

# ClassiPyGRB: Machine Learning-Based Classification and Visualization of Gamma Ray Bursts using t-SNE

Keneth Garcia-Cifuentes<sup>1\*</sup>, Rosa L. Becerra<sup>1\*</sup>, and Fabio De Colle<sup>1</sup>

<sup>1</sup> Instituto de Ciencias Nucleares, Universidad Nacional Autónoma de México, Apartado Postal 70-543, 04510 CDMX, México ¶ Corresponding author \* These authors contributed equally.

DOI: [10.xxxxxx/draft](https://doi.org/10.xxxxxx/draft)

## Software

- [Review](#)
- [Repository](#)
- [Archive](#)

Editor: [Dan Foreman-Mackey](#)

## Reviewers:

- [@wkerzendorf](#)
- [@dfm](#)

Submitted: 22 May 2023

Published: unpublished

## License

Authors of papers retain copyright and release the work under a Creative Commons Attribution 4.0 International License ([CC BY 4.0](#)).

## Summary

Gamma-ray burst (GRBs) are the brightest events in the universe. For decades, astrophysicists have known about their cosmological nature. Every year, space missions such as Fermi and SWIFT detect hundreds of them. In spite of this large sample, GRBs show a complex taxonomy in the first seconds after their appearance, which makes it very difficult to find similarities between them using conventional techniques.

It is known that GRBs originate from the death of a massive star or from the merger of two compact objects. GRB classification is typically based on the duration of the burst (Kouveliotou et al., 1993). Nevertheless, events such as GRB 211211A (Yang et al., 2022), whose duration of about 50 seconds lies in the group of long GRBs, has challenged this categorization by the evidence of features related with the short GRB population (the kilonova emission and the properties of its host galaxy). Therefore, a classification based only on their gamma-ray duration does not provide a completely reliable determination of the progenitor.

Motivated by this problem, Jespersen et al. (2020) and Steinhardt et al. (2023) carried out analysis of GRB light curves by using the t-SNE algorithm, showing that Swift/BAT GRBs database, consisting of light curves in four energy bands (15-25 keV, 25-50 keV, 50-100 keV, 100-350 keV), clusters into two groups corresponding with the typical long/short classification. However, in this case, this classification is based on the information provided by their gamma-ray emission light curves.

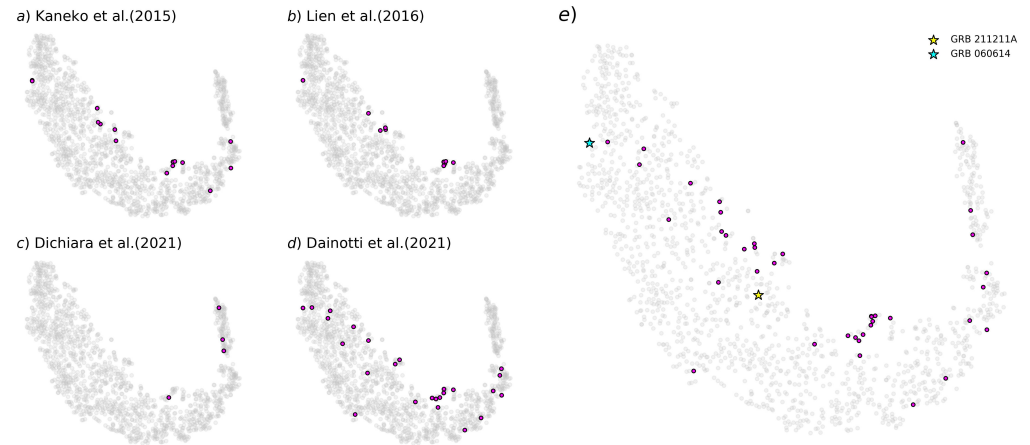
**ClassiPyGRB** is a Python 3 package to download, process, visualize and classify GRBs database from the [Swift/BAT Instrument](#) (up to July 2022). It is distributed over the GNU General Public License Version 2 (1991). We also included a noise-reduction and an interpolation tools for achieving a deeper analysis of the data.

## Statement of need

The Swift Burst Alert Telescope (BAT) is a wide-field, hard Gamma-ray detector that, since its launch in 2004, has played an important role in the high-energy astrophysical field. A preliminary query on the Astrophysics Data System (ADS) with the keyword "Swift/BAT" indicates that over 7000 research works have been uploaded to its platform (up to December 2023). Furthermore, the number of studies per year is increasingly, evidencing the relevance and impact of the use of the Swift/BAT database.

Although the Swift/BAT database is publicly available, for first-time users it might be a challenge to download and process the data. The data are stored in multiple directories, depending on the GRB. Moreover, the data are not pre-processed, and the user must perform the data manipulation and interpolation themselves. These issues make the data processing a

time-consuming task. ClassiPyGRB is a Python 3 package that aims to solve these problems by providing a simple and intuitive interface to download, process, visualize, and classify the photometric data of GRBs from the Swift/BAT database. ClassiPyGRB can also be used to promptly find similar GRBs with some specific feature, such as a bright component (Angulo-Valdez et al., 2024).



