IOCU
2026
Conference

The 3rd International Online Conference on Universe

4–6 March 2026 | Online

MDPI

Feature-Rich Representations of GRB Light Curves for Microlensing Classification

Mohammad Hossein Zhoolideh Haghighi
K. N. Toosi University of Technology

INTRODUCTION & AIM

Gamma-ray bursts (GRBs) are among the most energetic transient phenomena in the Universe, originating from catastrophic events such as massive star collapse or compact-object mergers. Their complex and rapidly varying light curves encode rich information about the underlying physical processes but also present significant challenges for analysis. Owing to their cosmological distances, GRBs are promising probes of gravitational lensing.

In the microlensing regime, gravitational lensing can produce multiple unresolved images of a single GRB, manifesting as repeated or "echoed" peaks within the same light curve. These echoes exhibit similar temporal variability but differ in magnification and arrival time, making them difficult to identify using traditional methods. Efficient and automated techniques are therefore essential for detecting such rare events.

In this work, in the continuation of our previous work, we investigate the capability and importance of extracted features from GRBs' light curve using **tsfresh** package, in distinguishing lensed GRBs. We train machine learning (ML) models to identify microlensed GRBs using simulated long-duration GRB light curves. After extracting temporal features from light curves and applying supervised ML classifiers, we are capable of detecting lensed from non-lensed GRBs based solely on their temporal structure. Our results demonstrate that extracted features from tsfresh package are more informative than ones extracted via **Cesium**.
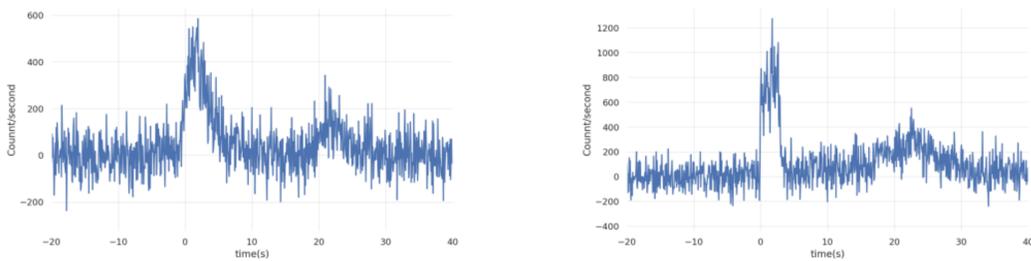
Figure 1. Left: simulated lensed light curve. Right: simulated non-lensed light curve.

METHOD

1. Data Generation
Using Monte Carlo simulations and a point-mass lens model, we generate a semi-synthetic dataset:
• Light curves for Lensed GRBs (with two echoed peaks)
• Light curves for Non-lensed GRBs

2. Feature Extraction
To detect microlensing signatures, we extract statistical and temporal features that characterize light-curve morphology, variability patterns, peak structure, and temporal distribution.

Feature extraction is performed using the tsfresh package, via the extract_features function.

This produces a feature matrix X, where each row represents one GRB, each column represents a computed feature.

3. Machine Learning Pipeline
Target variable:
• y = 1 → lensed
• y = 0 → Non-lensed

Dataset split:
• 80% Training
• 20% Testing

Feature scaling:
• Standardization using StandardScaler
• Fitted only on training data

Data Generation
(Using Monte Carlo simulations)

Feature Extraction
(using the tsfresh package)

Machine Learning Pipeline
(dataset split, Feature scaling, ..)

Classification Models
RF, SVC, NN, kNN, ...

