



RoBo6: Standardized MMT Light Curve Dataset for Rocket Body Classification

Daniel Kyselica

Department of Applied Informatics
Comenius University Bratislava
842 48 Bratislava, Slovakia
daniel.kyselica@fmph.uniba.sk

Marek Šuppa

Department of Applied Informatics
Comenius University Bratislava
842 48 Bratislava, Slovakia
marek.suppa@fmph.uniba.sk

Jiří Šilha

Division of Astronomy and Astrophysics
Comenius University Bratislava
842 48 Bratislava, Slovakia
jiri.silha@fmph.uniba.sk

Roman Ďurikovič

Department of Applied Informatics
Comenius University Bratislava
842 48 Bratislava, Slovakia
roman.durikovic@fmph.uniba.sk

Abstract

Space debris presents a critical challenge for the sustainability of future space missions, emphasizing the need for robust and standardized identification methods. However, a comprehensive benchmark for rocket body classification remains absent. This paper addresses this gap by introducing the RoBo6 dataset for rocket body classification based on light curves. The dataset, derived from the Mini Mega Tortora database, includes light curves for six rocket body classes: CZ-3B, Atlas 5 Centaur, Falcon 9, H-2A, Ariane 5, and Delta 4. With 5,676 training and 1,404 test samples, it addresses data inconsistencies using resampling, normalization, and filtering techniques. Several machine learning models were evaluated, including CNN and transformer-based approaches, with Astroconformer reporting the best performance. The dataset establishes a common benchmark for future comparisons and advancements in rocket body classification tasks.

1 Introduction

Humans have been utilizing the space for over 60 years for various purposes, launching more than 6000 missions [1]. The consequence of this is the increasing amount of space debris, which became a serious threat to future missions [2].

Each object reflects light, creating a measurable signal called a light curve, which serves as its unique footprint. The light curve is shaped by the object's shape, reflectivity, geometry, surface, and rotation thus encoding the objects appearance. Although traditional methods extract limited details from light curves, some potentially crucial parameters are often overlooked.

Machine learning can help scientists identify hazardous and unknown objects more efficiently than traditional methods, enhancing space safety [3]. Studies like [4] and [5] demonstrate its ability to distinguish between rocket bodies. However, existing solutions are not directly comparable. Models are often evaluated solely on proprietary datasets, and even when publicly accessible datasets are used, variations in sample filtration, preprocessing, normalization, and evaluation methods can lead to substantial inconsistencies in performance assessments. For instance, strict filtering might limit

evaluation to ideal scenarios, ignoring real-world complexities. Similarly, flawed evaluation strategies might falsely favor models that excel in dominant classes, masking their overall shortcomings.

A standardized benchmark dataset is hence essential to ensure fair comparison, identify the most effective approach and establish a foundation for consistent advancements in the field.

2 MMT Database

The primary source of the publicly available light curves dataset is the Russian Mini Mega Tortora (MMT) database [6, 7] a wide-field monitoring system operated by the Special Astrophysical Observatory of the Russian Academy of Sciences. In August 2024, the database contained 14, 888 objects and 502, 815 tracks (or light curves), with new records being added daily. The light curves of objects, identified by their NORAD ID, are stored in a sequence of measurements containing the time, standard magnitude, and phase angle. Figure 1 presents a sample light curve of the Falcon 9 rocket body. The database is widely used for the training and validation for object characterization and attitude determination, described in Section 3.

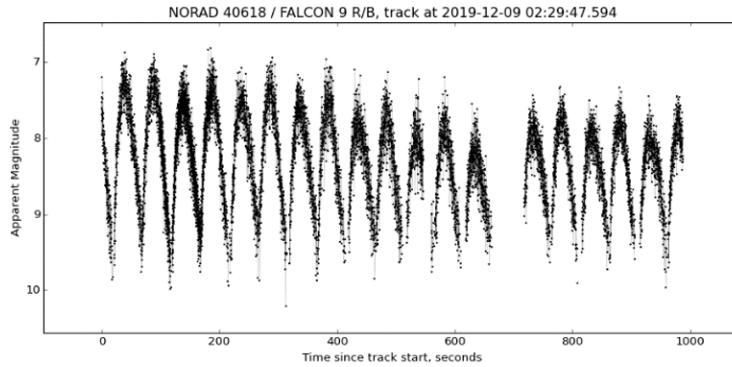


Figure 1: Example of Falcon 9 light curve from MMT database [7].

The data, however, exhibits several significant flaws: large gaps between observations, a limited number of observations per rotational period, and a low signal-to-noise ratio. These imperfections arise from physical processes, varying observation conditions, the high variability of the signal, and the inherent characteristics of the sensors. It is hence imperative for the raw data to be pre-processed before it can be used to train Machine learning models, as discussed in Section 4.

