Improving Solar Energetic Particle Event Prediction through Multivariate Time Series Data Augmentation

Pouya Hosseinzadeh , Soukaina Filali Boubrahimi , and Shah Muhammad Hamdi
Department of Computer Science, Utah State University, Logan, UT 84322, USA; pouya.hosseinzadeh@usu.edu, soukaina.boubrahimi@usu.edu, s.hamdi@usu.edu

Abstract

Solar energetic particles (SEPs) are associated with extreme solar events that can cause major damage to space- and ground-based life and infrastructure. High-intensity SEP events, particularly ~100 MeV SEP events, can pose severe health risks for astronauts owing to radiation exposure and affect Earth's orbiting satellites (e.g., Landsat and the International Space Station). A major challenge in the SEP event prediction task is the lack of adequate SEP data because of the rarity of these events. In this work, we aim to improve the prediction of ~30, ~60, and ~100 MeV SEP events by synthetically increasing the number of SEP samples. We explore the use of a univariate and multivariate time series of proton flux data as input to machine-learning-based prediction methods, such as time series forest (TSF). Our study covers solar cycles 22, 23, and 24. Our findings show that using data augmentation methods, such as the synthetic minority oversampling technique, remarkably increases the accuracy and F1-score of the classifiers used in this research, especially for TSF, where the average accuracy increased by 20%, reaching around 90% accuracy in the ~100 MeV SEP prediction task. We also achieved higher prediction accuracy when using the multivariate time series data of the proton flux. Finally, we build a pipeline framework for our best-performing model, TSF, and provide a comprehensive hierarchical classification of the ~100, ~60, and ~30 MeV and non-SEP prediction scenarios.

1. Introduction

Solar energetic particles (SEPs) are the outcomes of some extensive solar events such as solar flares and coronal mass ejections (CMEs). Large-scale eruptions occur when the Sun goes through different phases of strong activities. The solar flux and UV radiation during a solar flare cause extreme brightness near the Sun's surface that can be observable with special instruments. Consequently, part of the magnetic energy is converted into thermal energy that spreads through space. As a consequence, fast-moving energetic particles such as protons, electrons, and heavy ions are released from the Sun, spreading through space. These intense events are dangerous for astronauts, even within spacecraft, causing cancer and skin reactions. In particular, SEP events are hazardous when astronauts are outside of the station performing space walks. Furthermore, strong events are hazardous to electronic devices as well, resulting in major disruptions to communications on Earth. SEP events pose a serious challenge in achieving the goals of interplanetary missions; therefore, predicting and mitigating the risks associated with SEP events require advanced monitoring and a predictive system. Accurately predicting SEP events is a timely issue with NASA's recent aim to explore the Moon, which is planned to take place this decade. It is also important to have an accurate predictive system in place prior to human exploration on Mars, planned to take place in the next decade.

Some of the risks have been extensively analyzed in the context of a Mars crewed mission. As explained in McKenna-Lawlor et al. (2012), a scenario of a 30 day stay on Mars's surface and a 400 day cruise phase round trip to Mars shows that the latter could be dangerous for astronauts. The presence of an SEP event, particularly one characterized by a high-energy spectrum, could pose severe health risks to the crew during this time frame. Therefore, predicting SEP events in advance provides the decision-makers and astronauts with adequate time to take action and reduce the risks. Space agencies need to accurately predict these rare events to reduce the risks and ensure the safety of their missions (Fogtman et al. 2023).

Many SEP predictive models have been introduced since they were first observed and reported by Forbush (1946). SEP predictive models can be divided into two main categories: physics-based models and data-driven models. Physics-based models are rooted in our understanding of solar and space physics principles, utilizing scientific knowledge to make predictions. On the other hand, data-driven models leverage statistical and machine-learning (ML) techniques to discern patterns and relationships within observed data. SEP predictive models primarily distinguish between two types of events: gradual and impulsive. Gradual SEP events, such as those originating from phenomena like CMEs, are characterized by their slower and more progressive onset. This characteristic makes them relatively less challenging to predict (Luhmann et al. 2010; Tenishev et al. 2021). However, it is important to note that rapid onset SEPs can also originate from CME-driven shocks, particularly those occurring low in the solar corona, and forecasting such events presents unique challenges. SOLar Particle ENgineering COde (SOLPENCO) is a code that predicts the flux and fluence of the gradual SEP events (Aran et al. 2006). In addition, Sato et al. (2018) introduced a real-time warning system using a physics-based approach to predict SEP events. Physics-based models rely on a set of a priori

assumptions grounded in solar and space physics principles, which form the basis of their predictive capabilities. These assumptions, while valuable, introduce certain limitations in forecasting, such as simplifications in complex physical processes. It is important to acknowledge that physics-based models may encounter challenges in achieving real-time forecasting. Factors such as computational time and the use of non-real-time input observations can impact their ability to provide forecasts with the immediacy required for operational decision-making. This underscores one of the advantages of our study, as we leverage real-time observations from the Geostationary Operational Earth Satellites (GOES) spacecraft, which are vital for timely forecasting applications (Lario 2005).

Data-driven models (e.g., ML models) rely on historical observations to find precursors of SEP events and forecast future occurrences. There are a number of studies that use ML models to classify SEP events. Laurenza et al. provided short-term warnings for predictions of SEP events using a $\geq$ M2 soft X-ray (SXR) burst precursor, which resulted in a reduced false-alarm rate (Laurenza et al. 2009). To reduce false alarms, Stumpo et al. (2021) employed a methodology inspired by their Empirical model for Solar Proton Events Real-Time Alert (ESPERTA). In their approach, they implemented a cutoff based on flare heliolongitude and optimized the logistic regression model with input parameters such as flare heliolongitude, SXR fluence, and radio fluence at around 1 MHz. Additionally, they applied the synthetic minority over-sampling technique (SMOTE) algorithm to enhance their model's performance. These steps were taken by Stumpo et al. (2021) to mitigate the occurrence of false alarms in their forecasting process. Bain et al. explored a range of ML classifiers such as logistic regression, decision tree, and support vector machine to predict SEP events. They found that the false-alarm ratio depends on the ratio between the solar proton event (SPE) and non-SPE flares (Bain et al. 2018). Bourbahimi et al. indicated that multivariate time series of X-ray and proton flux channels are precursors to the predictions of >100 MeV SEP and non-SEP events and achieved a satisfactory result using an interpretable decision tree classifier (Boubrahimi et al. 2017). Furthermore, Aminalragia-Giamini et al. trained a deep-learning model using solar flare SXR measurements to predict the occurrences of SEP events (Aminalragia-Giamini et al. 2021). Finally, the University of Malaga Solar Particle Event Predicto model also uses ML as part of the prediction (Núñez & Paul-Pena 2020). The model has been running in RT at the Community Coordinated Modeling Center for many years. A comprehensive work that put more than 40 SEP predictive models together and reviewed the state-of-the-art approaches, including physics-based and data-driven models, has recently been released (Whitman et al. 2023).

SEP events are characterized by increases in proton flux above background levels, typically resulting from the acceleration of particles during solar flares and CME shocks. Observing the correlation between proton flux and X-ray emissions as time series data can be helpful in predicting the occurrence of SEP events. In this paper, we aim to employ time series data-driven models to classify SEP and non-SEP events. There are several SEP energy levels, such as >10, >20, >30, >60, and >100 MeV, that are of interest to researchers. For example, SEP events with a >100 MeV band can penetrate Earth's atmosphere and disrupt communications by affecting

electronic devices (Mewaldt et al. 2007). On the other hand, lower energy levels have a lower relative penetration power. In this work, we focus on predicting three classes of SEP events, $\sim$30 MeV (low), $\sim$60 MeV (medium), and $\sim$100 MeV (high), by training ML models.

In this work, non-SEP events are defined as X-ray events (solar flares) whose peak intensity is at least M3.0 but did not lead to any SEP event. It is known that ML models are data-hungry methods whose performances highly depend on the number of available samples to train (Marcus 2018; Hosseinzadeh et al. 2023). Unlike non-SEP events, SEP events are extremely rare, occurring less than 100 times for some energy levels in almost 30 yr. As a consequence, there are not enough data samples to train ML models. An effective way to deal with this issue is artificially increasing the training data size (Boubrahimi et al. 2016; Wen et al. 2020; Li et al. 2021; Alshammari et al. 2022; Bahri et al. 2022a, 2023a). Time series data augmentation has been successfully applied to improve the performance of multiple predictive tasks, such as forecasting, anomaly detection, and classification. Additionally, data augmentation increases the generalization ability of models by reducing overfitting (Iwana & Uchida 2021). There are several data augmentation techniques ranging from basic to advanced methods, such as statistical generative models, decomposition methods, and deep-learning methods (Wen et al. 2020). Moreover, the synthetic multivariate time series generation for flare forecasting has been a successful method to overcome such rare events, as explained in Chen et al. (2021).

The goal of this study is to leverage data augmentation methods to increase the size of the training data in order to overcome the small rare-event SEP event data set challenge. To achieve this goal, we collect SEP events and their corresponding X-ray and proton data from multiple catalogs and data sources. Then, we augment the size of the SEP catalogs using data augmentation techniques, namely, Gaussian noise, SMOTE, and adaptive synthetic (ADASYN). We then classify the events into SEP and non-SEP. We use three ML models: random convolutional kernal transform (ROCKET), shapelet transform (ST), and time series forest (TSF). We evaluate the results using k-fold cross-validation and predicting test data for different scenarios. This provides us with the opportunity to evaluate on more train-test splits. We analyze the results on multiple evaluation metrics such as accuracy, F1-score, TSS, and HSS2 used by multiple previous studies (Bobra & Couvidat 2015; Inceoglu et al. 2018). Later, we perform sequence analysis for all models to examine the effect of considering different observation window sizes on classification accuracy. By doing so, we can make sure which observation window size is more suitable for a specific energy band SEP prediction when using a specific classifier (e.g., ROCKET, SHAPELET, TSF). Finally, we developed a comprehensive hierarchical classification framework that combines the learning from all the binary models ($\sim$30, $\sim$60, and $\sim$100 MeV models).

The rest of this paper is organized as follows: In Section 2, we explain the process of data collection and visualize the data distribution. In Section 3, we outline the methodology used in our study, including data partitioning, data augmentation methods, classifiers, and evaluation metrics. In Section 4, we present our experimental results and discuss our findings. Finally, in Section 5, we conclude with a summary of our

findings, highlight the contributions of our study, and discuss potential future research directions on SEP event prediction.

2. Data Sources

We used three main data sources to conduct our study: (1) the Integrated Geostationary Solar Energetic Particle (GSEP) event catalog, (2) Heliophysics Events Knowledgebase (HEK) to determine non-SEP events, and (3) GOES proton channels to collect historical proton flux data.

Our SEP event list originates from the GSEP event catalog (Rotti et al. 2022), which is a comprehensive list of SEP events covering three solar cycles. The GSEP list contains >10, >30, >60, and >100 MeV energy bands. We extracted SEP events associated with ~30, ~60, and ~100 MeV energy bands ranging from 1986 to 2011. Then, we extracted non-SEP events using the HEK website,[1] which provides all events, including SEP and non-SEP. The decision not to include a model for ~10 MeV events was primarily based on the specific focus of our research objectives and the available data set. Our goal was to predict SEP events with higher energy levels (~30, ~60, and ~100 MeV) owing to their significance in space weather forecasting and their potential impact on space missions.

Our second data source is proton flux data measured by the Space Environment Monitor instrument on board GOES. According to the National Oceanic and Atmospheric Administration's data documentation,[2] proton flux channel P4 is related to SEP events ranging from 15 to 44 MeV energy bands. Furthermore, P5 and P6 are related to SEP events ranging from 39 to 82 MeV and from 84 to 200 MeV, respectively. Therefore, we considered proton flux data recorded by the Energetic Proton, Electron and Alpha Detectors channel P4 (P4_flux), P5 (P5_flux), and P6 (P6_flux) for ~30, ~60, and ~100 MeV, respectively. The average 5 minute data can be collected from https://www.ncei.noaa.gov/data/goes-space-environment-monitor/access/avg/.

Figure 1 shows images of the solar disk during solar flare occurrences with their corresponding proton flux data P6, P5, and P4. The images are captured on board the Solar Dynamics Observatory (SDO) of Atmospheric Imaging Assembly (AIA) 304 and are available from https://helioviewer.org/. The figure shows a 5 hr observation window prior to the parent solar flares' start time with a 5 minute average integral proton flux data cadence. The first three images correspond to three solar flares that triggered SEP events with ~100, ~60, and ~30 MeV energy bands. For instance, a solar flare recorded on 2012 May 19 at 04:17 showed an increase in proton flux above background to as high as the ~100 MeV integral energy band.

