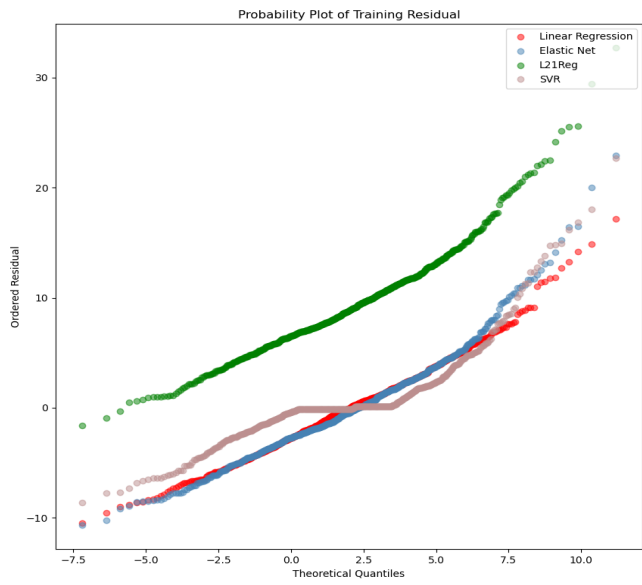
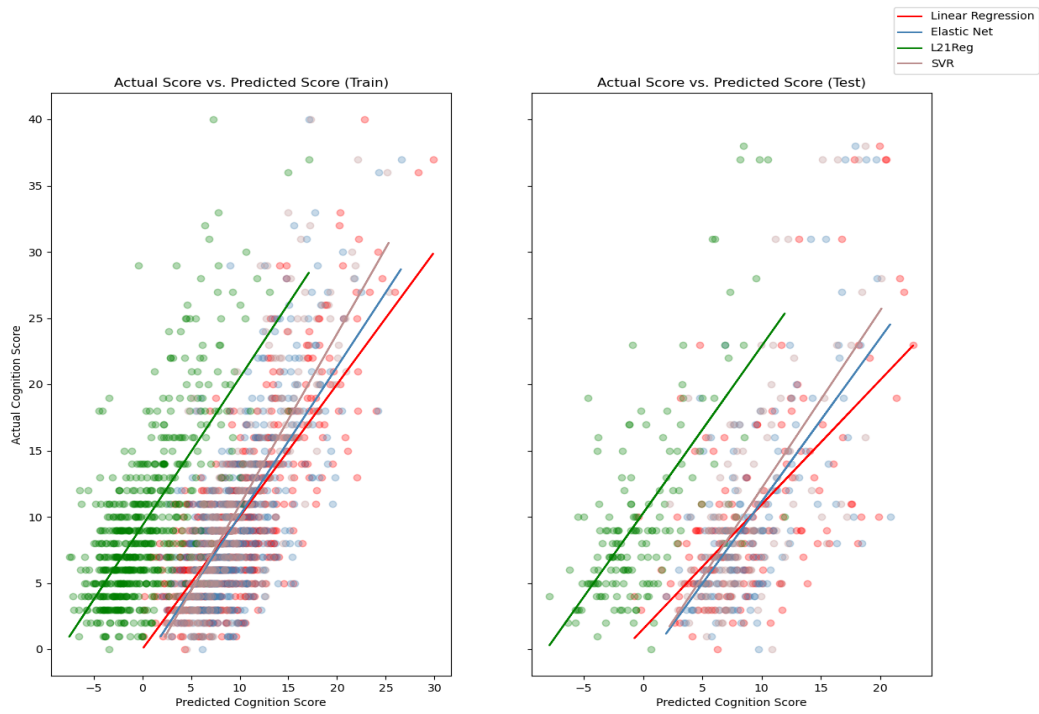