3 Machine learning research using MMT data

In recent years, there has been a trend to extract information about space objects from their light curves. A common kind of information to extract is the shape of an object. In this work, we focus on the shape classification problem with MMT be the main data source.

One of the first utilization of the database for machine learning was by the authors in [8]. They have set up a classification task using a custom 1-D convolutional network with 3 different target classes, namely: rocket bodies, debris, and satellites. It uses a sequence of 500 measurements from the light curve as input, with the reported test accuracy of around 75%.

A similar approach was used in [9] with the same input sequence strategy and 1-D CNN achieved a comparable accuracy of 75% with a different target class. To preserve more information in the input, in [4] the authors used phase angle data alongside magnitude measurements, resulting in two channel sequences of 1200 points in length, covering around 20 minutes of observation time. To achieve uniform size between light curves, truncation and zero padding was used. This helped the 1-D CNN network to distinguish between 8 classes (7 rocket body classes + bow-wing class) with a 90% accuracy. To enhance the performance, the model was pre-trained using simulated data. To ensure uniform temporal sampling, in the work [10] the inputs were resampled to a consistent frequency and a 1-D CNN and LSTM-based models were trained to classify object shapes specifically within the GEO orbit. As in the previous work, the data were padded and truncated to a uniform size.

Normalization was done by rescaling the sample between 0 and 1 by its maximum and minimum values.

In [5], a different approach was adopted to standardize input size while minimizing information loss. Light curves were folded by their rotational period, producing a new sample with 200 measurements each. This re-processing technique, combined with data augmentation approaches such as phase translation and noise addition, enabled a 1-D CNN to achieve 85% accuracy in a classification scenario involving 5 target classes of rocket bodies. Building on prior work, the authors in [11] implemented a method which involved segmenting light curves into individual rotational periods, and rescaling them into uniform size, resulting in a significantly expanded training dataset. Using this approach, a 1-D version of ResNet20 [12] network was employed to classify three rocket body classes, achieving 84% accuracy.

A contemporary approach to handling large sequences is using the Transformer[13] module. Astroconformer, a Transformer-based model introduced [14], was originally designed for surface gravity estimation from stellar light curves processing ,yet it can be easily adapted for a classification tasks.

While this section highlights only a subset of works in the subfield, it is evident that substantial differences exist in the preprocessing, evaluation, and normalization strategies employed across studies. These inconsistencies make direct comparison between methods highly challenging, if not impossible, emphasizing the need for a standardized approach to enable meaningful evaluation.

4 RoBo6 dataset

Prior work has predominantly focused on classifying objects such as rocket bodies, debris, or satellites based on their light curves. However, a more practically relevant challenge lies in distinguishing between objects with a similar shape, such as different types of rocket bodies. To the best of our knowledge, no standard benchmark currently exists for addressing this specific classification problem.

To address this gap, we curated a standardized dataset comprising six common rocket body populations : CZ-3B, Atlas 5 Centaur, Falcon 9, H-2A, Ariane 5 and Delta 4. The dataset is divided into a train set of 5,676 samples and a test set of 1,404 samples, with further details provided in Table 1.

Table 1: Number of samples per split and class.

	Ariane 5	Delta 4	CZ-3B	Atlas 5 Centaur	H-2A	Falcon 9
Train	660	233	2266	1029	623	865
Test	173	70	548	247	150	216

Each sample is characterized by the following fields: *label*, *ID*, *part* number, *period* (in nanoseconds), *mag*, *phase* and *time*. The last three fields refer to file paths that contain additional metadata, such as standard magnitude, phase angle, and time measurements. During preprocessing, samples can be divided into subsamples, with their sequence order recorded in the *part* field.

Since many machine learning models, such as CNNs, require grid structured data, each sample needs to be resampled onto a uniformly spaced grid with a manageable length. In order to convert the data into this format and to retrain as much information as possible, a series of filtering, splitting and resampling operations was performed over each sample. Standard normalization using mean and standard deviation was also applied.

The dataset is publicly available on the Hugging Face platform (<https://huggingface.co/datasets/kselica/RoBo6>) and was created from data downloaded in August 2024 from the official MMT website [7] using a custom Python script.

4.1 Gap-Based Splitting and Frequency Rescaling

The original dataset contains light curves that often span extended periods. However, in many cases, the gap between two consecutive measurements is larger than the rotational period of the object, presenting an opportunity to split such samples at these gaps. This process results in shorter light curves, as the intermediate gaps do not contain meaningful information.

Following this gap-based splitting, 95% of the samples were found to be shorter than 1,000 seconds. For the remaining longer samples, additional splits were made at this 1,000-second threshold. A sampling frequency of 10 Hz was chosen based on an analysis of time intervals between consecutive measurements. To standardize the dataset, all samples were rescaled and zero-padded to a fixed length of 10,000 points, representing 1,000 seconds of data.