After the data collection, we used data augmentation techniques to increase the number of SEP events to build a balanced data set with non-SEP events. Then, we used the new enriched data set to train ML-based predictive models. Figure 2 shows the data collection, augmentation, and classification pipeline.

Our Python source codes are made publicly available on GitHub,[3] with a copy of the most recent version deposited to Zenodo at doi:10.5281/zenodo.10463674.

We visualize the collected SEP and non-SEP time series proton flux data for ~100, ~60, and ~30 MeV energy bands using the t-distributed stochastic neighbor embedding (t-SNE) dimensionality reduction technique. t-SNE is a widely used technique for reducing dimensionality and visualizing large data sets by giving each data point a location in a two- or three-dimensional map (Van der Maaten & Hinton 2008). The first row of Figure 3 shows the t-SNE projection of our real data into a two-dimensional plane for all three energy bands (30, 60, and 100 MeV). Each data point in the panels corresponds to a single time series sample. These samples include real SEP events, synthetic SEP events, and non-SEP events. It is important to note that data points are not easily separable when no augmentation technique is used, as shown in the first row. In particular, ML models may confront a significant challenge in classifying ~100 MeV SEP events since the data points cannot be readily distinguished given the lack of clear boundaries between the data points. Likewise, it is difficult to visually separate ~30 and ~60 MeV SEP event data points from non-SEP events. Poor distinguishability makes models perform poorly in classification, resulting in a high false-alarm ratio and low accuracy levels. In Section 3.2, we discuss the data augmentation effects on the data distribution and projection.

It is essential to mention that in Figure 3 the observed differences in data distribution within the same energy band cases are a result of how t-SNE processes the data when considering different class combinations. When plotting t-SNE for real SEP and non-SEP samples, the projection reflects the distinction between these two classes. However, when we introduce synthetic SEP samples into the mix, we effectively have three distinct classes (real SEP, synthetic SEP, and non-SEP). This results in a variation in the t-SNE projection because the algorithm is now trying to capture the relationships and patterns between three different classes rather than two. It is also important to note that while the t-SNE projections may vary, the essential takeaways and insights from the data remain consistent across these different plots. The goal of these visualizations is to aid in understanding the separability of the classes and the clustering of data points, which can be valuable for our ML models.

3. Methods

This section introduces the methods that we used for data partitioning, data augmentation, classification, and evaluation. We divide this section into four subsections: (1) data partitioning, (2) data augmentation methods, (3) classification models, and (4) experimental evaluation.

3.1. Data Partitioning

In this subsection, we explain the approach used to split the data into partitions before and after using the data augmentation techniques. For all the partitioning scenarios, the non-SEP event data are real since they exist in abundance. We considered six different partitioning scenarios that we describe using the [Real/Synthetic] notation, as shown in Figure 4. Imbalanced [1/0] is used as a baseline to compare the results when the ML models are trained on real imbalanced data (i.e., no data augmentation technique is applied). In this scenario, the amount of non-SEP data is almost five times greater than the SEP data. Additionally, we considered five more scenarios to analyze the predictive models' results when the data are balanced (i.e., the number of SEP events is equal to the number of non-SEP events). First, we created a balanced [1/0] case

---
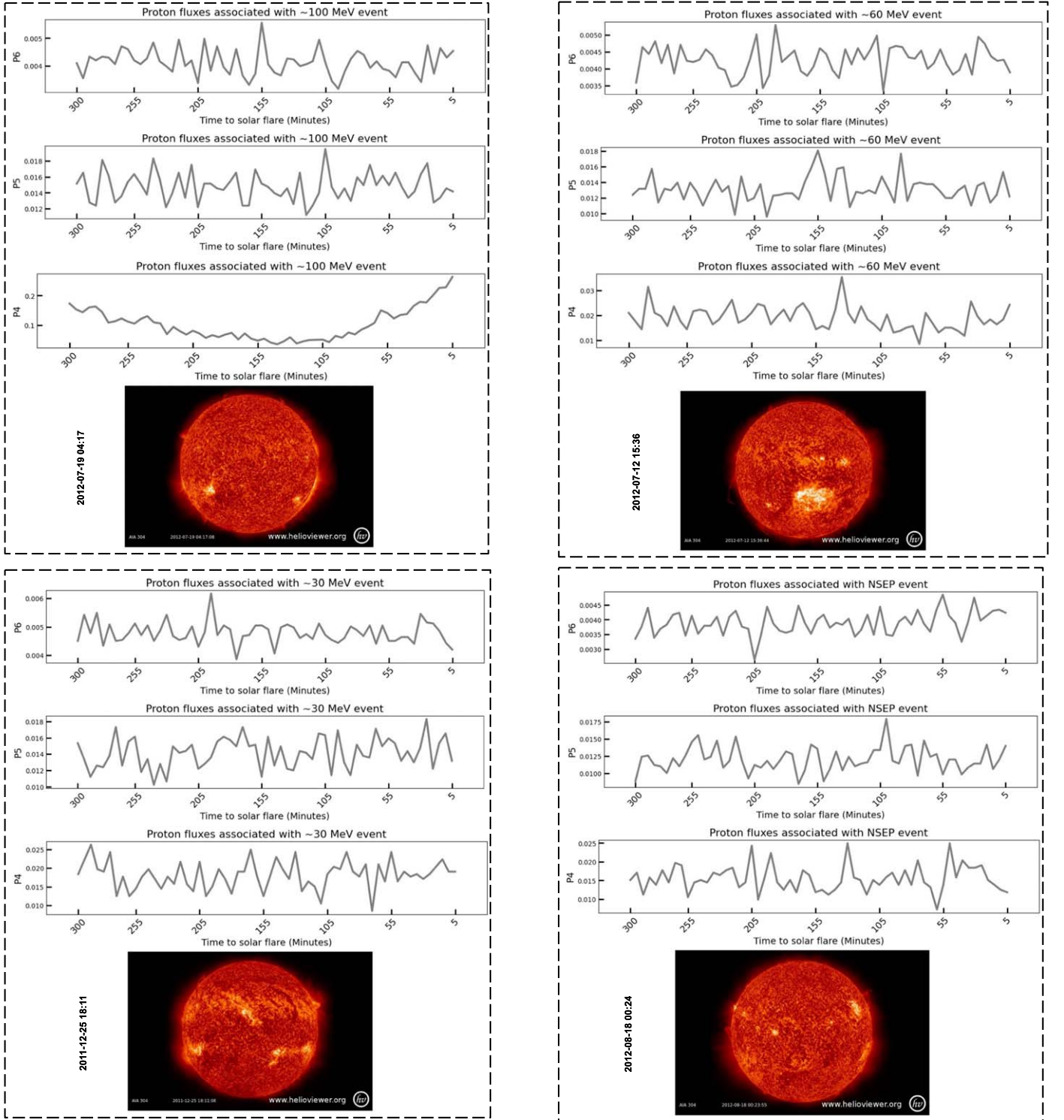
Figure 1. Solar flare images captured by SDO of AIA 304, accessible from https://helioviewer.org/. The 5 hr time series proton flux data P6, P5, and P4 channels are shown for ~100, ~60, and ~30 MeV and a non-SEP event.

where the number of SEP events and the number of non-SEP events are equal. Then, we leveraged data augmentation techniques to artificially create SEP events creating four more scenarios. The third [1/1] scenario consists of doubling the number of SEP event samples by creating 1N synthetic events. Note that all non-SEP events are real, and the data augmentation techniques are only applied to SEP events. Likewise, the fourth [1/2] scenario represents the case when there are N real

SEP event samples and $2N$ real non-SEP event samples. Similarly, we applied the same idea to achieve the fifth and sixth [1/3] and [1/4] scenarios. By analyzing the six partitioning scenarios, we can gain insights into how the ML models are interpreting and processing the augmented data and identify potential areas for improvement.

In order to facilitate a more thorough analysis, we increased the amount of data at a linear rate. In our study, we examine 93
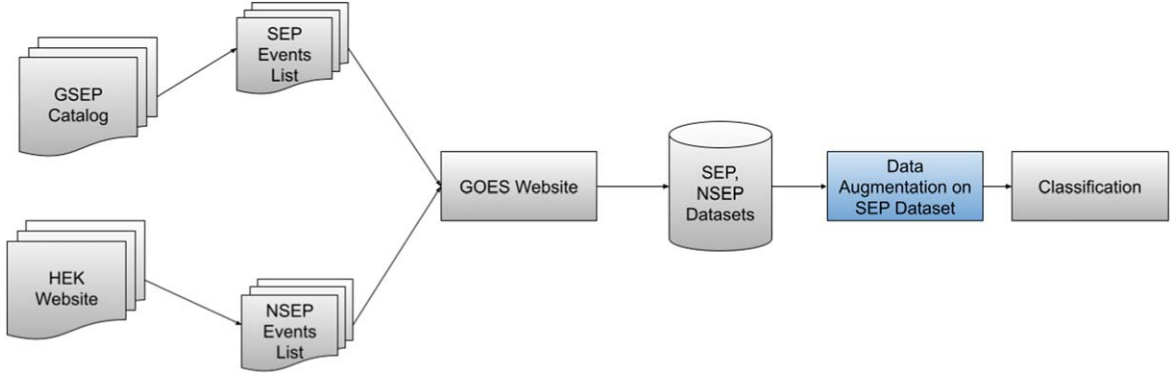
Figure 2. Data collection, augmentation, and classification pipeline.
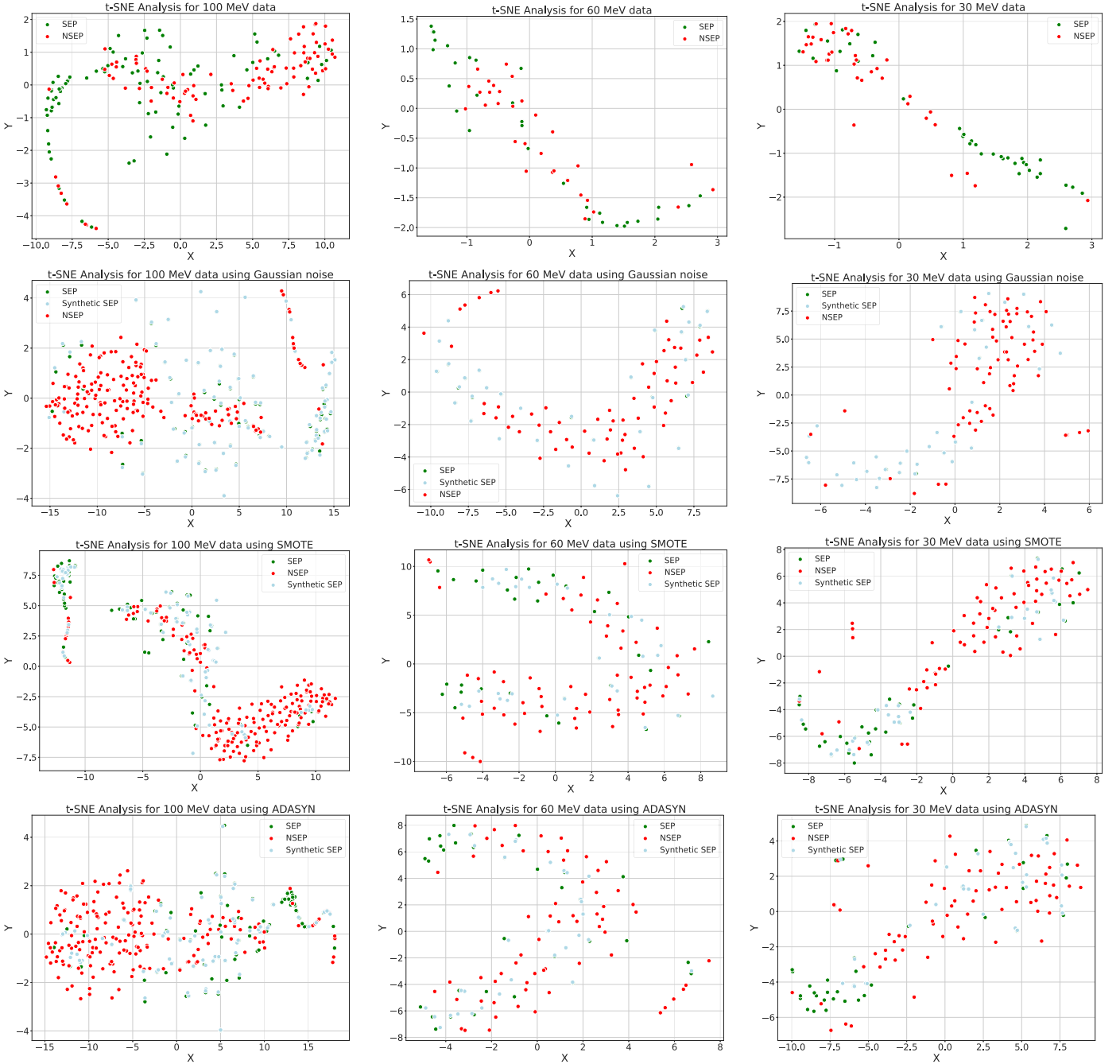
Figure 3. t-SNE-based two-dimensional projection of real and synthetic proton flux time series data for ∼30, ∼60, and ∼100 MeV SEP and non-SEP events.
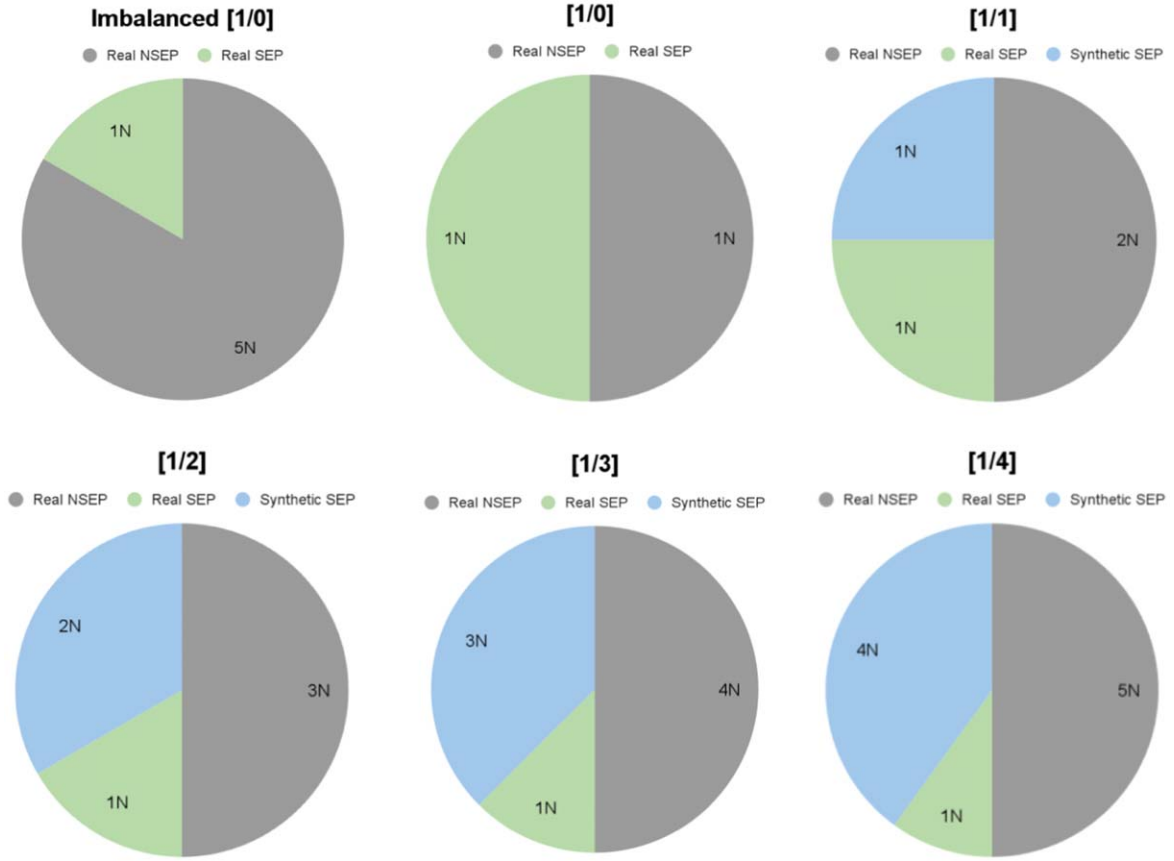
Figure 4. Comparison of the data augmentation models with the specific classifier used for evaluation.