4. Classification Models
We train and compare five supervised ML classifiers including: Random Forest (RF), Support Vector Classifier (SVC), k-Nearest Neighbors (kNN), Neural Network(NN), and Logistic Regression (LR).
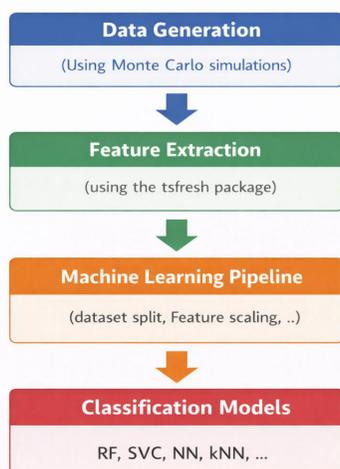
RESULTS & DISCUSSION

Model Performance Comparison

Best Model: Random Forest (RF)
•Accuracy: **94%**
•AUC: **0.99**
•F1-score → **0.94**
•Robust and well-balanced classification

Other Models
•NN → 84% accuracy, F1 → 0.83
•LR→ 73% accuracy , F1 → 0.73
•kNN → 67% accuracy, F1 → 0.67
•SVC → 60% accuracy, F1 → 0.55 (lowest performance)

Key Insight:
Ensemble methods (RF) outperform single estimators

Feature Importance

In our previous work some of the most Influential features were: mad, qso_log_chi2_qsonu, minimum, stetson_j, std, max_slope, amplitude which we selected manually, however, using tsfresh provide us with 1555 feature which are more informative for our classification task. We have shown top 20 Features in Figure 2.
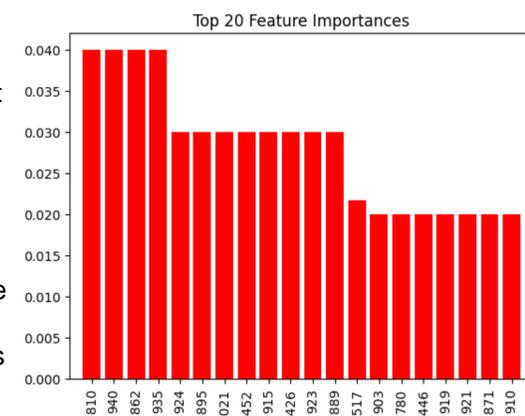
Figure 2. Top 20 most important features based on the trained RF classifier. Y-axis is importance of the feature and the X-axis is feature ID.
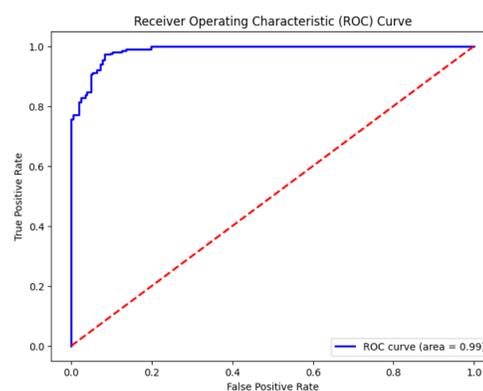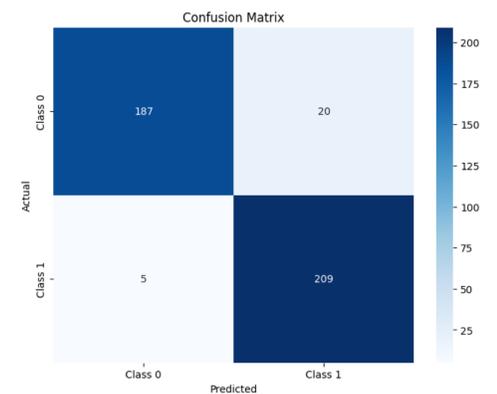
Figure 3. Confusion matrix and ROC curve for Random Forest classifier.

CONCLUSION

• Our ML approach successfully classifies lensed vs. non-lensed GRBs

• Extracting features using tsfresh package leads to more promising results

• The method is applicable to real-time GRB observations

• The best Performing Models is Random Forest

• The extracted features provide enough information to distinguish GRBs type.

• Model selection is critical for reliable microlensing detection

• The method can be applicable for real-time observations.

References

Zhoolideh et al. (2025), APJ, 992:189 (12pp), 2025 October 20
Christ, et al. (2018). tsfresh – A Python package. Neurocomputing, 307, 72-77.

zhoolideh@kntu.ac.ir