A Novel Model for Gamma-ray Source Classification using Automatic Feature Selection

Method: We are developing a novel framework using machine learning techniques with automatic feature selection algorithms for gamma-ray object classification.

Results: Using Random Forest (RF) algorithm for feature selection, we can improve the performance of classifying (a) pulsars/AGNs (accuracy >98%) and (b) young pulsar/MSPs (accuracy > 95%) in comparison with previous work.

Alex, P. Leung[1], Yuze Tong[1], Shengda Luo[1] & David C.Y. Hui[2]

[1]Macau University of Science & Technology

[2]Chungnam National University

1. Conventional classifications require some knowledge of the gamma-ray properties of different classes of objects which can be far from complete in view of the relatively short history of gamma-ray astronomy.
2. Instead of relying on a prior knowledge, automatic classification let the data “speak for themselves” and generate the classification model.
3. In the previous attempt of classifying gamma-ray sources with machine learning techniques (e.g. Saz Parkinson et al. 2016 ApJ 820 8), the power of automatic feature selection has not be fully exploited.
4. By coupling the classifiers with automatic feature selection algorithms, we aim to

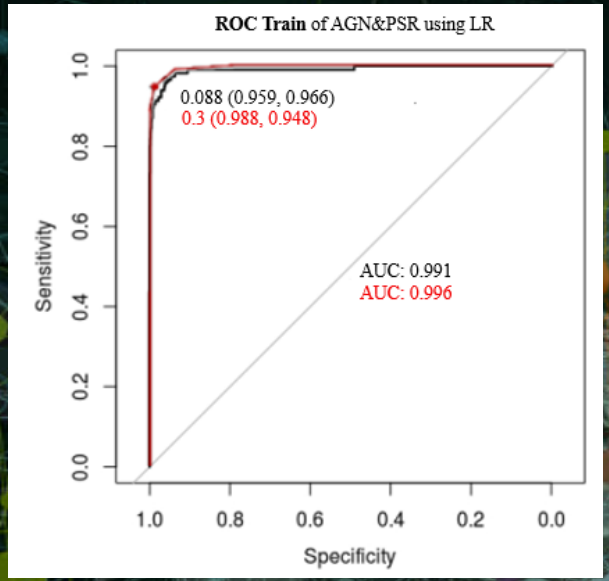
i) Improving the prediction accuracy

ii) Provide a more cost-effective prediction model

iii) Enhancing the discovery power in data mining

Pulsars vs AGNs

Comparisons of ROC curves for PSR/AGN classifications based on our model (red) and Saz Parkinson et al. (2016) (black) using logistic regression (LR) as classifer



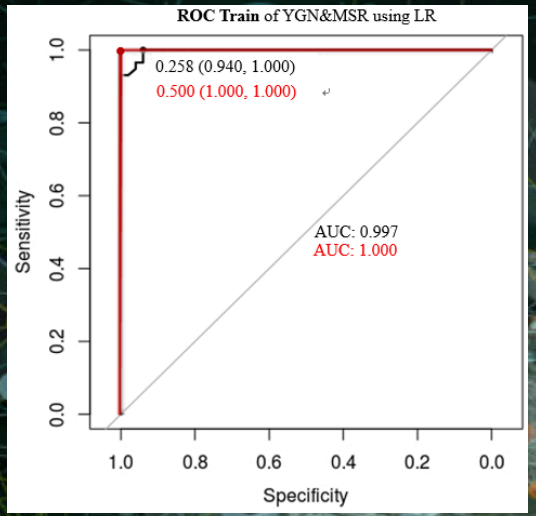
Ranked by importance(AGN&PSR)		
1	Variability_Index	20.868299
2	Signif_Curve	20.620902
3	Spectral_Index	17.407906
4	hr45	15.140595
5	Unc_Flux1000	14.315923
6	SED1000_3000	11.729596
7	Flux1000_3000	11.170445
8	hr23	11.036605
9	Unc_Energy_Flux100	10.177330

Prediction Model	Accuracy (LR)
Saz Parkinson et al. (2016)	94.9%
Our work	98.2%

A set of features automatically selected in our scheme without a prior knowledge and ranked accordingly.

Young Pulsar (YNG) vs Millisecond Pulsar (MSP)

Comparisons of ROC curves for YNG/MSP classification based on our model (red) and Saz Parkinson et al. (2016) (black) using logistic regression (LR) as classifer



Ranked by importance (MSP&YNG)		
1	Unc_Energy_Flux100	23.23109948
2	GLAT	13.57550697
3	Flux_Density	7.73807429
4	Signif_Curve	6.20134667
5	hr34	3.76813769
6	hr23	3.72374297
7	Spectral_Index	1.93633277
8	hr45	1.62792916

Prediction Model	Accuracy (LR)
Saz Parkinson et al. (2016)	90.7%
Our work	95.7%

Instead of 11 features adopted in Saz Parkinson et al. (2016), only 8 features are automatically selected in our scheme and ranked accordingly.