real SEP samples for the ∼100 MeV energy band and generate additional synthetic samples to increase the total number of SEP samples to 465, which includes both real and synthetic samples. Similarly, we examine 33 and 38 samples for ∼60 and ∼30 MeV bands, respectively. For the sake of fairness, the same number of non-SEP samples is taken into account for all scenarios listed above. Furthermore, non-SEP samples show a consistent pattern in terms of both quantity and value across all three energy bands in our data set. This consistency arises from the fact that a non-SEP event is defined as an event that does not involve any form of SEP. It it evident that the number of real SEP samples for ∼60 and ∼30 MeV energy bands is less than that for the ∼100 MeV energy band owing to inadequate data available from the GSEP list. Events >30 and >60 MeV in the GSEP catalog are indeed more abundant than events >100 MeV. However, as our analysis specifically addressed ∼30, ∼60, and ∼100 MeV events, the counts for ∼30 and ∼60 MeV were understandably not higher than those for ∼100 MeV since the overlapped events are discarded.

We show a subset of the SEP event list in Table 1, which shows the SEP start time, flare start time, flare peak time, flare GOES class, and energy level in MeV. The flare start time and flare peak time represent the start time of the flare and the time of maximum X-ray flux during flare occurrence, respectively. In this work, we consider the flare start time as the last observed time within the observation window, and it is the data from this observation window that are used as an input to the model. To ensure that the size of the observation window can vary without impacting the prediction time, we analyze different observation window sizes as discussed in Section 4.

In our study, the SEP events used in the data set are considered as isolated events. We define an SEP event as an event that occurs independently within the specified observation window, which is set before solar flares. We do not consider prior SEP events within the observation window as precursors to the event. Our approach focuses on analyzing the proton flux data in the observation window time frame to predict SEP events based on the characteristics of the data during that specific period. We acknowledge that in reality SEP events can sometimes occur in succession from particularly active regions, and the occurrence of one SEP event can be a key predictor of subsequent events. However, for the purpose of our study and to isolate the impact of the proton flux data within the chosen observation window, we treat each SEP event as an independent occurrence.

3.2. Data Augmentation Methods

In order to create synthetic SEP partitions, we use three time series data augmentation techniques: (1) Gaussian noise (Talavera et al. 2022), (2) SMOTE (Yang et al. 2021), and (3) ADASYN (He et al. 2008).

Gaussian noise consists of creating synthetic data by adding random noise to the time series data. By doing so, we can increase the robustness of the models and avoid overfitting since there will be larger training data available. Random noise is produced each time and is added to the real SEP data in order to create new SEP data. In other words, random noise is generated independently for each time step within the time series of real SEP data. Subsequently, this random noise is added to the corresponding time steps in the real SEP data,

Table 1
SEP Event List Based on GSEP Catalog

ID	SEP Start Time	Flare Start Time	Flare Peak Time	Flare GOES Class	Energy Level (MeV)
1	1986-02-06 08:35:00	1986-02-06 06:18:00	1986-02-06 06:25:00	X1.7	100
2	1986-02-07 13:00:00	1986-02-07 10:11:00	1986-02-07 10:29:00	M5.2	100
3	1986-02-10 21:00:00	1986-02-10 20:25:00	1986-02-10 20:48:00	C9.5	100
4	1986-02-14 10:35:00	1986-02-14 09:10:00	1986-02-14 09:21:00	M6.4	100
5	1986-05-04 10:45:00	1986-05-04 09:39:00	1986-05-04 10:07:00	M1.2	100
6	1986-03-06 18:05:00	1986-03-06 16:37:00	1986-03-06 17:02:00	C4.6	60
7	1988-01-02 23:00:00	1988-01-02 21:11:00	1988-01-02 21:35:00	X1.4	60
8	1989-03-17 18:20:00	1989-03-17 17:29:00	1989-03-17 17:37:00	X6.5	60
9	1989-11-26 11:00:00	1989-11-25 22:55:00	1989-11-25 23:24:00	X1.0	60
10	1990-03-19 05:55:00	1990-03-19 04:39:00	1990-03-19 05:08:00	X1.5	60
11	1986-02-04 09:25:00	1986-02-04 07:35:00	1986-02-04 07:40:00	X3.0	30
12	1986-02-05 02:00:00	1986-02-05 12:37:00	1986-02-05 12:53:00	M3.0	30
13	1987-11-07 22:00:00	1987-11-07 20:28:00	1987-11-07 20:30:00	M1.2	30
14	1989-03-08 04:00:00	1989-03-06 13:54:00	1989-03-06 14:15:00	X5.0	30
15	1989-04-10 21:00:00	1989-04-09 00:44:00	1989-04-09 00:59:00	X3.5	30

Note. This table contains a subset of events used in our work.

resulting in the creation of new synthetic SEP data points. This process is repeated for each instance of data augmentation to diversify the synthetic data set. The Gaussian probability density function is defined in Equation (1), where μ and σ are the mean and standard deviation of the distribution, respectively:

$$N(\chi: \mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2}(x - \mu)^2/\sigma^2\right). \quad (1)$$

The second method is SMOTE, which consists of generating synthetic data between each minority class sample and its k-nearest neighbor. SMOTE oversamples by creating synthetic samples instead of oversampling with replacement, a technique where instances from the minority class are duplicated randomly to balance the data set (Chawla et al. 2002). Although SMOTE has been successfully applied to time series data (Hostetter & Angryk 2020), there are some limitations such as working only on continuous data and a linearly dependent model that can cause a biased result. We acknowledge that SMOTE as a nonparametric data augmentation method may have limitations, particularly when dealing with autocorrelated time series. We carefully explored parametric procedures such as autoregressive integrated moving average (ARIMA; Singh & Ray 2021) and generalized autoregressive conditional heteroskedasticity (GARCH; Horv & Kokoszka 2003) as well. GARCH considers volatility clustering and time-varying volatility and has demonstrated success in modeling financial time series. These models can capture different aspects of time series patterns, including short-memory and long-memory components and volatility dynamics. Furthermore, resampling strategies for imbalanced time series forecasting is another data augmentation technique (Moniz et al. 2017). However, in our work, we observed that SMOTE outperforms these methods and also generates data that align well with the characteristics of the original solar proton flux time series as shown in Figures 17 and 18 in the Appendix.

The following sentences explain SMOTE in detail. It first computes the k-nearest neighbors for each minority class sample in X using a distance metric. The k-nearest neighbors are computed based on a distance metric applied to the features of each data point. This distinction ensures that the synthetic samples generated by the SMOTE method are relevant to the

original data distribution and feature space. For each minority class sample xi in X, k neighbors are randomly selected from its k-nearest neighbors. $N(xi)$ represents the set of selected neighbors. For each selected neighbor xj in $N(xi)$, it computes the difference vector $dij = xj - xi$. It generates m synthetic minority class samples by interpolating between xi and its selected neighbors. For each synthetic sample s, the equation is

$$s = xi + \text{rand}(0, 1) * dij, \quad (2)$$

where rand(0, 1) is a random number uniformly distributed between 0 and 1. Finally, it combines m synthetic minority class samples into the original feature matrix X.