1000000 787.0

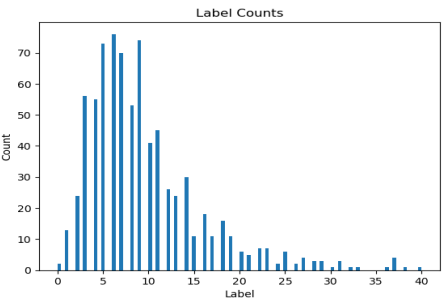
Р. (М. Е. \ Л' Р. : 17.000



Dataset and Model Prediction Summary: D2 = fdg_ADAS11

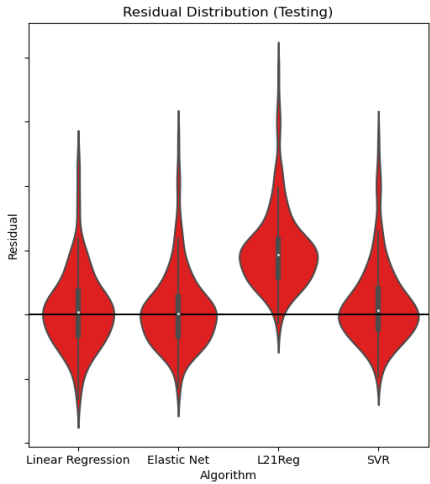
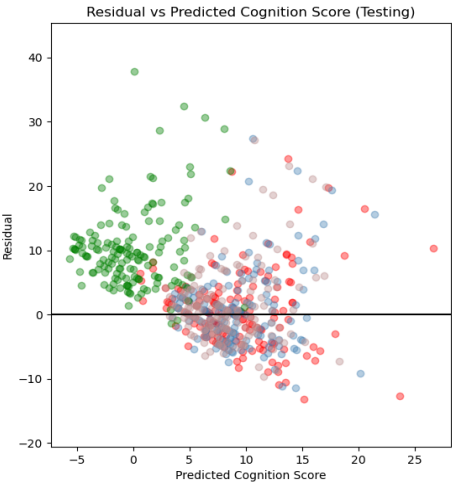
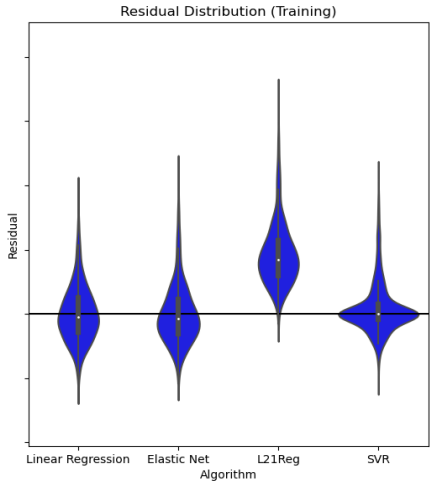
