# The University of Hong Kong

Metcalf Summer Internship @ LSR

# Machine Learning and the search for soft $\gamma$-ray pulsars
Progress Report

August 28, 2021

*Author*

Jiatong Li

*Supervisor*

Dr. Pablo Saz Parkinson

# Contents

# 1 Abstract

Our research relies on the fourth catalog of sources (4FGL). Since the Fermi Gamma-ray Space Telescope was launched, the Large Area Telescope (LAT) continuously performs the all-sky survey to study cosmological phenomena. Based on [1], the major improvements of 4FGL are:

- Introduced weights in the maximum likelihood analysis
- Accounted for the effect of energy dispersion. This small correction wasn't presented in the previous catalog

A number of researchers has used machine learning technique to analyze and classify Gamma-Ray sources such as [2]. However, few study has investigated whether we can have an algorithmic method to classify soft gamma-ray pulsars which have energy peaks that is under 1000 MeV. Therefore, our research focuses on using unsupervised learning machine learning to detect whether our 4FGL dataset contains different clusters. Ideally, one of the clusters should be soft gamma-ray pulsars. Then if we get a set of physical properties of a pulsar (e.g. P88Y2536), such as integral photon flux from 1 to 100 GeV and other values, we can effectively predict whether that is a soft gamma pulsar or not.

**Disclaimer**: this progress report contains only a rough step of my research. If you want to check the detailed code/mathematics, please check *Reference* and *Appendix*. If you have any questions about the code or the project, feel free to contact me directly. My email is jtlee@uchicago.edu.

# 2 Background

## 2.1 Introduction to gamma-ray astronomy

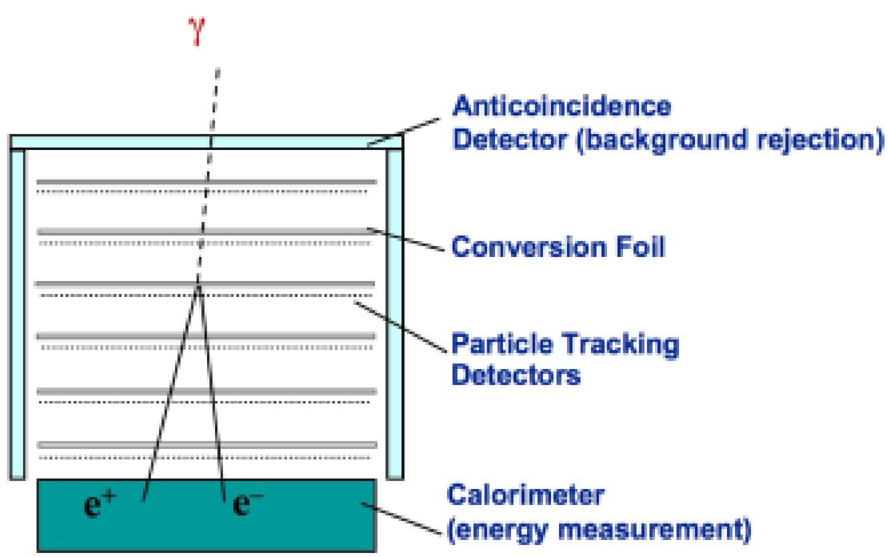


Figure 1: The schematic design of the Fermi LAT (Image: NASA)

According to [1], the 4FGL catalog were taken during the period 2008 Aug 4 to 2016 Aug 2. It covers the data from eight years. The current LAT data is Pass 8 P8R3 [3].

Since I have only 10 weeks to do this research, I can't fully finish it. My job focuses on whether **it is possible to use unsupervised learning algorithm to classify soft gamma-ray pulsars** by conducting Exploratory Data Analysis (EDA) and statistical modeling. If other students join the group in the future, they can pick up my work and run the algorithm.

Our key targets are pulsars which have energy peaks that are below 1000 MeV.

# 3 Methodology

## 3.1 Spatial Join and Match

In this research, our data are primarily from three sources.

- 4FGL source catalog ( 6500 rows)
- ATNF Pulsar Catalogue ( 3000 rows)
- Known soft gamma-ray pulsar ( 300 rows)

Obviously, any pulsar may appear in multiple datasets. A crucial step is to match and identify puslar. We used an astronomical data analysis software called Topcat to perform spatial join and match. The result can be summarized in one Venn diagram.

Each group has its own interesting behavior. For example, pulsars in Group 3 means the pulsar is identified in the 4FGL catalog, but they are not detected in the "known list" of soft gamma-ray pulsar. An interesting question to ask is "why this happens?". Perhaps the known list is obsolete. The documents of the seven groups can be found in Appendix (2).