The third technique is ADASYN, which has been frequently used in data augmentation tasks. The idea behind ADASYN is similar to SMOTE. While SMOTE solely focuses on the samples in the minority class, ADASYN focuses on the samples in the minority class that are difficult to classify. ADASYN generates synthetic samples based on the density distribution of the minority class and reduces the impact of outliers. The following sentences explain ADASYN in detail. According to He et al. (2008), it similarly computes the k-nearest neighbors for each minority class sample in X using a distance metric. However, ADASYN calculates the relative imbalance ratio, ri, for each minority class sample xi in X. The ratio is defined as $ri = Ni/N$, where Ni is the number of minority class samples in the k-nearest neighbors of xi. Then, it computes the weight for each minority class sample, where wi is defined as $wi = ri/\Sigma ri$. The model generates m synthetic minority class samples by selecting the k-nearest neighbors for each minority class sample with probabilities proportional to their weights. For each synthetic sample s, the formula is given by

$$s = xi + \text{rand}(0, 1) * (xj - xi), \quad (3)$$

where xj is a randomly selected neighbor of xi. Finally, it combines m synthetic minority class samples into the original feature matrix X. Both SMOTE and ADASYN are effective in addressing the class imbalance problem, but ADASYN's adaptive nature can provide better results in scenarios where the imbalance is severe or when the minority class samples are particularly difficult to classify.
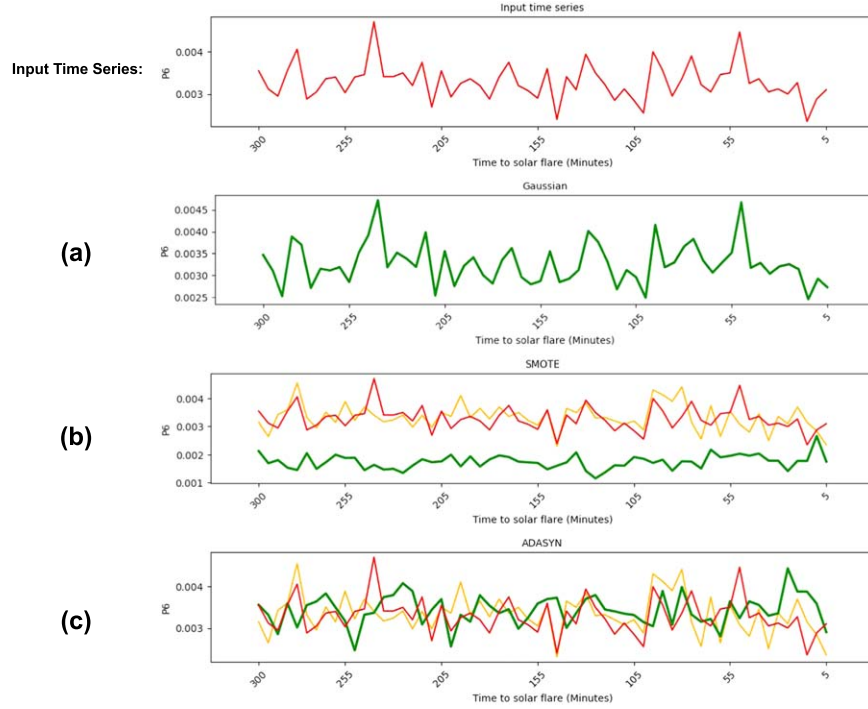
Figure 5. Input time series before and after using data augmentation methods with (a) Gaussian noise, (b) SMOTE, and (c) ADASYN. Original and augmented time series are shown in red and green, respectively. The time series shown in yellow shows the *k*-nearest neighbor.

An illustration of augmented time series data generated by the three methods is shown in Figure 5. The input time series represents real proton flux data a few hours prior to an SEP event. The figure shows the nearest neighbor time series for SMOTE and ADASYN that were used in the augmentation process in yellow. Figure 3 shows the two-dimensional t-SNE projection of the enriched data set using the three aforementioned data augmentation methods. The real SEP samples are shown in green, and the synthetic ones are shown in blue. Non-SEP data are entirely real and are shown in red. The second row of Figure 3 shows the data distribution after using Gaussian noise to increase the number of SEP samples. The blue circles represent synthetic SEP data where it can be observed that they nearly follow the real SEP data distribution. Likewise, SMOTE and ADASYN show a similar pattern by generating synthetic SEP data points close to the real data points. It is important to note that the Gaussian noise method generates data points that are relatively closer to the real SEP data points. Our hypothesis is that by applying data augmentation techniques to enrich our data set, we can help classifiers achieve better prediction accuracy.

3.3. Classification Models

This subsection explains the details of three ML-based classification methods used to predict SEP events. Our data consist of time series of the proton flux of multiple energy band channels. Working with time series data is challenging owing to the temporal dependencies between the order and the timing of the observations. For this reason, we used time series classification models for their ability to deal with high-dimensional data taking into account the chronological order of the data. In this paper, we used three time series classification models: ROCKET, ST, and TSF. The following models have been provided to be successfully used in multiple

prior studies that involve multivariate time series data (Bagnall et al. 2017; Ruiz et al. 2021).

ROCKET transforms time series data into features using a very large number of random convolutional kernels, typically on the order of tens of thousands of kernels (Dempster et al. 2020). The idea behind convolutional kernels in ROCKET is similar to traditional convolutional neural networks that summarize temporal segments into a lower-dimensional segment using a kernel function (Girshick 2015). By applying random convolutional kernels, ROCKET is able to capture a wide range of temporal patterns at different scales. The randomization and ensembling process enhances the robustness and generalization capabilities of the feature extraction in the ROCKET classifier. The latter has gained significant popularity mainly owing to its remarkable speed and accuracy, making it a desired technique in various classification applications (Dhariyal et al. 2023).

The second classifier is the ST, which has been frequently used in different fields of study (Arul & Kareem 2021; Bahri et al. 2022b, 2022c; Li et al. 2022a, 2022b). Research on shapelets has gained considerable attention, primarily because of their highly interpretable nature (Bahri et al. 2022b). ST is a technique used in time series classification that extracts discriminative and important features. It involves a process of selecting and comparing subsequences within the time series to identify those that are most representative of a specific class (Ye & Keogh 2009). Shapelet-based learners provide a visual representation of the pattern that triggers the decision made by the classifier (Boubrahimi et al. 2020).

TSF is the third classifier that we trained for the SEP/non-SEP classification task. TSF is an ensemble approach that utilizes multiple decision trees grouped together, where each tree is based on interval-based random sampling from the time series data. TSF randomly selects different intervals from the time series for training *r* decision trees. The individual decision
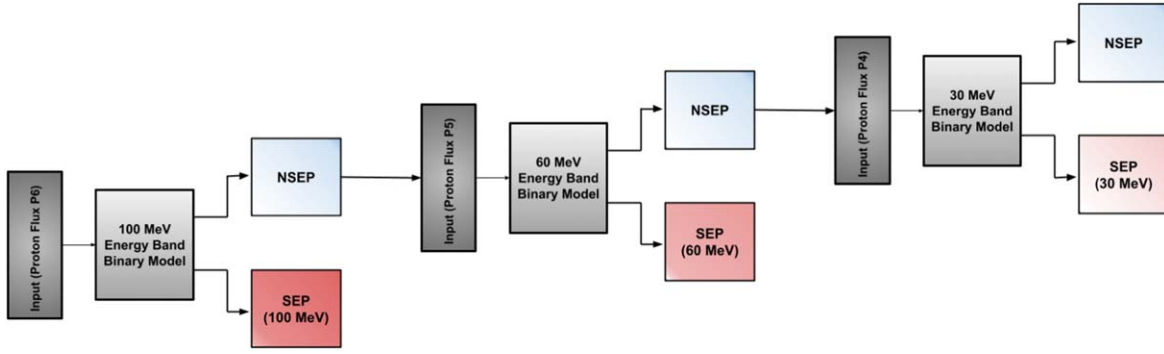
Figure 6. Hierarchical framework diagram.

trees are trained on features generated from a total of the square root of m intervals with a minimum length p (m being the length of the time series; Rigatti 2017). The features constitute the summary statistics of the time series intervals (i.e., mean, standard deviation, and slope). TSF outperforms other state-of-the-art baseline models, such as one-nearest-neighbor classifiers with dynamic time warping (Deng et al. 2013).

Finally, we have developed a hierarchical framework that processes and predicts SEP events using a hierarchical approach. Figure 6 shows a diagram of our proposed hierarchical framework for classifying the three energy bands. The model starts by predicting whether a 100 MeV event will occur. In case there is no predicted 100 MeV event, the model moves to predict the occurrence of a 60 MeV event. In case there is a predicted 60 MeV event, the model ends the prediction process. Otherwise, the model predicts the occurrence of 30 MeV events.

3.4. Experimental Evaluation

In this subsection, we explain the metrics used in this paper to evaluate both balanced and imbalanced classification. In the first phase, we evaluate the classifiers' performance before and after applying data augmentation techniques when trained on balanced data sets using accuracy and F1-score metrics. Accuracy is a measure that reveals how many correct predictions the model has made as defined in Equation (4). On the other hand, F1-score is a combination of precision and recall that identifies positive cases and minimizes false-positive and false-negative results. F1-score, precision, and recall are defined in Equations (7), (5), and (6), respectively:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \quad (4)$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (5)$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (6)$$

$$\text{F1-score} = 2 * \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}}. \quad (7)$$

Due to the substantial imbalance in SEP event numbers across different energy bands and the non-SEP class, the accuracy and F1-score metrics may exhibit elevated values. This is a result of a classifier consistently predicting the same class, specifically the majority non-SEP class. Accordingly, we used two more metrics better suited for the imbalance classification problem: true skill statistic (TSS) and updated

Heidke skill score (HSS2). These metrics have been particularly used in the task of the flare forecasting problem to assess the models in Chen et al. (2021). The TSS and HSS2 metrics are defined by

$$\text{TSS} = \frac{\text{TP}}{\text{TP} + \text{FN}} - \frac{\text{FP}}{\text{FP} + \text{TN}} \quad (8)$$

$$\text{HSS2} = \frac{2 * ((\text{TP} * \text{TN}) - (\text{FN} * \text{FP}))}{(\text{TP} + \text{FN}) * (\text{FN} + \text{TN}) + (\text{FP} + \text{TN}) * (\text{TP} + \text{FP})}, \quad (9)$$

where TP, FP, TN, and FN represent true positive, false positive, true negative, and false negative, respectively.

4. Experimental Results

In this section, we show different analyses of the SEP and non-SEP classification and discuss the results in light of the metrics defined in the previous section. As explained in Section 3.2, we used three data augmentation techniques to generate synthetic samples from the existing input data, increasing the size of the data set. By doing so, these techniques help prevent the overfitting problem that can occur when an ML model is trained on a small-sized data set. We first show the results of the classification on the balanced data in terms of accuracy and F1-score. We used a k-fold cross-validation process to evaluate the models' variance across different folds. The k-fold cross-validation technique divides the data set into k equal folds and uses $k - 1$ folds to train the models and one fold to test the models. For example, a 10-fold cross-validation will use nine folds for training the model and the remaining one for testing. This procedure is performed 10 times, with each fold utilized once as the test set. We used 10-fold cross-validation to evaluate the 100 MeV model. Likewise, we used k-fold cross-validation for assessing 60 and 30 MeV energy bands with K equal to 7 since the number of data samples is comparatively smaller. We divide this section as follows: analysis on the balanced data set, analysis on the imbalanced data set, comparison between data augmentation methods, comparison between univariate and multivariate time series models, four-class classification, hierarchical classification framework, and sequence analysis.

4.1. Analysis on Imbalanced and Balanced Data Sets

First, we compared the accuracy and F1-score of the imbalanced and balanced time series data set when used to train the three classifiers. The process is carried out for predictions in all three energy bands for SEP and non-SEP
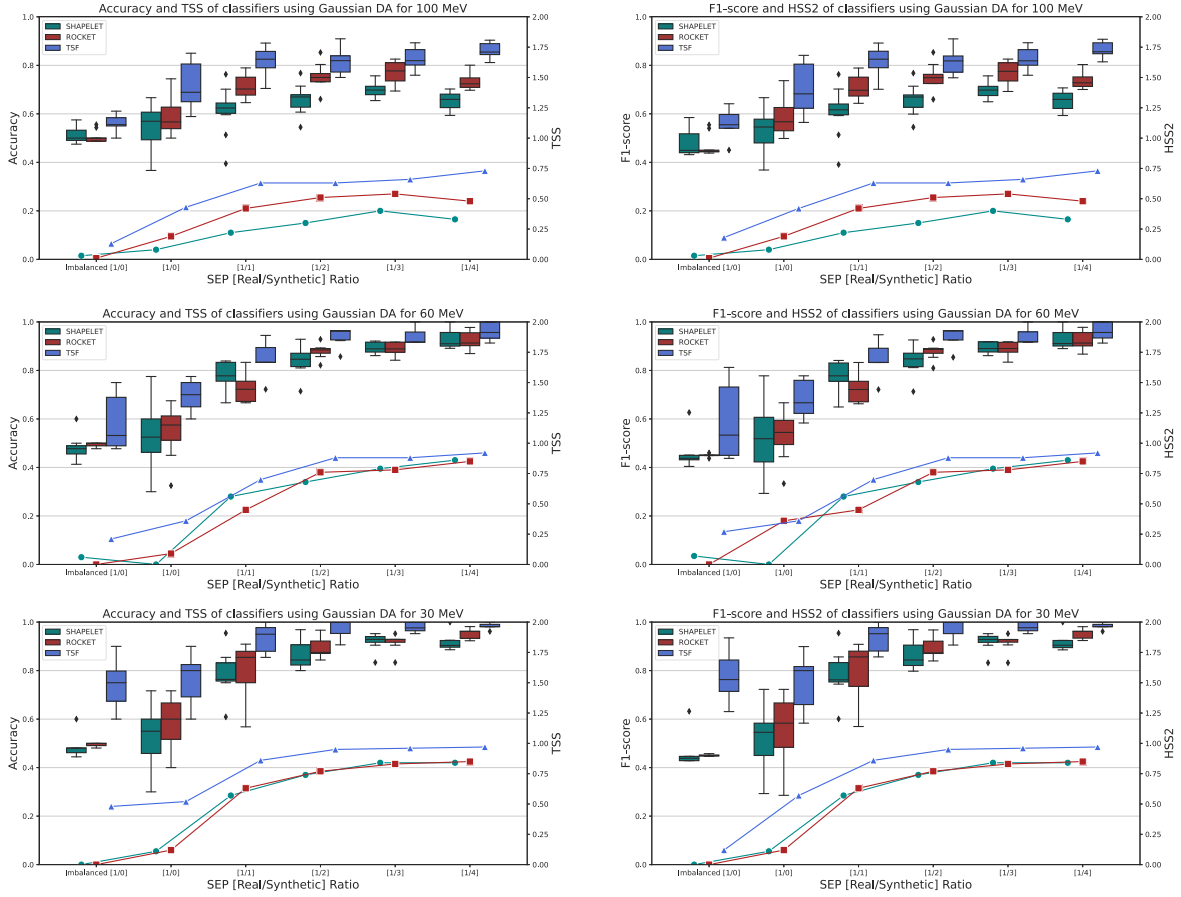
Figure 7. Evaluation of SEP and non-SEP prediction of ROCKET, SHAPELET, and TSF. The box plots show the distribution of *k*-fold cross-validation in terms of accuracy and F1-score before and after applying Gaussian data augmentation techniques. The line plots show TSS and HSS2 before and after applying Gaussian data augmentation techniques. [1/0] represents the results on the balanced data set without using DA, and the rest represent the results on data augmentation partition according to their corresponding SEP [Real/Synthetic] ratio.