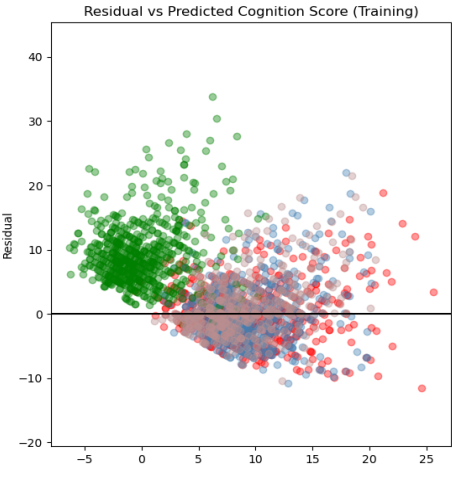
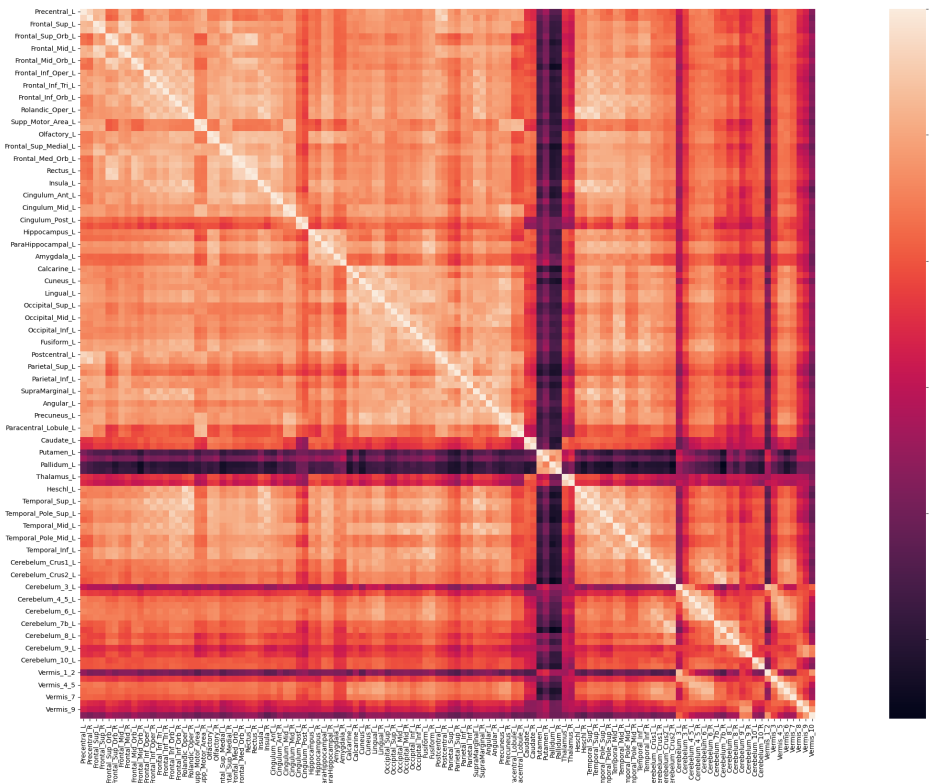


Dataset and Model Prediction Summary: D3 = vbm_ADAS11

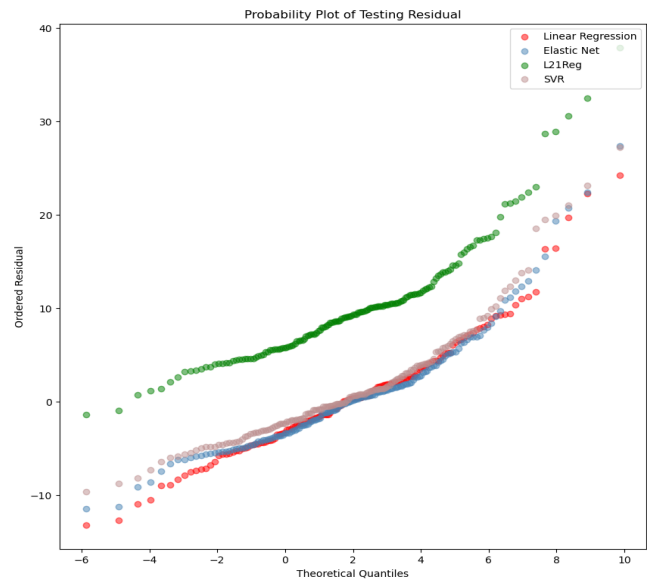
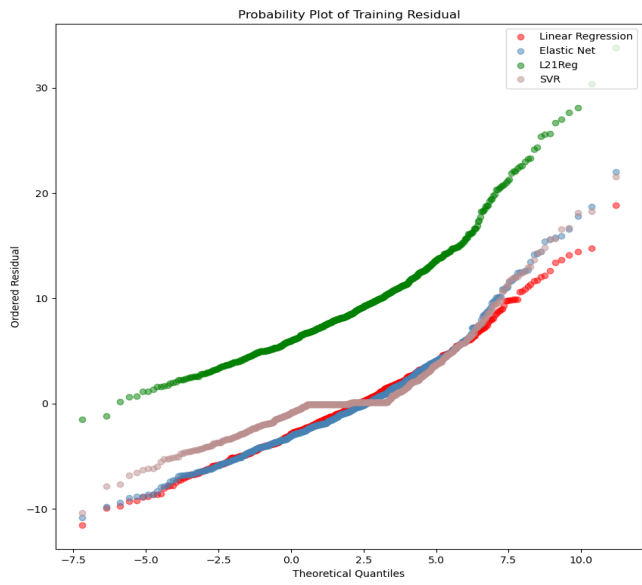
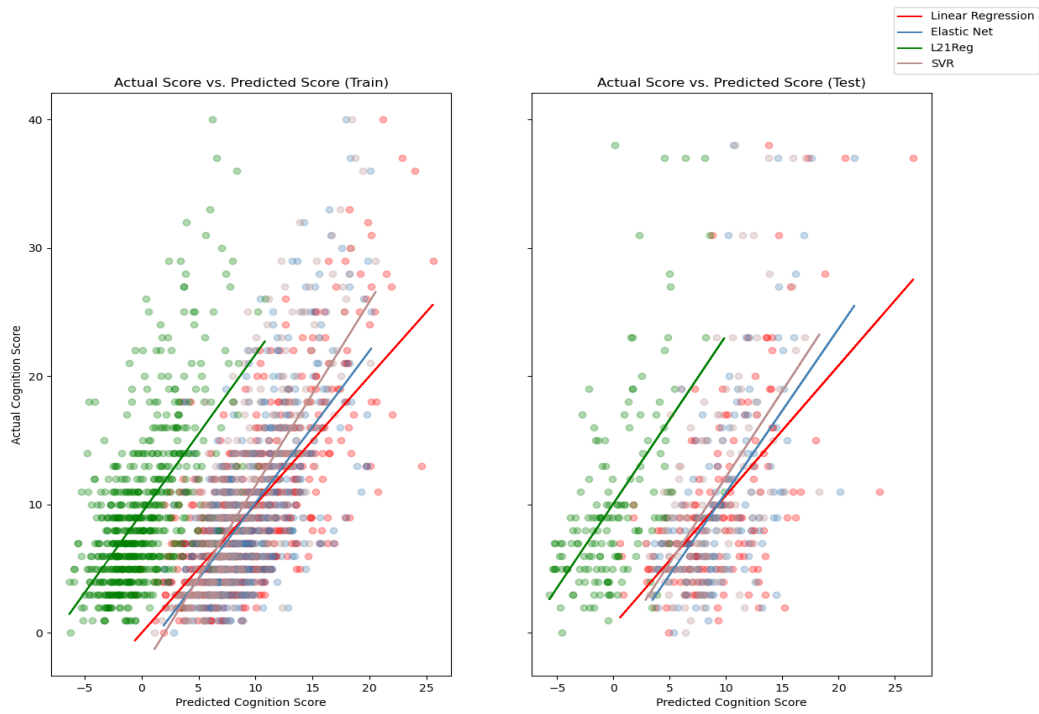


