

Introduction

- SEP events are rare but can severely affect space-based and ground technologies.
- Accurate SEP prediction is challenging due to strong class imbalance and complex pre-flare signals.

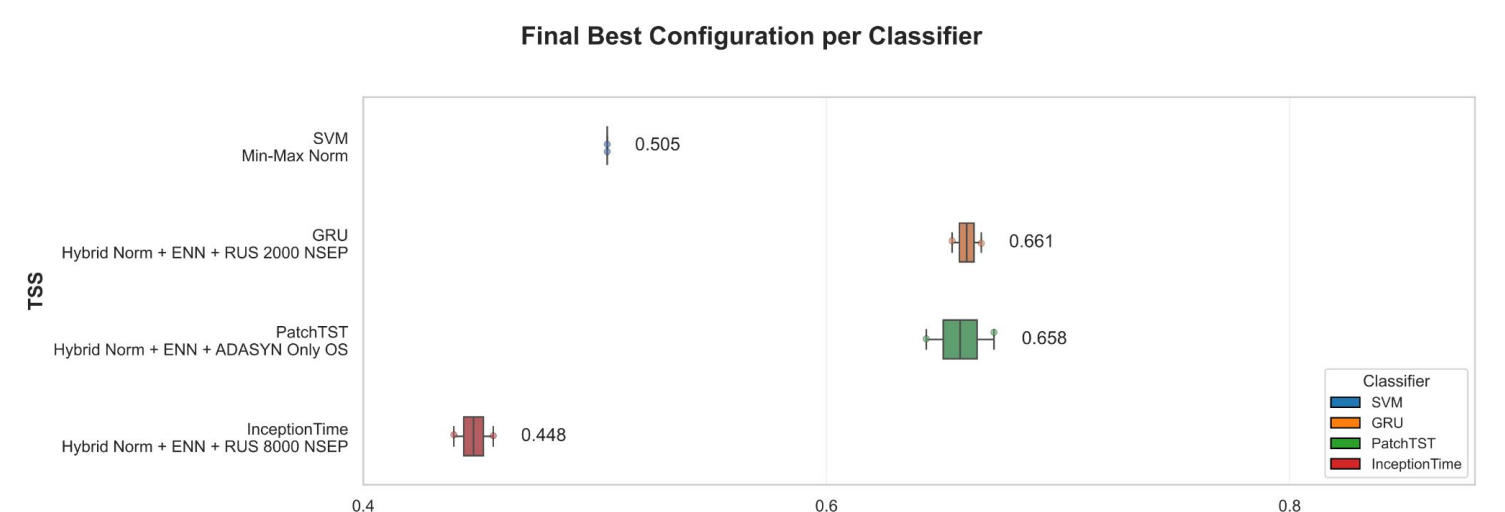
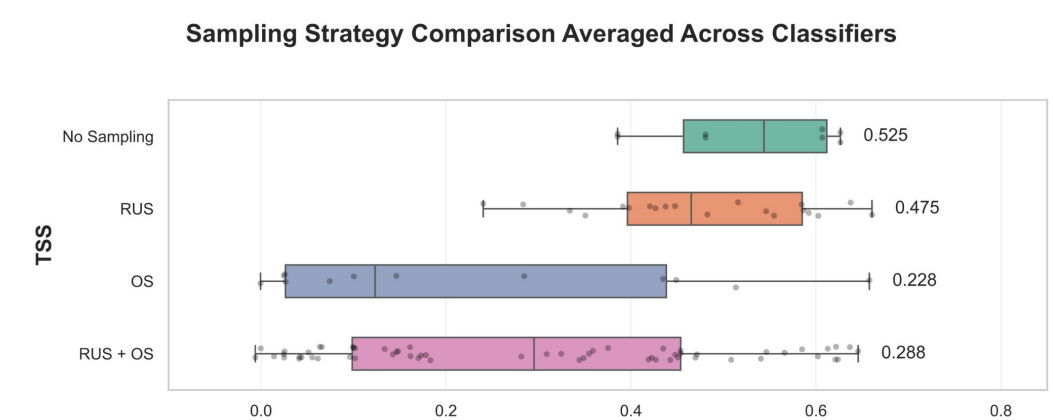
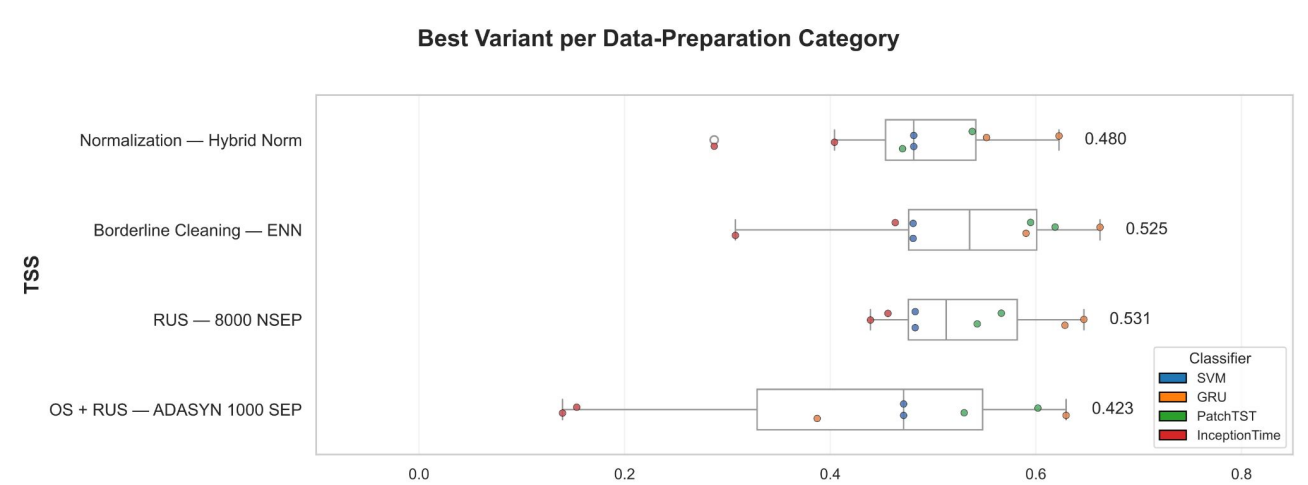
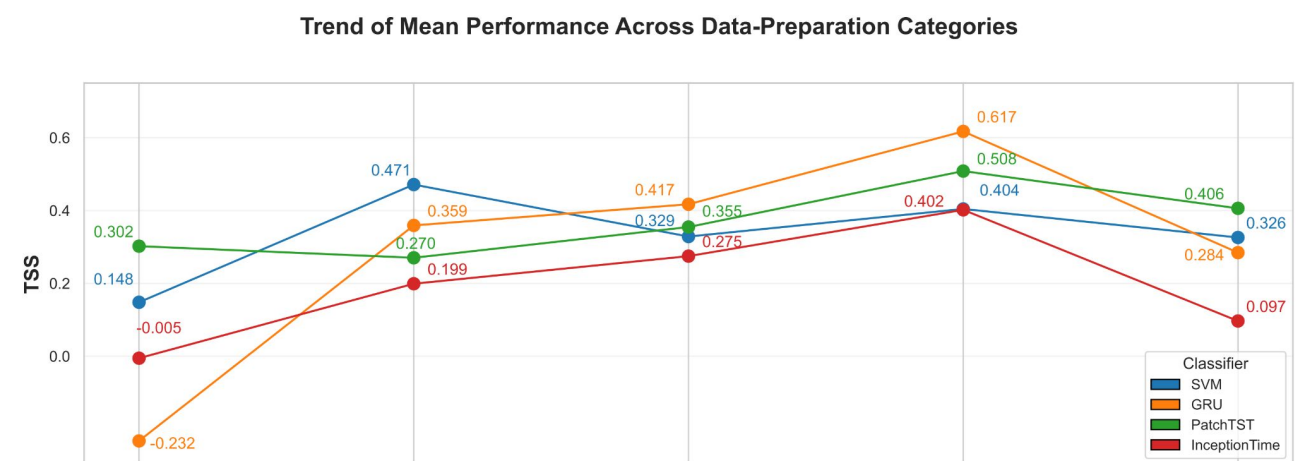
The Dataset