events. As shown in Figure 7, the imbalanced time series data (defined as Imbalanced [1/0]) yielded poor classification results for the case of all the energy bands. Likewise, the balanced [1/0] showed weak performance. However, using the Gaussian noise data augmentation technique resulted in improvements in accuracy and F1-score. As shown in Figure 7, adding a single synthetic SEP partition enhanced the accuracy and F1-score of all classifiers. We note that the accuracy and F1-score continually improve as the number of synthetic folds is increased. The box plots show that as we add more synthetic SEP folds, the robustness of the results improves as measured by the accuracy and F1-score box plots' variance.

Figure 8 reveals that despite having a larger data set, distinguishing between SEP and non-SEP events for the 100 MeV energy band is a more challenging classification task. Conversely, the classification of SEP/non-SEP events in the 60 and 30 MeV energy bands appears to be relatively easier, particularly when incorporating synthetic data. Figures 8 and 9 indicate that the performance of the SMOTE and ADASYN techniques is consistent and comparable to that of the Gaussian method, regardless of the experimental conditions. Although the Gaussian data augmentation technique was found to be more effective for distinguishing between SEP/non-SEP events in the 60 and 30 MeV energy bands, the SMOTE and ADASYN techniques outperformed Gaussian when it came to the 100 MeV energy band.

Figure 7 also shows average TSS and HSS2 metrics for imbalanced and balanced data sets as line plots. This shows that the TSS and HSS2 scores are close to 0.8 for the case of the 100 MeV energy band and exceed 0.9 for the case of the 60 and 30 MeV energy bands when using the Gaussian noise data augmentation technique. In the absence of data augmentation techniques, TSS and HSS2 scores are notably low, particularly for SHAPELET and ROCKET methods. Chen et al. achieved a TSS score improvement from an average of 0.04 to 0.81 when using data augmentation on the solar flare prediction model (Chen et al. 2021). The latter results are consistent with our prediction improvement shown in Figures 7, 8, and 9. Figure 8 shows that the performance of 100 MeV classification is significantly improved on a single partition [1/1] generated by SMOTE. When predicting the 60 and 30 MeV energy bands, employing Gaussian noise data augmentation yielded better results. Figures 7, 8, and 9 show that assigning more synthetic data partitions increases both TSS and HSS2 in all the scenarios. As evidenced by these results, the classification of SEP/non-SEP events can be highly challenging. Without leveraging data augmentation techniques, the average classification accuracy across all three energy bands for the best-performing model cannot exceed 70%.

In all experimental scenarios, TSF achieved superior performance compared to both ROCKET and SHAPELET. Additionally, it is worth noting that ROCKET generally outperforms SHAPELET. These results indicate that the data
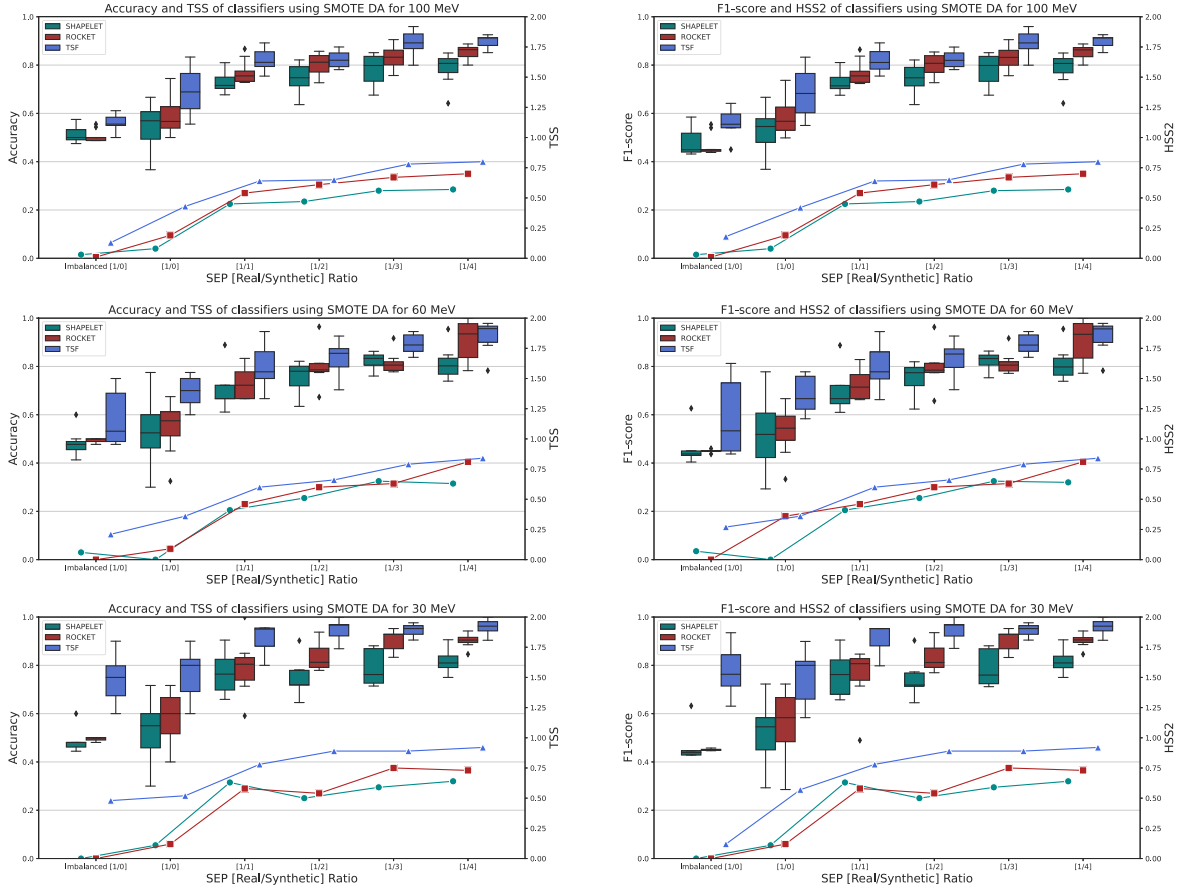
Figure 8. Evaluation of SEP and non-SEP prediction of ROCKET, SHAPELET, and TSF. The box plots show the distribution of *k*-fold cross-validation in terms of accuracy and F1-score before and after applying SMOTE data augmentation techniques. The line plots show TSS and HSS2 before and after applying SMOTE data augmentation techniques. [1/0] represents the results on the balanced data set without using DA, and the rest represent the results on data augmentation partition according to their corresponding SEP [Real/Synthetic] ratio.

augmentation techniques can significantly enhance the predictive capabilities of the ML-based models. Although the SHAPELET classifier is not the most accurate, it is an explainable method. The prominent features extracted by the SHAPELET classifier for 100, 60, and 30 MeV binary SEP/non-SEP classification using a 5 hr observation window size are shown in Figure 10 in red. In each energy band scenario, we display an important shapelet that leads to SEP events. We also show that the most important shapelet triggers non-SEP events. These shapelets play a pivotal role in our SEP event prediction models. The significance of Figure 10 lies in its demonstration of the SHAPELET classifier's ability to identify subtle yet discriminative patterns associated with SEP events. These shapelets serve as key discriminative features that contribute to the classifier's decision-making process. While they may appear intricate, they are important in improving our understanding of SEP event prediction.

Among the classifiers presented in this study, SHAPELET was ranked third, whereas TSF consistently achieved the best performance across all the tested scenarios. The figures demonstrate that all of the data augmentation techniques employed were effective in improving accuracy and F1-score. Nonetheless, it is noteworthy that SHAPELET was limited in its ability to enhance results in the last partition. As mentioned earlier, ST is a feature extraction method that extracts informative subsequences from a set of time series data. One potential reason for SHAPELET's poor performance compared

to the other classifiers discussed in this study may be attributed to the length of the observation window. Increasing the size of the observation window could potentially enhance the feature extraction capabilities of the SHAPELET classifier. The feature extraction step of ROCKET, which involves random convolutional kernels, showed promising results but did not achieve the same level of performance as TSF.

4.2. Comparison between Data Augmentation Methods

The next experiment examines the models' performance with respect to the data augmentation methods in balanced data sets. Figure 11 shows a comparison of the three data augmentation methods. The average and median accuracy of the balanced classification without using data augmentation methods are represented with horizontal gray and black dashed lines, respectively. The comparison between different data augmentation methods is shown in the first column of Figure 11. The results suggest that the Gaussian technique resulted in better predictions for 60 and 30 MeV energy bands. As can be observed in Figure 11, data augmentation has increased the accuracy of SEP/non-SEP classification in all scenarios. The last row of Figure 11 illustrates the accuracy of TSF for the three energy bands separately. The use of data augmentation techniques resulted in a 20% increase in accuracy for the classification of 100 MeV, increasing it from an average of 70% to an average of 90%. Likewise, the classification of
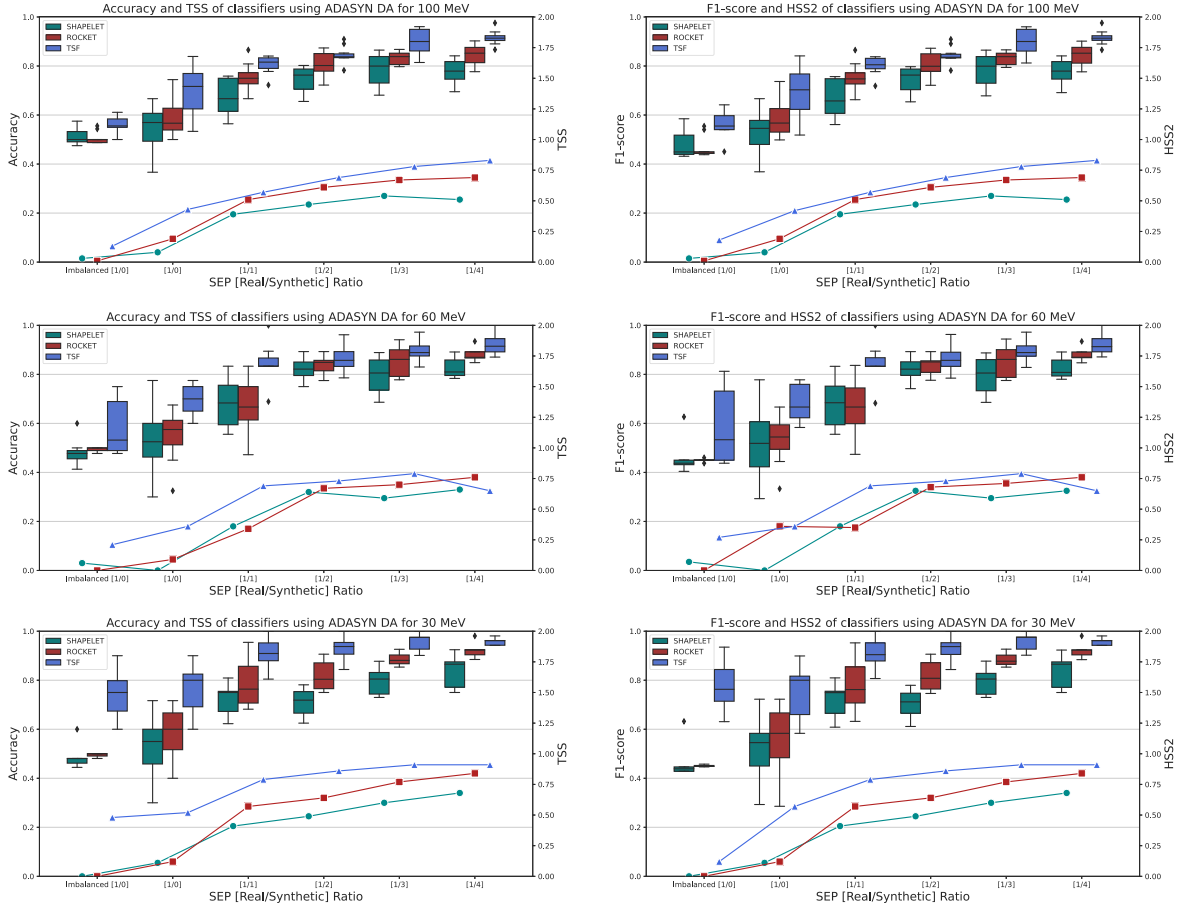
Figure 9. Evaluation of SEP and non-SEP prediction of ROCKET, SHAPELET, and TSF. The box plots show the distribution of *k*-fold cross-validation in terms of accuracy and F1-score before and after applying ADASYN data augmentation techniques. The line plots show TSS and HSS2 before and after applying ADASYN data augmentation techniques. [1/0] represents the results on the balanced data set without using DA, and the rest represent the results on data augmentation partition according to their corresponding SEP [Real/Synthetic] ratio.