**Figure 1:** t-SNE visualization map obtained for the noise-reduced dataset binned at 64 ms with  $pp = 30$ . GRBs colored in magenta are classified as Extended Emission by previous works. Image taken from Garcia-Cifuentes et al. (2023)

ClassiPyGRB allows researchers to query light curves for any GRB in the Swift/BAT database simply and intuitively. The package also provides a set of tools to preprocess the data, including noise/duration reduction and interpolation. Moreover, The package also provides a set of facilities and tutorials to classify GRBs based on their light curves, following the method proposed by Jespersen et al.(2020) and Garcia-Cifuentes et al.(2023) (see e.g., Figure 1). This method is based on dimensionality reduction of the data using t-distributed Stochastic Neighbour Embedding (t-SNE), where the user can visualize the results using a Graphical User Interface (GUI). The user can also plot and animate the results of the t-SNE analysis, allowing to perform a deeper hyperparameter grid search. The package is distributed over the GNU General Public Licence Version 2 (1991).

Although Swift/BAT offers basic functionality for data retrieval and analysis, it is still necessary to implement a package that allows the user to access and use the data in a simple way. In our case, ClassiPyGRB is completely focused on GRB science. It complements what other Python packages oriented to the Swift/BAT instrument do, such as:

- 1) The NITRATES pipeline (DeLaunay & Tohuvavohu, 2022), designed for maximum likelihood-driven discovery and localization of Gamma Ray Bursts.
- 2) BatAnalysis (Parsotan et al., 2023), a package specializing in photometry from all sources observed by the BAT instrument, not just in GRBs.

Thus, ClassiPyGRB distinguishes itself by offering a comprehensive solution that addresses the entire workflow of GRBs, from data retrieval to classification. This package facilitates the entire process of GRB analysis, ensuring accessibility, efficiency, and robustness for researchers in the field of gamma-ray astronomy.

## Methodology and Structure of ClassiPyGRB

ClassiPyGRB mainly consists of three parts:

69 1) Retrieval and visualization of data from Swift/BAT: We implement an easy and fast  
70 code to download and plot an existing GRB post-processed data (e.g., Figure 2). There  
71 is the possibility to modify the time resolution of the data files (2ms, 8ms, 16ms, 64ms,  
72 256ms, 1s and 10s) and analyze the data by selecting only some of the energy bands.

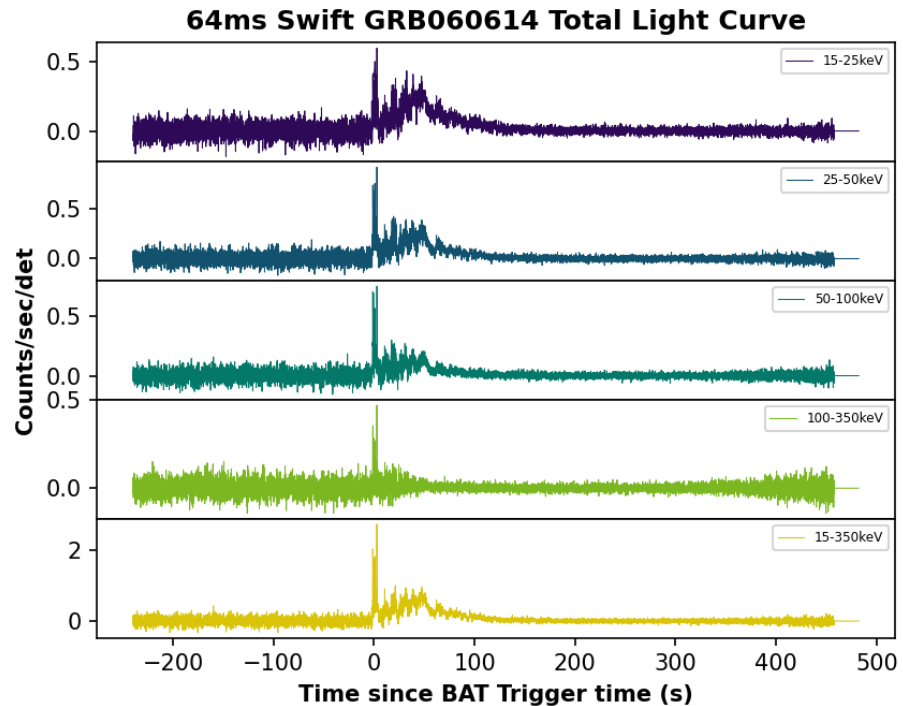


Figure 2: Light curve of GRB 060614A. Image taken from Garcia-Cifuentes et al. (2023)

- 73 2) Data processing. ClassiPyGRB is able to:
- 74 ■ Constrain the light curves by their duration  $T_{100}$ ,  $T_{90}$  or  $T_{50}$ .
  - 75 ■ Normalize the flux.
  - 76 ■ Standardize the temporal interval of all the sample (by zero-padding).
  - 77 ■ Improve the signal/noise (S/N) ratio applying the non-parametric noise reduction technique called FABADA (Sanchez-Alarcon & Ascasibar Sequeiros, 2022) to each band for every single light curve.
  - 78 ■ Interpolate the flux between two specific times.
  - 79
  - 80

81 3) Visualization and plotting of t-SNE maps. ClassiPyGRB produces 2D visualization maps colored by the duration of GRBs.