Dataset Counts Summary:	
instances:	787.0
features:	116.0
categorical_features:	0.0
quantitative_features:	116.0
missing_values:	0.0
missing_percent:	0.0

Top ML Algorithm Results (Averaged Over CV Runs):	
Best (Max Error):	Elastic Net = 19.847
Best (Mean Absolute Error):	Elastic Net = 3.968
Best (Mean Squared Error):	Elastic Net = 28.463
Best (Explained Variance):	Elastic Net = 0.306
Best (Median Absolute Error):	SVR = 3.156
Best (Pearson Correlation):	Linear Regression = 0.526



Dataset and Model Prediction Summary: D3 = vbm_ADAS11



Average Model Prediction Statistics (Rounded to 3 Decimal Points)

D1 = av45_ADAS11

ML Algorithm	Max Error	Mean Absolute	Mean Squared	Median Absolute	Explained Variance	Pearson Correlation
Linear Regression	21.428	4.357	33.374	3.505	0.174	0.491
Elastic Net	22.199	4.001	29.396	3.17	0.288	0.544
L21Reg	31.475	9.496	118.951	8.667	0.287	0.544
SVR	23.214	3.948	30.365	2.937	0.279	0.536

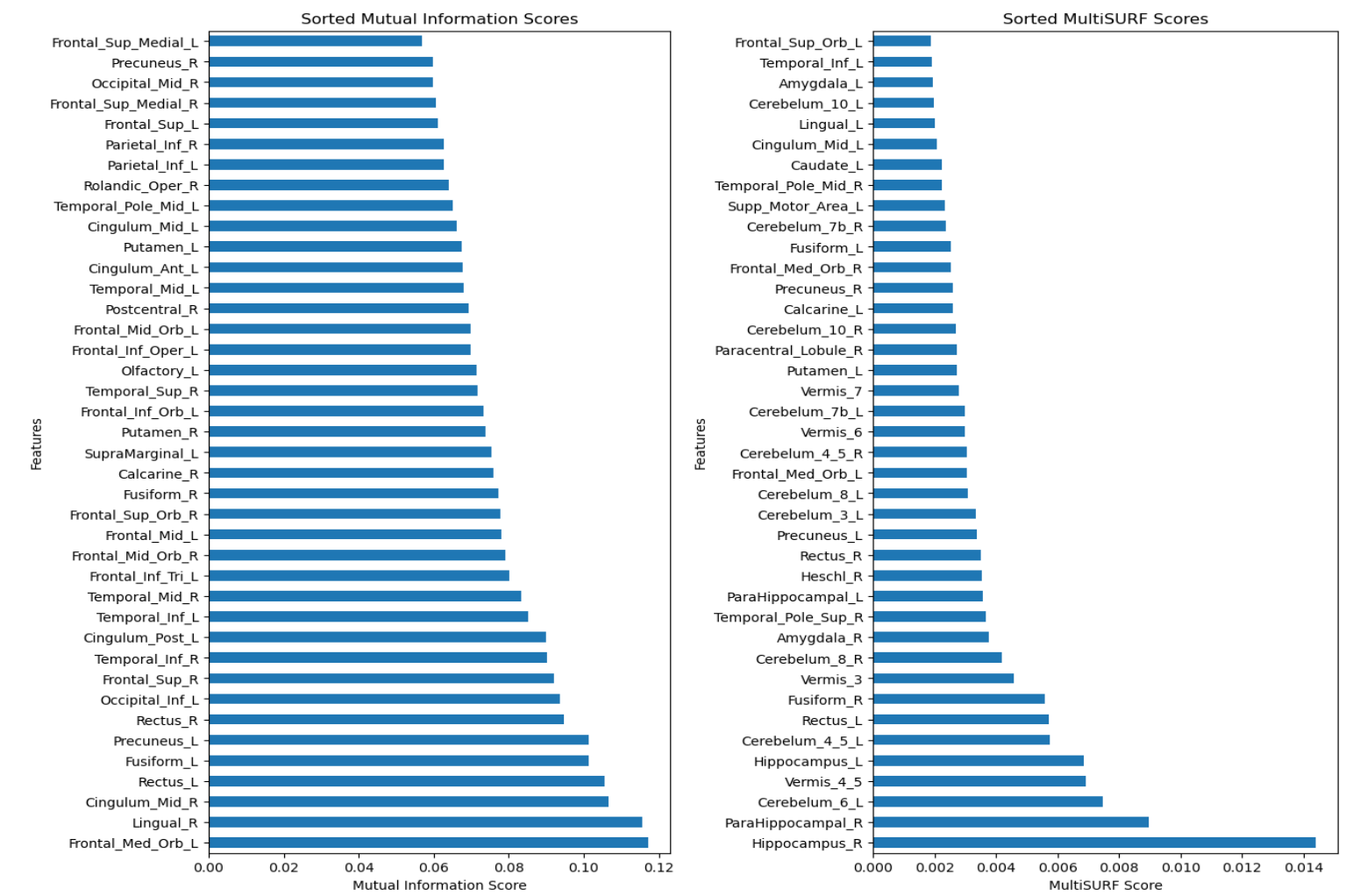
D2 = fdg_ADAS11

ML Algorithm	Max Error	Mean Absolute	Mean Squared	Median Absolute	Explained Variance	Pearson Correlation
Linear Regression	17.29	3.966	26.616	3.268	0.346	0.625
Elastic Net	18.646	3.692	24.608	2.796	0.401	0.644
L21Reg	28.378	9.488	113.793	9.006	0.402	0.644
SVR	18.647	3.737	26.304	2.786	0.363	0.617

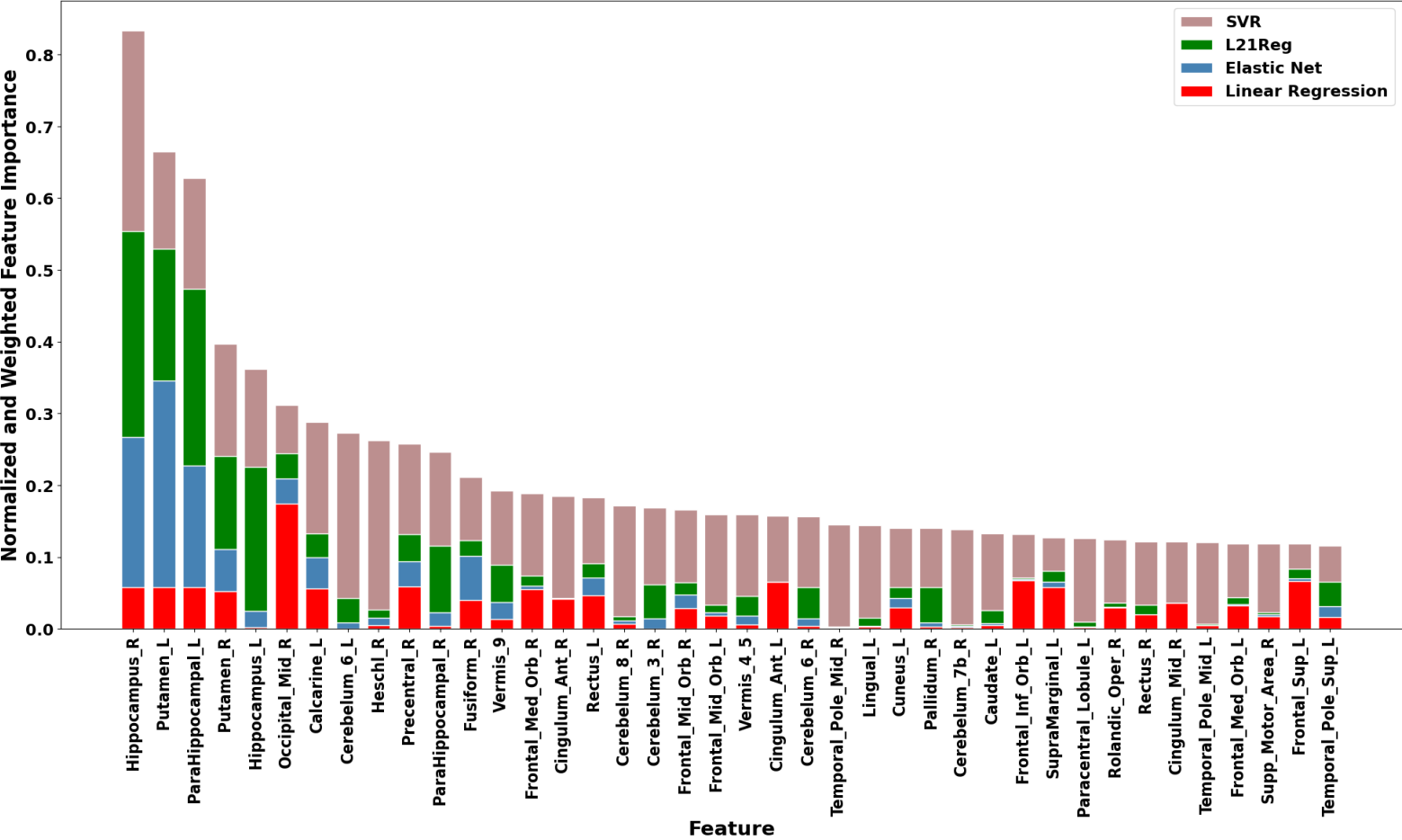
D3 = vbm_ADAS11

ML Algorithm	Max Error	Mean Absolute	Mean Squared	Median Absolute	Explained Variance	Pearson Correlation
Linear Regression	20.356	4.249	31.294	3.302	0.218	0.526
Elastic Net	19.847	3.968	28.463	3.227	0.306	0.562
L21Reg	30.471	9.488	118.762	8.734	0.29	0.544
SVR	21.329	4.025	30.581	3.156	0.272	0.539