SEP events with 30 MeV energy bands reached an average accuracy of 98% from an average accuracy of 75%.

4.3. Comparison between Univariate and Multivariate Time Series Models

Up to this point, we utilized univariate time series data in our binary classification. The main idea for using univariate time series data is that a specific proton flux channel is associated with a particular energy band. For instance, we used proton fluxes of channel P6 for classifying SEP events with 100 MeV energy bands. Similarly, we used proton fluxes of channels P5 and P4 to classify 60 and 30 MeV, respectively. This assumption, however, may have potentially hindered further prediction improvement since other proton flux channels may have an impact on predicting specific energy bands. Therefore, in this part, we create multivariate time series data using all three proton flux data (i.e., P6, P5, and P4) to analyze their impact on the binary classification. We employed three multivariate time series strategies to compare with the univariate time series strategy, which employs a 5 hr proton flux data (60 values) prior to the SEP (or non-SEP) parent solar flare. The following list explains the three strategies:

1. *Strategy 1*. In this strategy, we concatenate proton flux data from channels P6, P5, and P4. While this approach offers a comprehensive view of the proton flux across

multiple energy bands, it comes at the cost of potentially losing the timing information inherent in sequential time series data. Combining these channels creates a data set with 180 (3×60) values, obscuring the sequential order of observations. This might impact the model's ability to capture temporal dependencies in the data, as the sequence of events is essential in understanding solar phenomena.

2. *Strategy 2*. This strategy aggregates proton flux values from the three channels. While this simplifies the input data into a single series of 60 values, it does result in a loss of spectral information. The aggregation sums the values, and as a consequence, the model may not distinguish between the individual energy bands. This strategy might be less effective in capturing the spectral differences that are essential for identifying distinct SEP events.

3. *Strategy 3*. Here we generate summary statistics from P6, P5, and P4 channels, resulting in a concise representation of the data. The statistics, including average, median, and standard deviation, provide a condensed view of the proton flux characteristics. While this strategy reduces the dimensionality of the data and maintains some spectral information, it might lead to a loss of fine-grained details. However, it allows the model to focus on essential statistical features that may be indicative of SEP events.
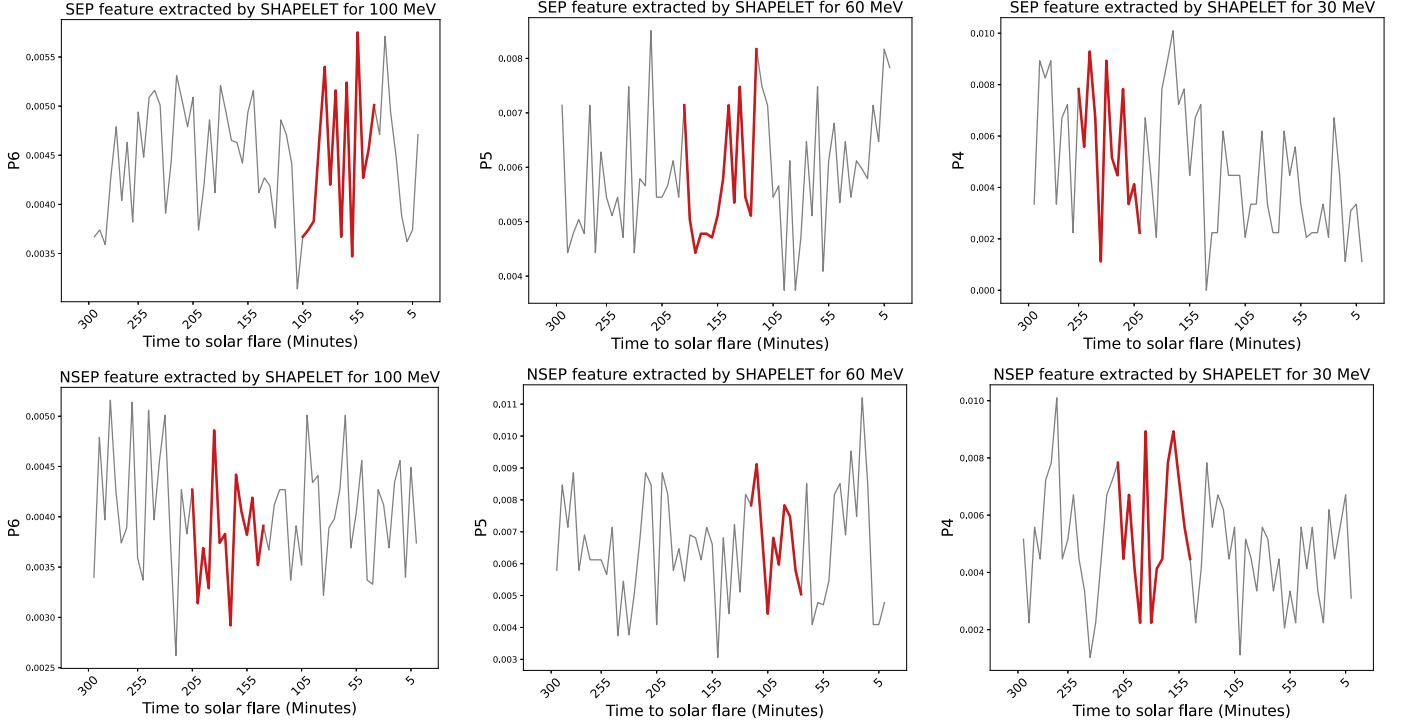
Figure 10. Important features extracted by the SHAPELET classifier for 100, 60, and 30 MeV binary classification using a 5 hr observation window size. The first row shows the SEP class, and the bottom row represents the non-SEP (NSEP) class.
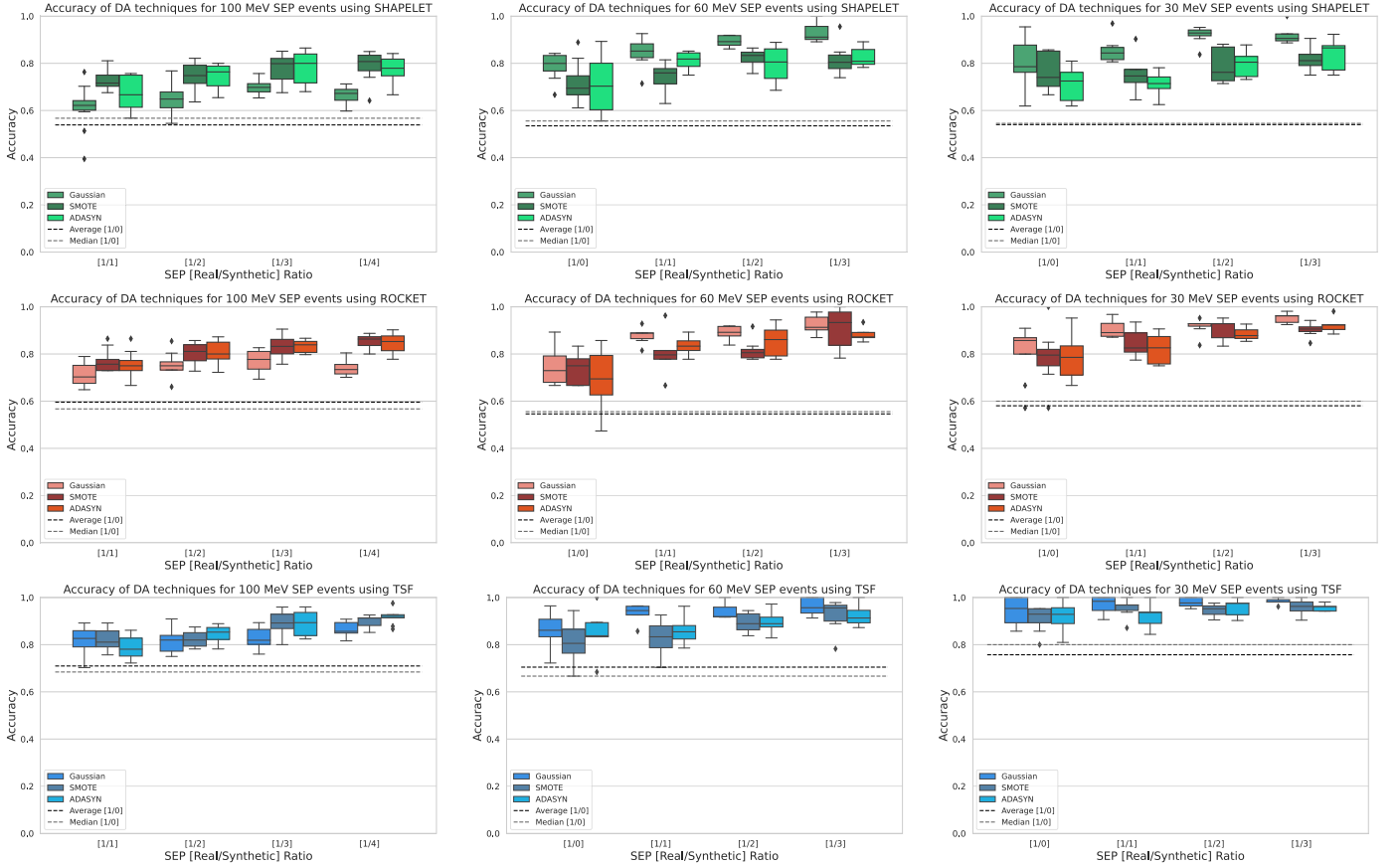

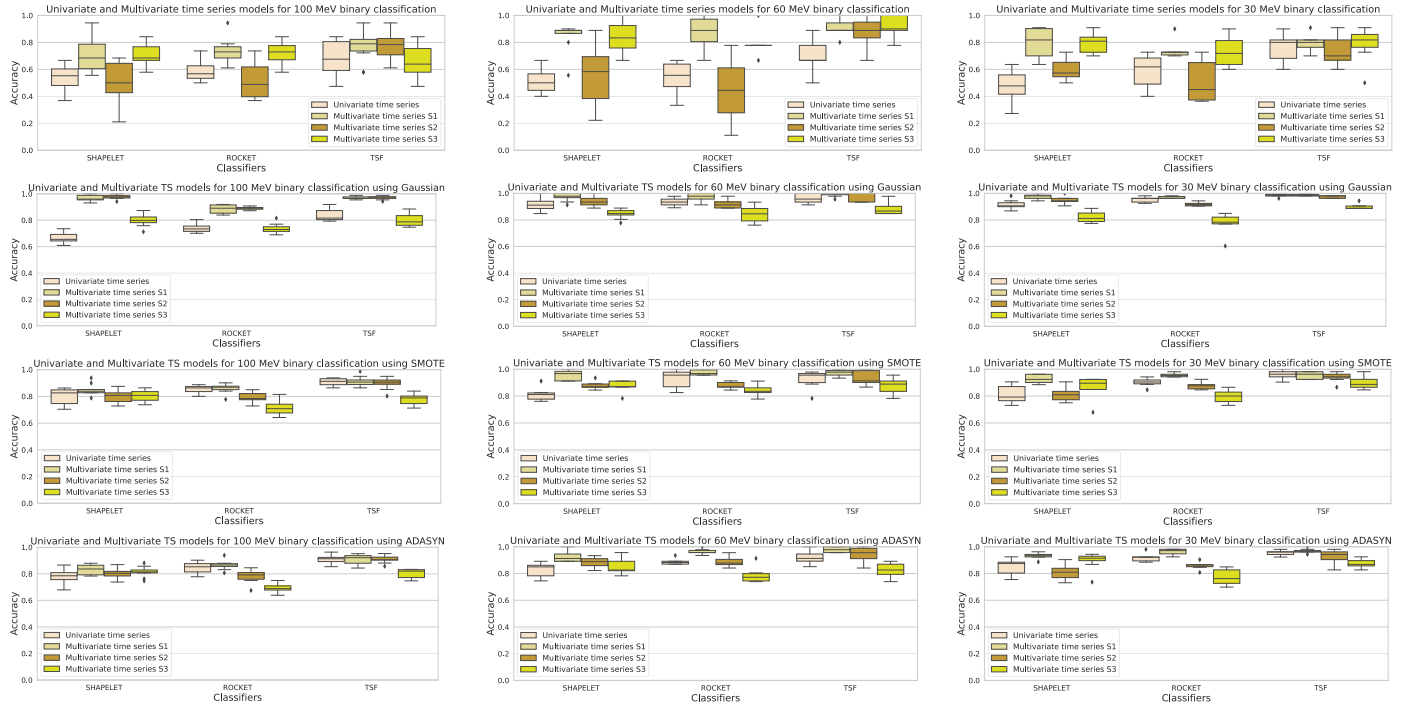**Figure 11.** Comparison of the data augmentation models with the specific classifier used for evaluation.