    - 82
    - 83 ■ It includes an intuitive graphic interface.
    - 84 ■ It is possible to add either of the two features to the t-SNE maps or to visualize the raw data.
    - 85
    - 86 ■ Manipulation of the t-SNE parameters.
    - 87 ■ Visualization of the light curves from the events with and without noise-reduction.
    - 88 ■ There is the possibility of working only on the desired bands of Swift/BAT.
    - 89 ■ Specific events can be searched for and highlighted on the display.

90 Any plots created with ClassiPyGRB can be customized by the user.

91 Note: Algorithms such as t-SNE visualization maps are very sensitive to any change in the  
92 perplexity and learning rate parameters. Therefore, as is the case when using any of these

93 visualization techniques derived from machine learning, care must be taken in the correct  
94 interpretation of the data.

95 This repository requires Python 3.8 or higher, and the necessary packages will be automatically  
96 handled during installation. Other packages will be required optionally in Documentation (i.e.,  
97 Jupyter).

## 98 Acknowledgements

99 KSGC acknowledges support from CONACyT fellowship. RLB acknowledges support from  
100 CONACyT postdoctoral fellowship.

## 101 References

- 102 Angulo-Valdez, C., Becerra, R. L., Pereyra, M., Garcia-Cifuentes, K., Vargas, F., Watson, A. M.,  
103 De Colle, F., Fraija, N., Butler, N. R., Dainotti, M. G., Dichiara, S., Lee, W. H., Troja, E.,  
104 Bloom, J. S., González, J. J., Kuttyrev, A. S., Prochaska, J. X., Ramirez-Ruiz, E., & Richer,  
105 M. G. (2024). Machine-learning enhanced photometric analysis of the extremely bright  
106 GRB 210822A. *Monthly Notices of the Royal Astronomical Society*, 527(3), 8140–8150.  
107 <https://doi.org/10.1093/mnras/stad3624>
- 108 DeLaunay, J., & Tohuvavohu, A. (2022). Harvesting BAT-GUANO with NITRATES (Non-  
109 Imaging Transient Reconstruction and Temporal Search): Detecting and Localizing the  
110 Faintest Gamma-Ray Bursts with a Likelihood Framework. *The Astrophysical Journal*,  
111 941(2), 169. <https://doi.org/10.3847/1538-4357/ac9d38>
- 112 Garcia-Cifuentes, K., Becerra, R. L., De Colle, F., Cabrera, J. I., & del Burgo, C. (2023).  
113 Identification of Extended Emission Gamma-Ray-Bursts Candidates using Machine Learning.  
114 *arXiv e-Prints*, arXiv:2304.08666. <https://doi.org/10.48550/arXiv.2304.08666>
- 115 Jespersen, C. K., Severin, J. B., Steinhardt, C. L., Vinther, J., Fynbo, J. P. U., Selsing,  
116 J., & Watson, D. (2020). An Unambiguous Separation of Gamma-Ray Bursts into  
117 Two Classes from Prompt Emission Alone. *Astrophysical Journal Letters*, 896(2), L20.  
118 <https://doi.org/10.3847/2041-8213/ab964d>
- 119 Kouveliotou, C., Meegan, C. A., Fishman, G. J., Bhat, N. P., Briggs, M. S., Koshut, T. M.,  
120 Paciesas, W. S., & Pendleton, G. N. (1993). Identification of Two Classes of Gamma-Ray  
121 Bursts. *Astrophysical Journal Letters*, 413, L101. <https://doi.org/10.1086/186969>
- 122 Parsotan, T., Laha, S., Palmer, D. M., Lien, A., Cenko, S. B., Krimm, H., & Markwardt, C.  
123 (2023). BatAnalysis: A Comprehensive Python Pipeline for Swift BAT Survey Analysis.  
124 *The Astrophysical Journal*, 953(2), 155. <https://doi.org/10.3847/1538-4357/ace325>
- 125 Sanchez-Alarcon, P. M., & Ascasibar Sequeiros, Y. (2022). Fully Adaptive Bayesian Algorithm  
126 for Data Analysis, FABADA. *arXiv e-Prints*, arXiv:2201.05145. <https://doi.org/10.48550/arXiv.2201.05145>
- 127
- 128 Steinhardt, C. L., Mann, W. J., Rusakov, V., & Jespersen, C. K. (2023). Classification  
129 of BATSE, Swift, and Fermi Gamma-Ray Bursts from Prompt Emission Alone. *The*  
130 *Astrophysical Journal*, 945(1), 67. <https://doi.org/10.3847/1538-4357/acb999>
- 131 Yang, J., Ai, S., Zhang, B.-B., Zhang, B., Liu, Z.-K., Wang, X. I., Yang, Y.-H., Yin, Y.-H., Li,  
132 Y., & Lü, H.-J. (2022). A long-duration gamma-ray burst with a peculiar origin. *Nature*,  
133 612(7939), 232–235. <https://doi.org/10.1038/s41586-022-05403-8>