4.2 Filtering of Low-Quality Samples

Low-quality samples are filtered based on two criteria empirically determined criteria consistent with values commonly used in prior studies, such as [8].

The first criterion sets the minimum number of measurements in one sample to 100. The second criterion evaluates the coverage of the apparent rotational period using a folded light curve, which is created by folding the data based on the object’s apparent rotational period [15] and reshaped to 100 points. To ensure sufficient data quality, the folded light curve must have at least 75% coverage by measurements.

4.3 Evaluation Strategy and Metric Selection

Selecting an effective evaluation method is essential for the precise assessment of model performance. A viable strategy involves partitioning data by objects into train and test sets to gauge the model’s competence in classifying new objects. Nevertheless, this method is challenging with multiple classes due to the limited object count and a significantly imbalanced light curve distribution, potentially resulting in an inadequately representative test set and insufficient samples for reliable evaluation. As shown in Table 2, the disparity in data distribution is particularly evident in the case of the CZ-3B rocket, where just 14 objects account for more than 60% of all light curve tracks.

Table 2: Number of tracks per CZ-3B rocket.

#Track Range	N.o. objects	N.o. tacks	Dataset Coverage
326 - 391	1	391	7.87 %
261 - 326	2	546	10.99 %
196 - 261	6	1303	26.22 %
131 - 196	5	782	15.74 %
66 - 131	10	907	18.25 %
1 - 66	63	1040	20.93 %

In real-world scenarios, the same object can be observed multiple times and must be consistently identified. To simulate this, the dataset can be split by track ID, ensuring all samples derived from the same light curve during preprocessing are assigned to the same set. Stratified splitting is employed to maintain consistent class distributions across the training and testing sets.

Given the imbalanced nature of the dataset, accuracy alone can be a misleading metric, since strong performance on classes with many samples may overshadow poor performance on underrepresented classes. To address this issue, the F1 macro score was chosen as the primary evaluation metric, as it takes into account the cardinality of all classes and thus provides a more balanced assessment of model performance across the dataset.

4.4 Model Training and Performance Evaluation

To demonstrate the utility of the dataset, five selected models [4], [12], [8], [5] and [16] were trained on it. Each model was trained for 50 epochs using the Adam optimizer with a learning rate of 0.001. Training parameters specified in the respective publications were followed; in cases where parameters were not provided, default settings were used. The complete list of parameters and further preprocessing details can be found in Table 4. The results, summarized in Table 3, align closely with those reported in the respective publications. Astroconformer, despite being developed for stellar light curves, proved to be the best-performing model in this benchmark comparison.

Table 3: Models performance.

Model	Accuracy	F1	Precision	Recall
ALLWORTH [4]	0.559 ± 0.044	0.478 ± 0.038	0.491 ± 0.033	0.531 ± 0.024
RESNET [12]	0.694 ± 0.023	0.600 ± 0.034	0.738 ± 0.026	0.584 ± 0.033
FURFARO [8]	$0.628 \pm \mathbf{0.009}$	$0.552 \pm \mathbf{0.013}$	0.570 ± 0.017	$0.552 \pm \mathbf{0.013}$
YAO [5]	0.672 ± 0.017	0.604 ± 0.023	0.622 ± 0.029	0.601 ± 0.020
ASTROCONFORMER [16]	0.725 ± 0.011	0.684 ± 0.015	$0.702 \pm \mathbf{0.010}$	0.677 ± 0.019

The hyperparameters used for the training the respective models are detailed in Table 4. In an attempt to reproduce the published models as closely to their published state as possible, a specific preprocessing step was incorporated specifically for these models:

- **ALLWORTH** [4] Observations from the initial 20 minutes were rescaled to a uniform grid consisting of 1,200 points.
- **FURFARO** [8] - The first 500 points were utilized as input.
- **YAO** [5] - The light curve was folded with the apparent rotation period to 200 point grid.

Table 4: Parameters used in the training process for each model.

	Input size	Channels	Batch size	Scheduler
Default	10 000	Mag	32	x
ALLWORTH[4]	1 200	Mag + Phase	256	x
RESNET[12]	10 000	Mag	32	x
FURFARO[8]	500	Mag	32	x
YAO[5]	200	Mag	32	x
ASTROCONFORMER[16]	10 000	Mag	32	✓

Only one paper offers a basis for comparison. We trained the Allworth method [4] on our dataset and compared the results with those reported in the original publication. However, the comparison is not entirely fair, as the training scenarios differ. Specifically, the authors in [4] used a balanced dataset with only 500 samples per class. Table 5 shows the result for the overlapping classes between the two datasets. The source code for the experiments is publicly available at https://github.com/kyselica12/RoBo6_Model_Comparison.