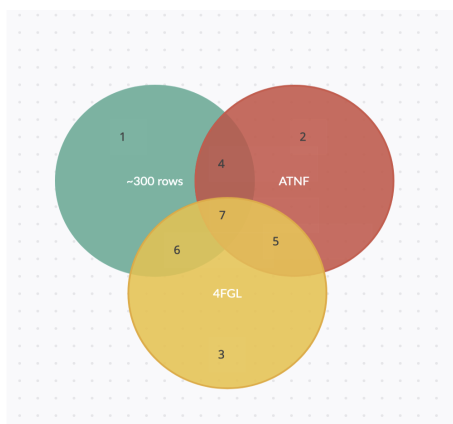


Figure 2

## 3.2 Feature Selection

The original dataset contains 79 columns. We need to first reduce the dimension so that the features (columns) in the training dataset are more relevant to the target variable (e.g. non-GeV pulsar) we want to investigate. (check appendex for the source code)

The algorithm I used is called Recursive Feature Elimination (RFE). Here is a list of important variables we selected.

1. Flux1000
2. PL_Index
3. LP_Index

4. LP_SigCurv
5. PLEC_Exp_Index
6. PLEC_Flux_Density
7. Frac_Variability

Note: mathmetical definition of the aforementioned variables can be found at [1].

### 3.3 Data pre-processing and Data Exploration

After we got a list of target variables, our objective is to plot pairplot and visualize them. If the clusters have clear separations, then we can concludes that running unsupervised algorithm is possible.

The first step is to drop all the irrelevant variables (columns). Next, we need to eliminate rows that contain NULL value. Otherwise, we will get errors during the data visualization process. After such eliminations, we got 5584 rows.

One of the variables in the 4FGL catalog is called "CLASS1" which indicates the class designation for the associated sources. For example, "PSR" stands for "Pulsar" and "SNR" stands for "Supernova remnant". We need to keep this variable in our dataframe so that we can investigate each subclass. In order to help us focus on soft gamma-pulsars. I created a new label called "Soft" which means the pulsar of that row has an energy peak that is below 1000 GeV since this group is what we are interested in.

### 3.4 Distribution Histogram

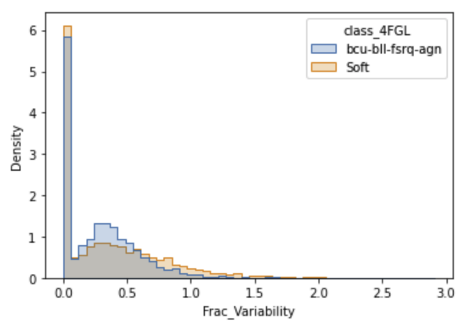


Figure 3

I have developed a script which can help us inspect any variables with respect to the class designation. For example, in Figure 3, we plotted the distribution of "Frac_Variability" with two subclasses we are interested in "bcu-bll-fsrq-agn" and "Soft". We can inspect whether the two subclasses are separated or not.

### 3.5 Pairplot

In EDA, I plotted the pair plot in a grid. Essentially, we put each variable in the x-axis and y-axis. The diagonal direction is little bit different. It shows the distribution histogram of that variable. Check the Appendix (2) for the full picture. Since our dataset contains lots of data points. Plotting them all at once will consume to much CPU and computing resources. I preprocessed the data and reduced its dimension. In order to plot all the rows, it is more appropriate to use GPU resources.

## 4 Conclusions

Through the pairplot presented in the Appendix, we can clearly see that there are separations between each subclass. Therefore, we can conclude that using unsupervised machine learning algorithm is possible. Moreover, by doing the EDA, we gained some interested insights. The next steps are:

1. Using GPU to plot the full pair scatter plot.
2. Running unsupervised learning algorithm on the dataset. Quality control and check whether the result makes sense.

## 5 Acknowledgement

It is my great pleasure to thank my advisor Dr. Pablo Saz Parkinson for all the help.

# 6 Appendix A: Code

1. Source code github repo
2. Data folder

## References


[1] S. Abdollahi, F. Acero, M. Ackermann, M. Ajello, W. B. Atwood, M. Axelsson, L. Baldini, J. Ballet, G. Barbiellini, D. Bastieri, and et al., “Fermi large area telescope fourth source catalog,” *The Astrophysical Journal Supplement Series*, vol. 247, no. 1, p. 33, Mar 2020. [Online]. Available: http://dx.doi.org/10.3847/1538-4365/ab6bcb

[2] P. M. Saz Parkinson, H. Xu, P. L. H. Yu, D. Salvetti, M. Marelli, and A. D. Falcone, “Classification and ranking offermilat gamma-ray sources from the 3fgl catalog using machine learning techniques,” *The Astrophysical Journal*, vol. 820, no. 1, p. 8, Mar 2016. [Online]. Available: http://dx.doi.org/10.3847/0004-637X/820/1/8

[3] W. Atwood, A. Albert, L. Baldini, M. Tinivella, J. Bregeon, M. Pesce-Rollins, C. Sgrò, P. Bruel, E. Charles, A. Drlica-Wagner, A. Franckowiak, T. Jogler, L. Rochester, T. Usher, M. Wood, J. Cohen-Tanugi, and S. Zimmer, “Pass 8: Toward the full realization of the fermi-lat scientific potential,” 2013.