Figure 12. Comparison between univariate and multivariate time series models for SEP/NSEP binary classification in 100, 60, and 30 MeV, separately. Multivariate time series models include three strategies: concatenation, aggregation, and statistics.

In our exploration of different strategies for multivariate time series data, we carefully considered the implications of each strategy on the physical meaning and interpretability of the input data. These considerations underscore the trade-offs inherent in our multivariate time series strategies. The choice of strategy depends on the specific goals of the predictive model. Strategies 1 and 2 offer a broader view of the proton flux, whereas Strategy 3 focuses on key statistical measures. Our goal was to evaluate these strategies rigorously to determine which one strikes the right balance between maintaining crucial physical information and achieving superior classification performance. Figure 12 indicates that in nearly all scenarios multivariate time series strategies 1 and 2 outperform the univariate time series strategy. Consequently, our multivariate models improve the performance of SEP/non-SEP binary classification.

4.4. Four-class Classification

Previously, we focused on binary classification, distinguishing between the occurrence and absence of SEP events. In this subsection, we shift our focus to multiclass classification, a task encompassing not only event prediction but also the assignment of specific energy band labels to SEP events (100, 60, 30 MeV), along with a non-SEP classification. This shift is motivated by the advantages offered by multivariate time series data, as elaborated in Section 4.3. In this context, multivariate time series data entail the fusion of information from three distinct proton flux channels: P4, P5, and P6. These channels are employed to construct multivariate time series proton flux data for each event belonging to the four-class categorization, which includes the three SEP energy bands (100, 60, 30 MeV) and the non-SEP category.

Four different scenarios are defined to analyze the performance of classification before and after using data augmentation techniques. These scenarios serve as the basis for assessing classification performance both before and after the application of data augmentation techniques. As shown in Figure 13, in the baseline scenario without any data augmentation, the models exhibit relatively lower accuracy when confronted with the complexities of the four-class classification problem. Here the models are challenged not only to distinguish between SEP and non-SEP events but also to correctly categorize them into distinct energy bands.

Figure 13, along with the results presented, provides a quantitative assessment of our strategies. It highlights the effectiveness of our models in the challenging four-class classification scenario. However, it is important to note that the four-class classification model, while demonstrating promise, is yet to reach the same levels of performance achieved in the binary classification scenario. These experiments yield valuable insights into the models' performance, their limitations, and avenues for further refinement.

4.5. Hierarchical Classification Framework

In the previous sections, we discussed the classification results of different models for the binary classification case (i.e., specific energy band events against non-SEP events) and the four-class classification case. In this subsection, we train a hierarchical model that uses the binary classification methods in a hierarchical approach, as defined in Section 3.3. To provide a thorough understanding of the model's potential, we consider the model's performance in two key contexts: one with no data augmentation techniques, and another after the application of the SMOTE data augmentation method on univariate time series data. These contexts help us assess the model's adaptability and its response to enhanced data. As shown in Figure 14, it is evident that, across the board, applying data augmentation techniques contributes to improved accuracy for
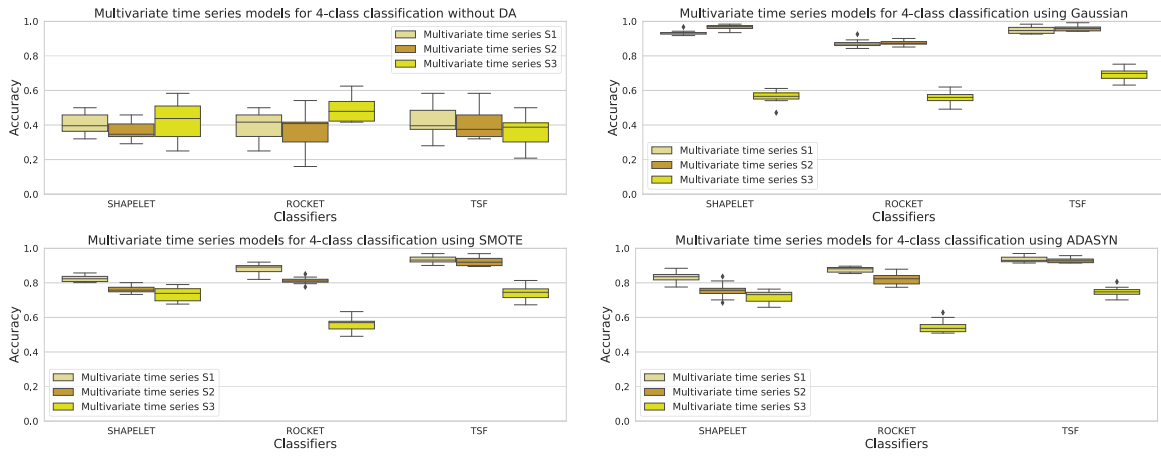
Figure 13. Evaluation of different strategies of multivariate time series data for four-class classification with NSEP and 100, 60, and 30 MeV labels. Results are demonstrated before and after using data augmentation techniques.
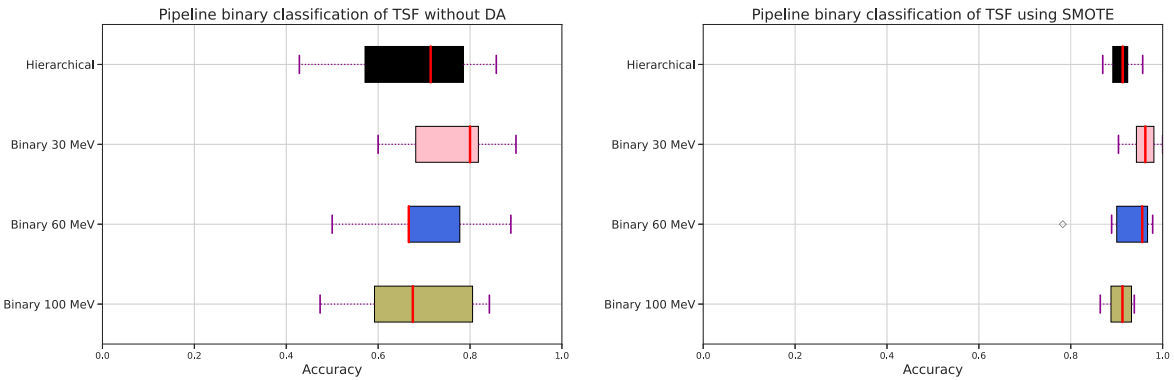

Figure 14. Accuracy of binary TSF classification before and after using SMOTE data augmentation.

all models. This augmentation process, a recurring theme throughout our experiments, is crucial in enhancing predictive capabilities. Data augmentation provides the models with a broader knowledge base, which in turn refines their predictive accuracy.

In certain practical applications, such as space weather forecasting, resource allocation is a critical consideration. The hierarchical model offers the advantage of being able to allocate resources based on the predicted energy band. High-energy events may require a more urgent response or special measures compared to lower-energy events. Hierarchical classification allows for a stratified response, ensuring that resources are allocated in alignment with the potential severity of the event.

While the hierarchical model builds on the foundation of binary classification trained models, it is worth noting that it exhibits a slightly lower accuracy compared to individual binary models. This variance is not a reflection of inefficiency but an inherent characteristic of the hierarchical model's operational dynamics. Specifically, the hierarchical model employs a decision hierarchy, which entails a sequential evaluation process. Predictions are made step by step, starting with the highest energy band and progressing to the lower energy bands. This structured approach, while powerful, has a notable consequence in the event of a high-energy band prediction turning out to be a false positive. In such instances, where an erroneous high-energy prediction occurs, the hierarchical model halts the prediction process prematurely.

As a result, the prediction may be stopped before accurately identifying lower energy bands. This specific mode of operation introduces a minor reduction in the overall accuracy of the model.

4.6. Sequence Analysis

In this section, we perform a sequence analysis on various observation window sizes using multiple models and data augmentation methods. As shown in Figure 15, we initially carried out a sequence analysis for SHAPELET, ROCKET, and TSF classifiers without utilizing any data augmentation techniques. We considered observation windows of size 5, 10, and 15 hr. Therefore, the number of values in each sequence will be 60, 120, and 180, respectively. We conducted individual analyses for 100, 60, and 30 MeV binary classifications. The same analysis was performed on the classification when using Gaussian, SMOTE, and ADASYN.

While the intuitive expectation was that larger observation windows would consistently improve classification results, our findings provided a more nuanced perspective. The impact of observation window size on model performance is multifaceted. We observed that the effectiveness of a larger observation window depends on several factors, including the specific energy band under consideration and the complexity of the underlying proton flux patterns. Notably, for lower energy bands such as 30 and 60 MeV classifications, which often exhibit more extended durations and intricate temporal patterns, larger observation windows do offer performance

Figure 15. Analysis of different models in terms of average accuracy of the k-fold cross-validation. SHAPELET, ROCKET, and TSF are used for classifying SEP and non-SEP events using observation windows of size 5, 10, and 15 hr. These heatmaps include classification without data augmentation techniques and also with three data augmentation techniques.

benefits. In these scenarios, the models benefit from the increased contextual information and richer history of proton flux fluctuations.

However, for the 100 MeV classification, the advantages of a larger observation window proved to be less consistent. Although, in certain situations, the application of a 15 hr observation window improved the model's accuracy, in other cases the models exhibited superior performance with a 5 hr window. This variability implies that the impact of observation window size is intricately associated with the specific characteristics of the events under prediction.

The second sequence analysis experiment we conducted is using the multivariate time series binary classification for 100 MeV events using the 15 hr optimal window size. As shown in Figure 16, the classifiers achieved remarkable results, with accuracy levels exceeding 74%. The average accuracy for SHAPELET, ROCKET, and TSF was recorded as 74%, 78%, and 77%, respectively. The figure also shows improvements achieved by using the Gaussian data augmentation method. It can be observed that Figure 16 shows improvements in 100 MeV event classification compared to the results achieved in Figure 7.

5. Conclusion and Future Work

In this study, we explored the prediction of SEP events in various energy bands (∼100, ∼60, and ∼30 MeV) using ML-based time series classification models. We have assessed the predictive accuracy across different scenarios, emphasizing the added value of data augmentation techniques in improving model performance. While SHAPELET and ROCKET benefited significantly from these enhancements, TSF consistently exhibited superior performance across the three energy bands. Additionally, the use of multivariate time series data has led to a substantial increase in the accuracy of SEP event predictions.

Our work aligns with the broader landscape of space weather forecasting, a field critical for safeguarding space missions, astronauts, and technological assets. To contextualize our contributions and acknowledge the relevant advancements in the space weather domain, we refer to notable studies that have explored diverse aspects of SEP event prediction and space weather modeling.