Table 5: Comparison of the Allworth method[4] on our dataset and the original paper.

<i>Rocket Bodies</i>	F1 score		Precision		Recall	
	Our	Original	Our	Original	Our	Original
Atlas 5 Centaur	0.39	0.74	0.53	0.71	0.33	0.78
CZ-3B	0.65	0.63	0.77	0.61	0.57	0.65
Delta 4	0.11	0.81	0.19	0.78	0.09	0.85
Falcon 9	0.49	0.80	0.48	0.81	0.49	0.79
H-2A	0.56	0.79	0.46	0.82	0.72	0.77

5 Conclusion

We introduce the RoBo6 dataset, a new benchmark for rocket body classification based on light curves. Derived from the MMT database, it includes six classes of rocket bodies and addresses inconsistencies in prior datasets through preprocessing and standardization. By enabling easy comparison across models and providing robust training and testing splits, RoBo6 facilitates consistent evaluations and accelerates advancements in this research area. Models trained on the dataset demonstrated its suitability for both traditional and transformer-based architectures, with Astroconformer achieving the highest performance. We hope this dataset fosters advancements in space object classification and promotes sustainable practices in space exploration.

Acknowledgements

This research was partially supported by grant APVV-21-0114.

References

- [1] Space debris by the numbers. https://www.esa.int/Space_Safety/Space_Debris/Space_debris_by_the_numbers. Accessed: 2024-09-09.
- [2] About space debris. https://www.esa.int/Space_Safety/Space_Debris/About_space_debris. Accessed: 2024-09-09.
- [3] Yaobing Xiang, Jiangbo Xi, Ming Cong, Yun Yang, Chaofeng Ren, and Ling Han. Space debris detection with fast grid-based learning. In *2020 IEEE 3rd International Conference of Safe Production and Informatization (IICSPI)*, pages 205–209, 2020.
- [4] James Allworth, Lloyd Windrim, James Bennett, and Mitch Bryson. A transfer learning approach to space debris classification using observational light curve data. *Acta Astronautica*, 181:301–315, 2021.
- [5] LU Yao and ZHAO Chang-yin. The basic shape classification of space debris with light curves. *Chinese Astronomy and Astrophysics*, 45(2):190–208, 2021.
- [6] S Karpov, E Katkova, G Beskin, A Biryukov, S Bondar, E Davydov, E Ivanov, A Perkov, and V Sasyuk. Massive photometry of low-altitude artificial satellites on mini-mega-tortora. *Revista Mexicana de Astronomía y Astrofísica*, 48:112–113, 2016.
- [7] Mmt website. <http://mmt9.ru/satellites>. Accessed: 2024-07-29.
- [8] Roberto Furfaro, Richard Linares, and Vishnu Reddy. Space objects classification via light-curve measurements: deep convolutional neural networks and model-based transfer learning. In *AMOS Technologies Conference, Maui Economic Development Board*, pages 1–17, 2018.
- [9] Richard Linares, Roberto Furfaro, and Vishnu Reddy. Space objects classification via light-curve measurements using deep convolutional neural networks. *The Journal of the Astronautical Sciences*, 67(3):1063–1091, 2020.
- [10] Emma Kerr, Gabriele Falco, Nina Maric, David Petit, Patrick Talon, E Geistere Petersen, Chris Dorn, Stuart Eves, Noelia Sánchez-Ortiz, R Dominguez Gonzalez, et al. Light curves for geo object characterisation. In *8th European Conference on Space Debris*, pages 9–20, 2021.
- [11] Daniel Kyselica, Linda Jurkasová, Roman Ďurikovič, and Jiří Šilha. Astronomical objects classification by convolutional neural network algorithms layers. In *2022 New Trends in Signal Processing (NTSP)*, pages 1–8. IEEE, 2022.
- [12] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.
- [13] A Vaswani. Attention is all you need. *Advances in Neural Information Processing Systems*, 2017.
- [14] Billy Shrive, Don Pollacco, Paul Chote, James A Blake, Benjamin F Cooke, James McCormac, Richard West, Robert Airey, Alex MacManus, and Phineas Allen. Classifying leo satellite platforms with boosted decision trees. *RAS Techniques and Instruments*, 3(1):247–256, 2024.
- [15] Jiří Šilha, Matej Zigo, Tomáš Hrobár, Peter Jevčák, and Martina Verešvárska. Light curves application to space debris characterization and classification. *complexity*, 10:3, 2021.
- [16] Jia-Shu Pan, Yuan-Sen Ting, and Jie Yu. Astroconformer: The prospects of analysing stellar light curves with transformer-based deep learning models. *Monthly Notices of the Royal Astronomical Society*, 528(4):5890–5903, 2024.