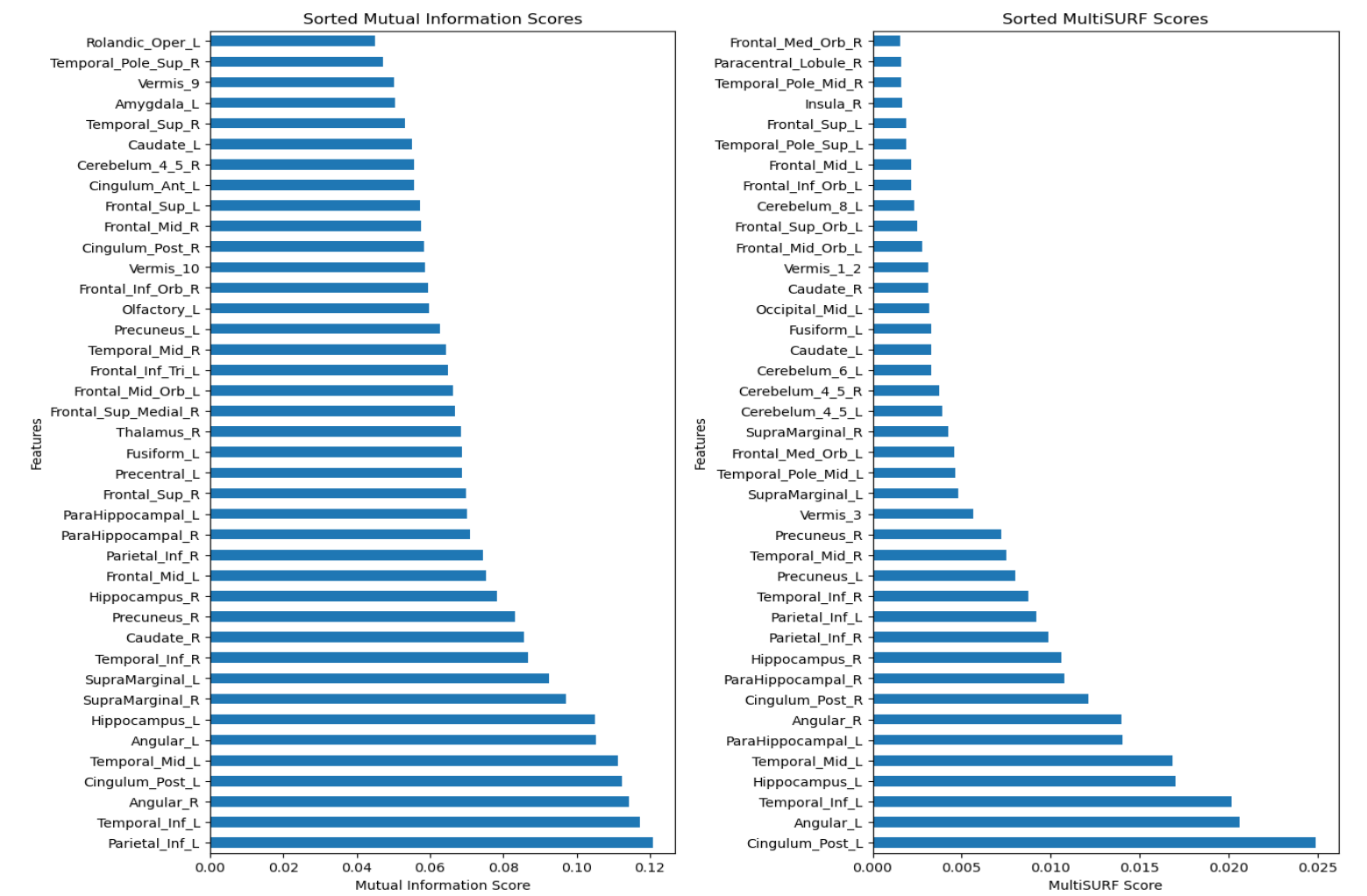
Feature Importance Summary: D1 = av45_ADAS11



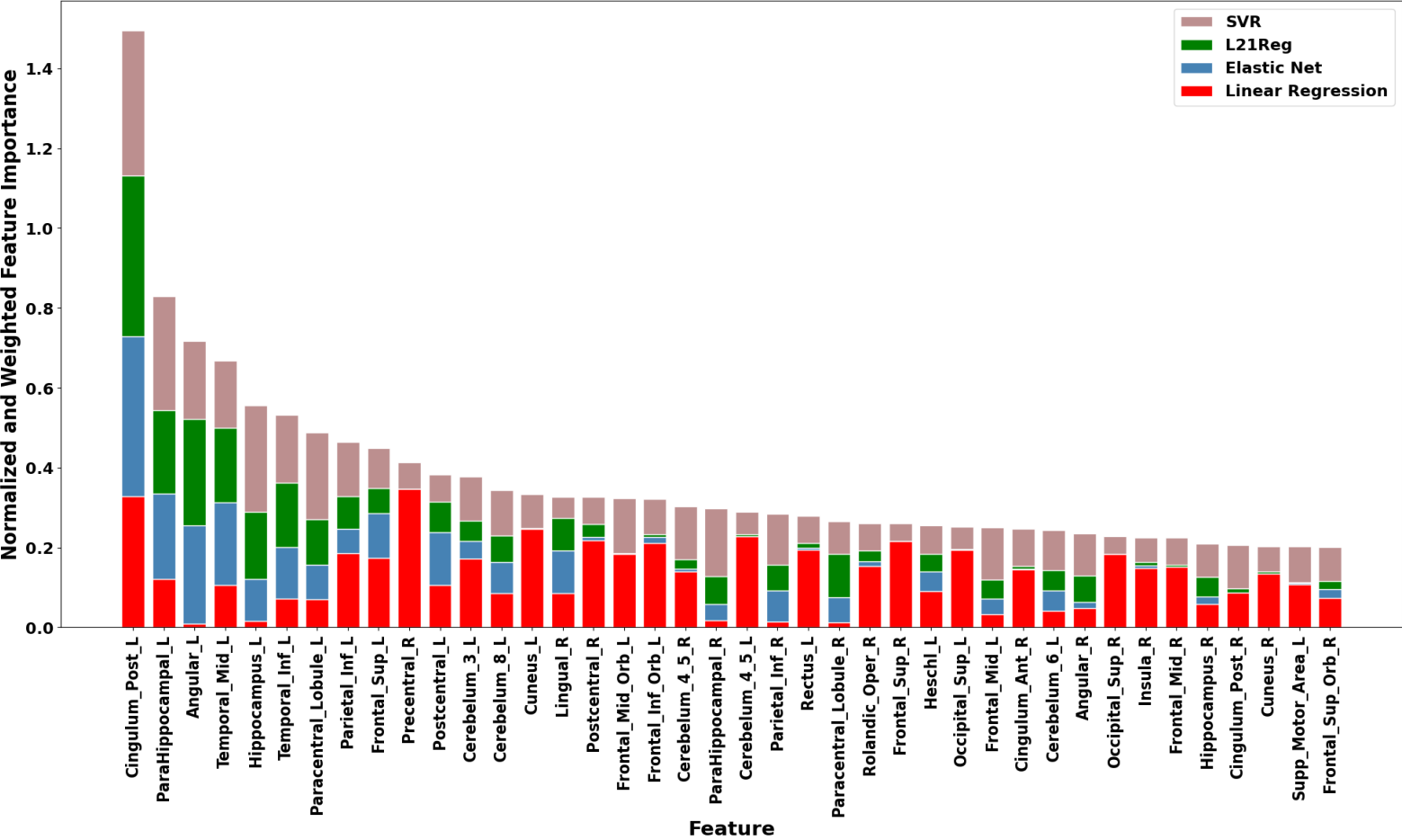
Composite Feature Importance Plot (Normalized and Performance Weighted)