[4] W. H. L. Kuiper, “The soft $\gamma$-ray pulsar population: a high-energy overview,” *Monthly Notices of the Royal Astronomical Society*, vol. 449, p. 3827–3866, June 2020. [Online]. Available: https://doi.org/10.1093/mnras/stv426

[5] A. A. Abdo, M. Ajello, A. Allafort, L. Baldini, J. Ballet, G. Barbiellini, M. G. Baring, D. Bastieri, A. Belfiore, R. Bellazzini, B. Bhattacharyya, E. Bissaldi, E. D. Bloom, E. Bonamente, E. Bottacini, T. J. Brandt, J. Bregeon, M. Brigida, P. Bruel, R. Buehler, M. Burgay, T. H. Burnett, G. Busetto, S. Buson, G. A. Caliandro, R. A. Cameron, F. Camilo, P. A. Caraveo, J. M. Casandjian, C. Cecchi, Çelik, E. Charles, S. Chaty, R. C. G. Chaves, A. Chekhtman, A. W. Chen, J. Chiang, G. Chiaro, S. Ciprini, R. Claus, I. Cognard, J. Cohen-Tanugi, L. R. Cominsky, J. Conrad, S. Cutini, F. D'Ammando, A. de Angelis, M. E. DeCesar, A. D. Luca, P. R. den Hartog, F. de Palma, C. D. Dermer, G. Desvignes, S. W. Digel, L. D. Venere, P. S. Drell, A. Drlica-Wagner, R. Dubois, D. Dumora, C. M. Espinoza, L. Falletti, C. Favuzzi, E. C. Ferrara, W. B. Focke, A. Franckowiak, P. C. C. Freire, S. Funk, P. Fusco, F. Gargano, D. Gasparrini, S. Germani, N. Giglietto, P. Giommi, F. Giordano, M. Giroletti, T. Glanzman, G. Godfrey, E. V. Gotthelf, I. A. Grenier, M.-H. Grondin, J. E. Grove, L. Guillemot, S. Guiriec, D. Hadasch, Y. Hanabata, A. K. Harding, M. Hayashida, E. Hays, J. Hessels, J. Hewitt, A. B. Hill, D. Horan, X. Hou, R. E. Hughes, M. S. Jackson, G. H. Janssen, T. Jogler, G. Jóhannesson, R. P. Johnson, A. S. Johnson, T. J. Johnson, W. N. Johnson, S. Johnston, T. Kamae, J. Kataoka, M. Keith, M. Kerr, J. Knödlseder, M. Kramer, M. Kuss, J. Lande, S. Larsson, L. Latronico, M. Lemoine-Goumard, F. Longo, F. Loparco, M. N. Lovellette, P. Lubrano, A. G. Lyne, R. N. Manchester, M. Marelli, F. Massaro, M. Mayer, M. N. Mazziotta, J. E. McEnery, M. A. McLaughlin, J. Mehault, P. F. Michelson, R. P. Mignani, W. Mitthumsiri, T. Mizuno, A. A. Moiseev, M. E. Monzani, A. Morselli, I. V. Moskalenko, S. Murgia, T. Nakamori, R. Nemmen, E. Nuss, M. Ohno, T. Ohsugi, M. Orienti, E. Orlando, J. F. Ormes, D. Paneque, J. H. Panetta, D. Parent, J. S. Perkins, M. Pesce-Rollins, M. Pierbattista, F. Piron, G. Pivato, H. J. Pletsch, T. A. Porter, A. Possenti, S. Rainò, R. Rando, S. M. Ransom, P. S. Ray, M. Razzano, N. Rea, A. Reimer, O. Reimer, N. Renault, T. Reposeur, S. Ritz, R. W. Romani, M. Roth, R. Rousseau, J. Roy, J. Ruan, A. Sartori, P. M. S. Parkinson, J. D. Scargle, A. Schulz, C. Sgrò, R. Shannon, E. J. Siskind, D. A. Smith, G. Spandre, P. Spinelli, B. W. Stappers, A. W. Strong, D. J. Suson, H. Takahashi, J. G. Thayer, J. B. Thayer, G. Theureau, D. J. Thompson, S. E. Thorsett, L. Tibaldo, O. Tibolla, M. Tinivella, D. F. Torres, G. Tosti, E. Troja, Y. Uchiyama, T. L. Usher, J. Vandenbroucke, V. Vasileiou, C. Venter, G. Vianello, V. Vitale, N. Wang, P. Weltevrede, B. L. Winer, M. T. Wolff, D. L. Wood, K. S. Wood, M. Wood, and Z. Yang, “The second fermi large

area telescope catalog of gamma-ray pulsars," *The Astrophysical Journal Supplement Series*, vol. 208, no. 2, p. 17, sep 2013. [Online]. Available: https://doi.org/10.1088/0067-0049/208/2/17