- NOAA/GOES pre-flare SEP dataset.
- 17,794 samples: 17,625 Non-SEP and 169 SEP events.
- Each sample: 24-hour multivariate time series.
- 288 time steps \times 10 physical features.
- Binary labels: Non-SEP $<$ 10 MeV and SEP \geq 10 MeV.
- Chronological 70/10/20 train-validation-test split per class to avoid temporal leakage.

Methodology

- This work evaluates data preparation strategies for improving SEP prediction from multivariate time series data.
- **Normalization:** Tested Z-score, Min-Max, Log, and Hybrid normalization after data splitting to evaluate their impact compared to raw data. Hybrid normalization applies Log + Robust scaling to skewed flux, density, temperature, and X-ray features, while using Z-score normalization for velocity-related features.
- **Borderline cleaning:** Applied Tomek Links, Edited Nearest Neighbors (ENN), and NearMiss Undersampling to remove ambiguous Non-SEP samples located near SEPs, reducing overlap between the majority and minority classes.
- **Random undersampling (RUS):** Reduced the dominant Non-SEP class to different training sizes (8000, 4000, 2000, 1000, and 500 samples) while keeping all available SEP samples.
- **Oversampling:** Increased the rare SEP class using SMOTE, ADASYN, and TimeGAN.
- **Oversampling + undersampling:** Combined SEP oversampling with Non-SEP undersampling to create more balanced training sets and evaluate whether hybrid balancing improves rare-event detection.
- **Model evaluation:** Trained SVM, GRU, PatchTST, and InceptionTime models and compared performance using TSS.

Experiments & Conclusions



- Normalization significantly improves performance compared with raw data. Hybrid normalization is the strongest normalization method.
- ENN is the best borderline-cleaning method, while NearMiss significantly reduced prediction performance.
- RUS can improve performance when the majority class is reduced carefully, especially with 8000 Non-SEP samples; however, aggressive undersampling reduces performance by removing useful Non-SEP information.
- Oversampling alone or combined with RUS reduces performance, suggesting that synthetic SEP generation is unreliable, likely due to the extremely small number of real SEP samples available.
- GRU and PatchTST achieve the strongest final results, while InceptionTime performs lower.
- Hybrid Norm + ENN + RUS by keeping 2,000 Non-SEP samples, with GRU as the classifier, achieved the highest performance (TSS = 0.661).