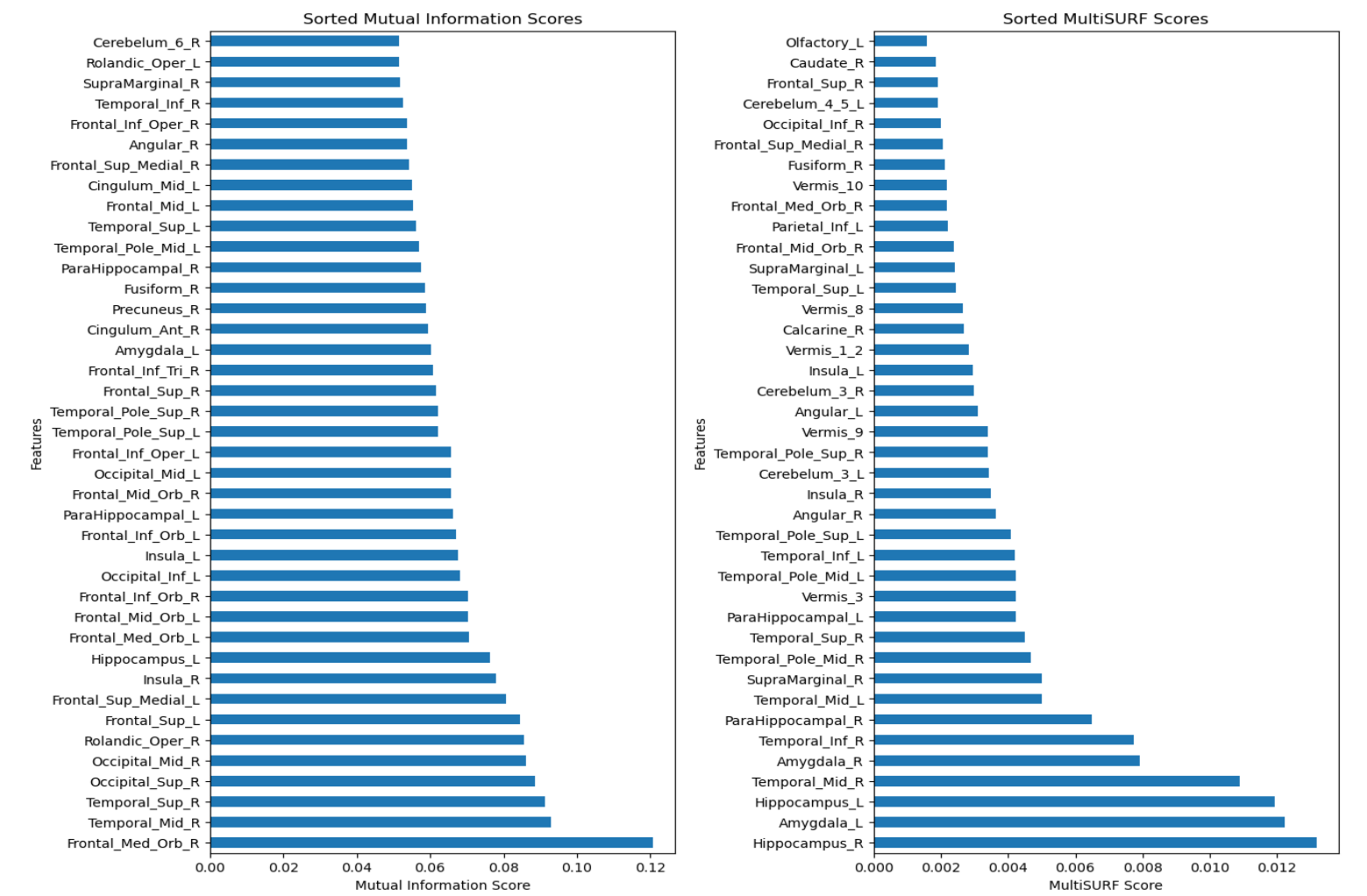
Feature Importance Summary: D2 = fdg_ADAS11



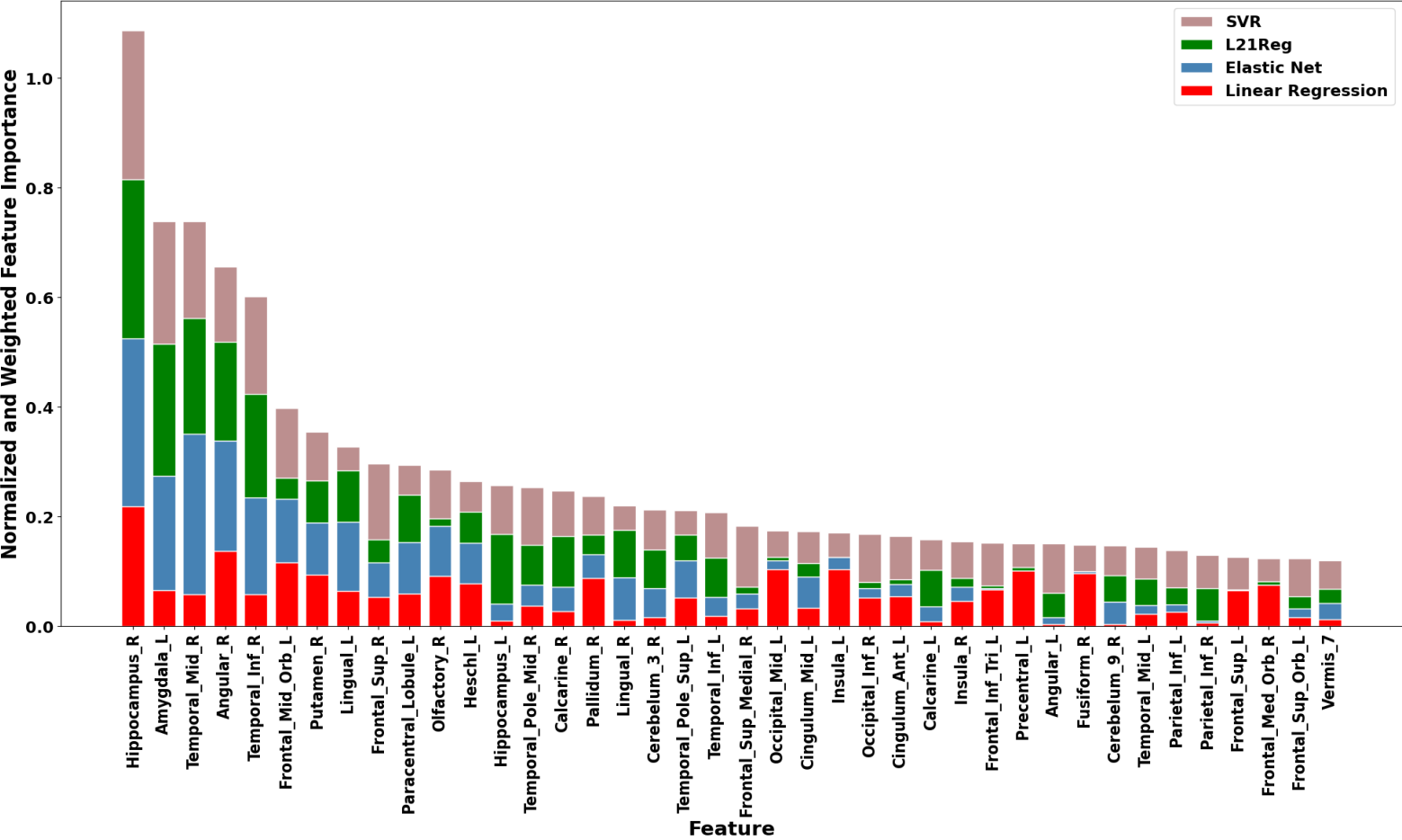
Composite Feature Importance Plot (Normalized and Performance Weighted)



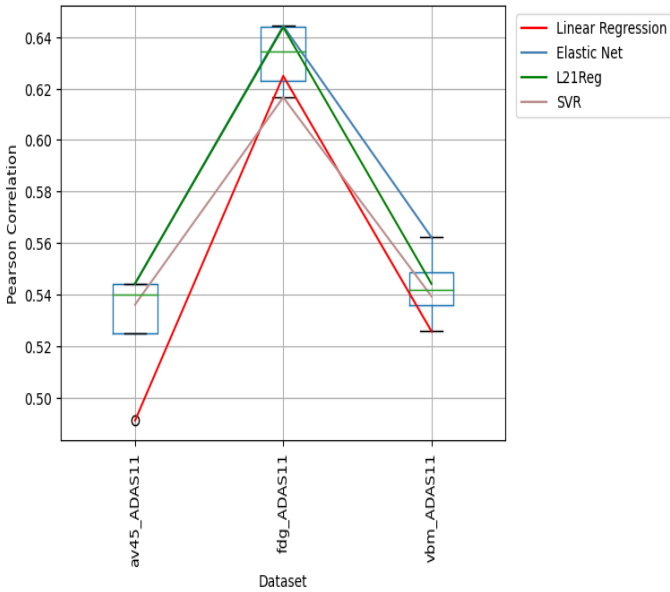
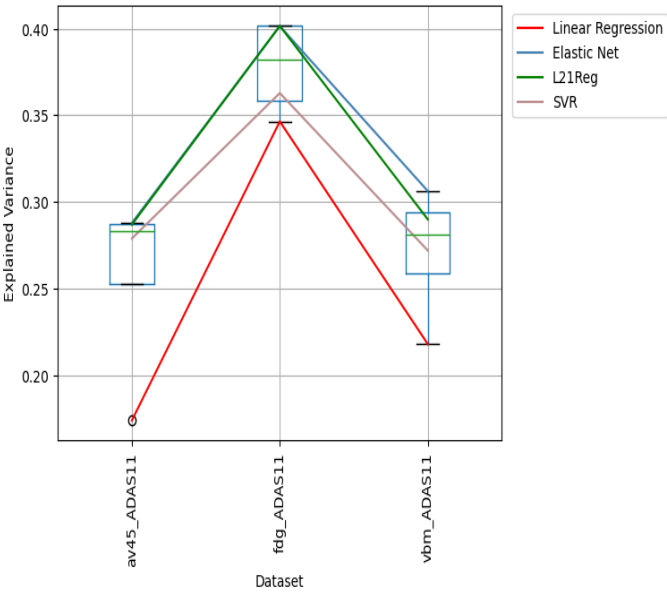
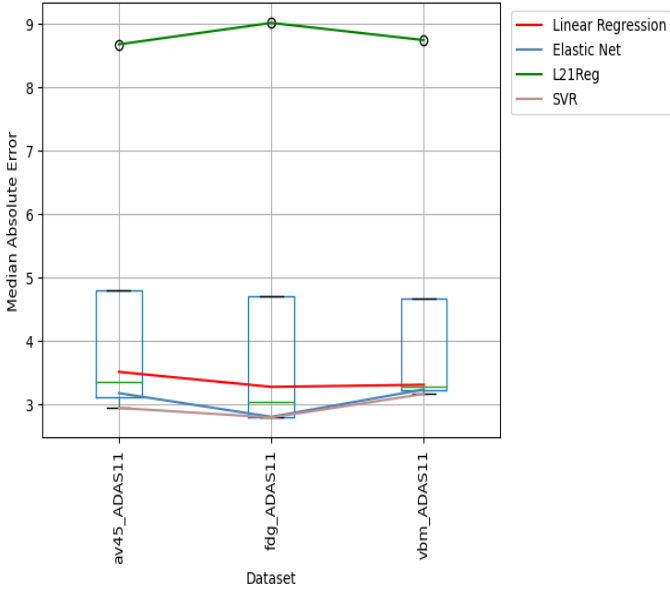
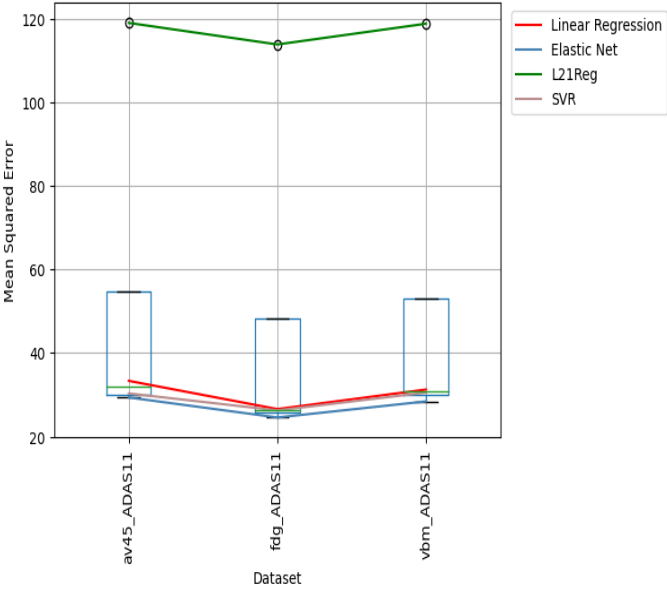
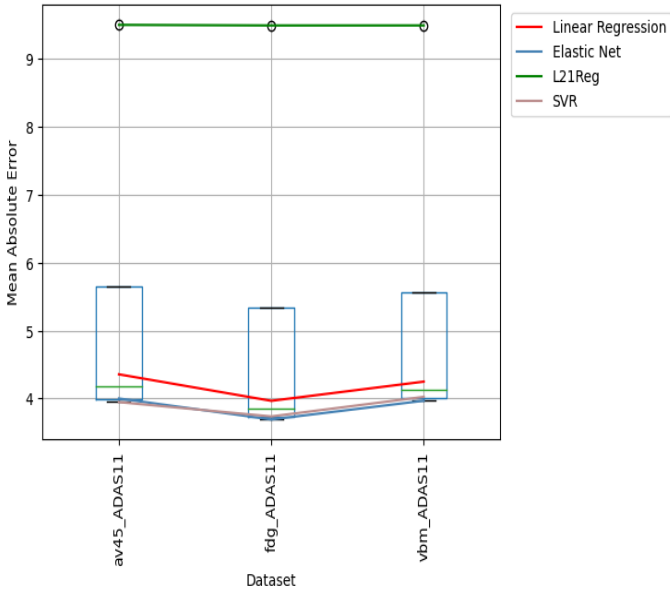
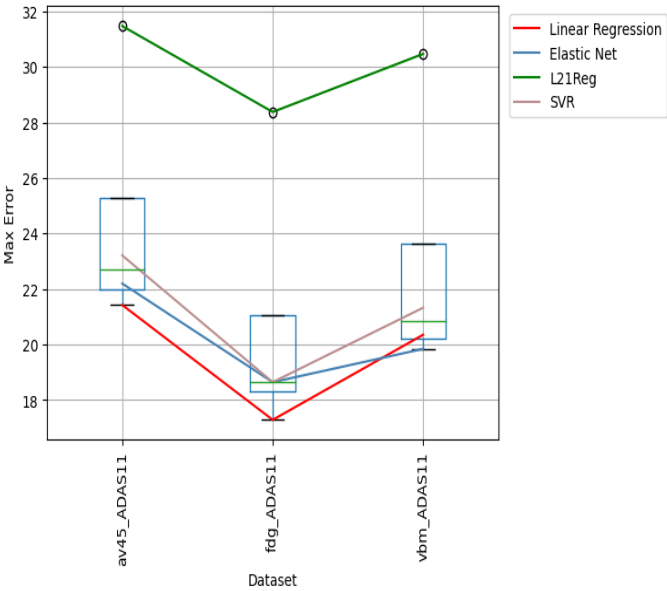
Feature Importance Summary: D3 = vbm_ADAS11



Composite Feature Importance Plot (Normalized and Performance Weighted)



Compare ML Performance Across Datasets



Using Best Performing Algorithms (Kruskall Wallis Compare Datasets)

Datasets:
D1 = av45_ADAS11
D2 = fdg_ADAS11
D3 = vbm_ADAS11

index	P-Value	Best Alg D1	Mean D1	Best Alg D2	Mean D2	Best Alg D3	Mean D3
Max Error	0.4819	L21Reg	31.4748	L21Reg	28.3776	L21Reg	30.4709
Mean Absolute Error	1.0	L21Reg	9.4957	L21Reg	9.4875	L21Reg	9.4879
Mean Squared Error	0.8781	L21Reg	118.9512	L21Reg	113.7927	L21Reg	118.7621
Median Absolute Error	0.6907	L21Reg	8.6667	L21Reg	9.0057	L21Reg	8.734
Explained Variance	0.075	Elastic Net	0.288	L21Reg	0.4015	Elastic Net	0.3064
Pearson Correlation	0.0226	Elastic Net	0.5443	Elastic Net	0.6443	Elastic Net	0.5623

Pipeline Runtime Summary

av45_ADAS11		fdg_ADAS11	
Pipeline Component	Time (sec)	Pipeline Component	Time (sec)
Exploratory Analysis	1.66	Exploratory Analysis	1.49
Preprocessing	0.08	Preprocessing	0.07
Mutual Information	24.11	Mutual Information	24.29
MultiSURF	104.99	MultiSURF	103.99
Feature Selection	0.52	Feature Selection	0.4
Linear Regression	2.69	Linear Regression	43.47
Elastic Net	29.88	Elastic Net	87.87
L21Reg	359.44	L21Reg	359.44
SVR	714.36	SVR	879.44
Stats Summary	108.74	Stats Summary	19.47

vbm_ADAS11	
Pipeline Component	Time (sec)
Exploratory Analysis	1.44
Preprocessing	0.07
Mutual Information	23.92
MultiSURF	107.56
Feature Selection	0.39
Linear Regression	46.2
Elastic Net	119.61
L21Reg	359.44
SVR	771.17
Stats Summary	393.39