The study by Núñez & Paul-Pena (2020) is crucial in understanding the use of solar flare proxies as predictors for SEP events, showcasing the potential for anticipatory models.
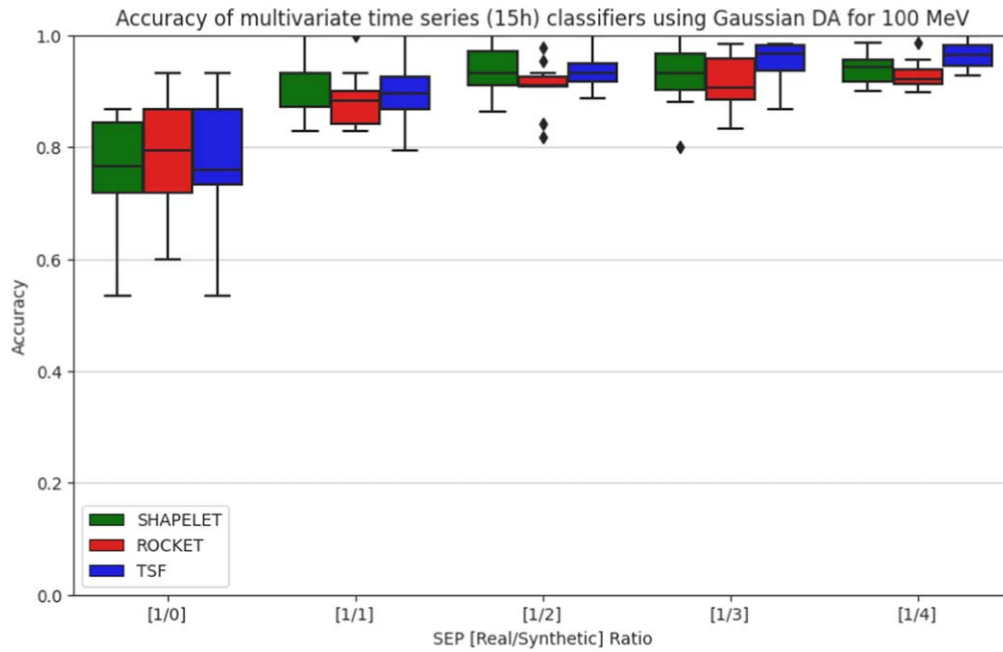
Figure 16. Performance of multivariate time series classifiers for predicting 100 MeV events using a 15 hr observation window. It also shows improvements achieved by using the Gaussian data augmentation method.

Posner and their research team, as discussed in their work (Posner & Strauss 2020), delve into early warnings for SEP events during Mars missions, showcasing the utility of advance alerts for radiation exposure management. Damiani et al. (2009) shed light on the mesospheric and stratospheric response to SEPs, highlighting the complexities of atmospheric chemistry induced by these events. The study by Richardson et al. (2018) presents a formula that predicts SEP intensity based on CMEs, offering a valuable empirical tool for forecasting SEP events.

In this broader context, our research contributes to the ongoing efforts in space weather prediction and fortifies the foundation for improved SEP event forecasting. Our future work will encompass real-time prediction, parameter prediction, and the integration of complementary data sets, aligning with the evolving landscape of space weather modeling.

As space exploration endeavors continue to evolve, the accuracy and reliability of SEP event forecasts assume paramount importance. By combining the latest advancements with a keen understanding of historical data, our work aims to enhance the safety and success of future space missions.

Appendix

Figure 17 depicts a comprehensive comparison of time series models utilizing different data augmentation methods (ARIMA, GARCH, SMOTE, and ADASYN) across four distinct partitions for the ∼100 MeV data set. This comparative analysis offers a nuanced understanding of the performance variations among these techniques, shedding light on their efficacy in handling the inherent complexities of SEP event prediction. Furthermore, Figure 18 presents a visual representation comparing the input time series (highlighted in red) with the SMOTE-generated time series (displayed in green). This comparative visualization serves as a compelling demonstration of the alignment between the original solar proton flux data and the synthetic data generated by the SMOTE technique. By showcasing the similarities between the input and augmented data, this figure underscores the ability of SMOTE to generate synthetic data that closely mirror the characteristics and patterns of the original time series, thereby validating its utility in enhancing forecasting models for solar proton flux.
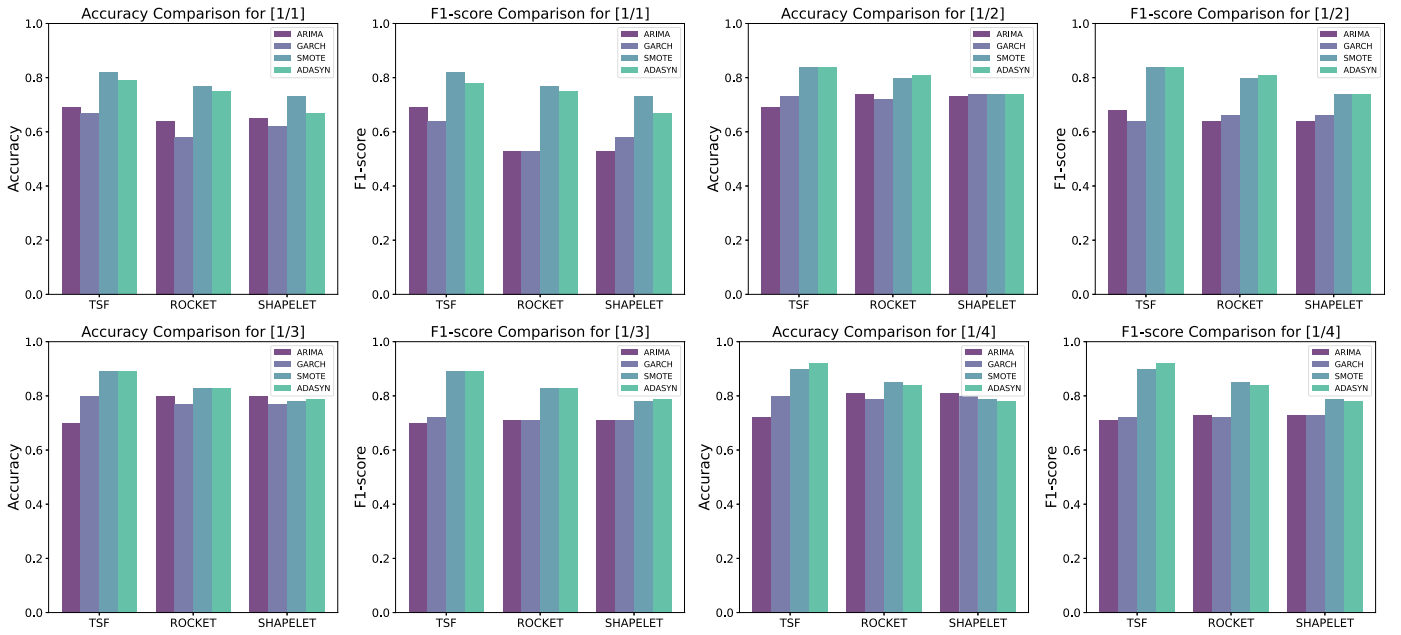
Figure 17. Comparison between the performance of time series models using data augmentation techniques with ARIMA, GARCH, SMOTE, and ADASYN in four different partitions for a ~100 MeV data set.
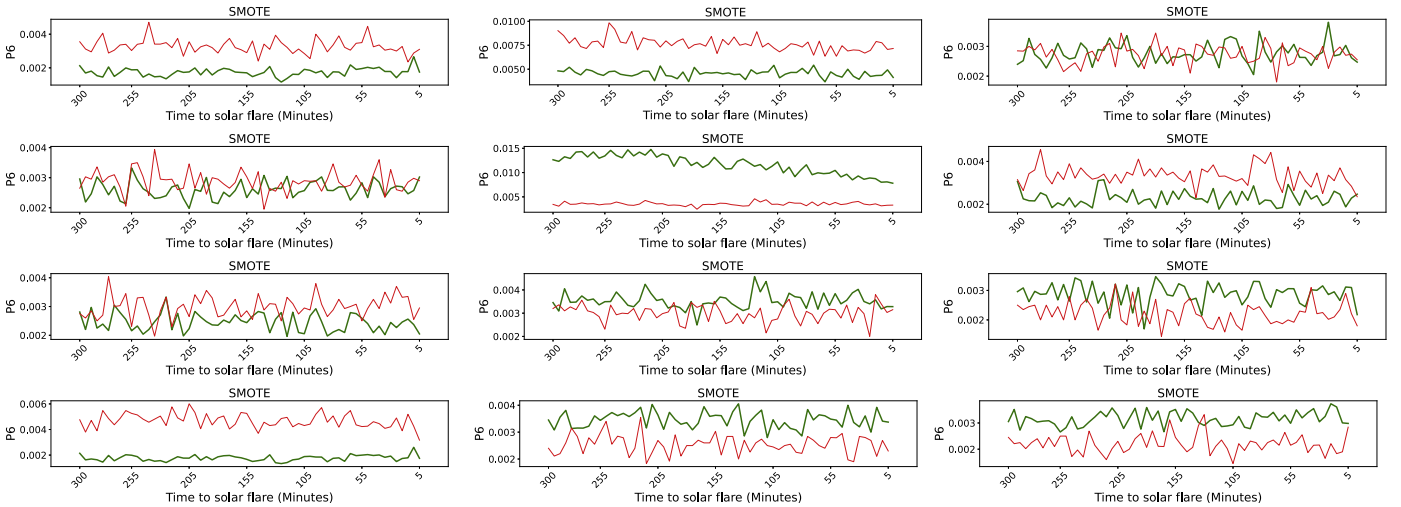

Figure 18. Comparison of input time series (in red) and SMOTE-generated time series (in green) highlighting their similarities.

ORCID iDs

Pouya Hosseinzadeh https://orcid.org/0000-0001-8045-2709
Soukaina Filali Boubrahimi https://orcid.org/0000-0001-5693-6383
Shah Muhammad Hamdi https://orcid.org/0000-0002-9303-7835

References

Alshammari, K., Hamdi, S. M., & Boubrahimi, S. F. 2022, in 2022 IEEE Int. Conf. on Big Data (Piscataway, NJ: IEEE), 4796
Aminalragia-Giamini, S., Raptis, S., Anastasiadis, A., et al. 2021, JSWSC, 11, 59
Aran, A., Sanahuja, B., & Lario, D. 2006, AdSpR, 37, 1240
Arul, M., & Kareem, A. 2021, EngSt, 228, 111564

Bagnall, A., Lines, J., Bostrom, A., Large, J., & Keogh, E. 2017, Data Min. Knowl. Disc., 31, 606
Bahri, O., Li, P., Boubrahimi, S. F., & Hamdi, S. M. 2022a, in 2022 IEEE Int. Conf. on Big Data (Piscataway, NJ: IEEE), 242
Bahri, O., Boubrahimi, S. F., & Hamdi, S. M. 2022b, arXiv:2208.10462
Bahri, O., Hosseinzadeh, P., Filali Boubrahimi, S., & Hamdi, S. M. 2022c, SHINE 2022 Workshop, 181
Bahri, O., Li, P., Boubrahimi, S. F., & Hamdi, S. M. 2023a, in Big Data Analytics and Knowledge Discovery, ed. R. Wrembel (Cham: Springer), 42
Bain, H., Brea, P., & Adamson, E. 2018, AGUFM, 2018, SM31D–3530
Bobra, M. G., & Couvidat, S. 2015, ApJ, 798, 135
Boubrahimi, S. F., Aydin, B., Kempton, D., & Angryk, R. 2016, in 2016 IEEE Int. Conf. on Big Data (Piscataway, NJ: IEEE), 3149
Boubrahimi, S. F., Aydin, B., Martens, P., & Angryk, R. 2017, in 2017 IEEE Int. Conf. on Big Data (Piscataway, NJ: IEEE), 2533
Boubrahimi, S. F., Hamdi, S. M., Ma, R., & Angryk, R. 2020, in 2020 IEEE Int. Conf. on Big Data (Piscataway, NJ: IEEE), 493
Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. 2002, J. Artif. Intell. Res., 16, 321

Chen, Y., Kempton, D. J., Ahmadzadeh, A., & Angryk, R. A. 2021, in Artificial Intelligence and Soft Computing. ICAISC 2021, ed. L. Rutkowski (Cham: Springer), 296

Damiani, A., Diego, P., Laurenza, M., Storini, M., & Rafanelli, C. 2009, AdSpR, 43, 28

Dempster, A., Petitjean, F., & Webb, G. I. 2020, Data Min. Knowl. Disc., 34, 1454

Deng, H., Runger, G., Tuv, E., & Vladimir, M. 2013, Inform. Sci., 239, 142

Dhariyal, B., Le Nguyen, T., & Ifrim, G. 2023, Data Min. Knowl. Disc., 37, 1010

Fogtman, A., Baatout, S., Baselet, B., et al. 2023, npjMG, 9, 8

Forbush, S. E. 1946, PhRv, 70, 771

Girshick, R. 2015, in 2015 IEEE Int. Conf. on Computer Vision (Piscataway, NJ: IEEE), 1440

He, H., Bai, Y., Garcia, E. A., & Li, S. 2008, in 2008 IEEE Int. Joint Conf. on Neural Networks (IEEE World Congress on Computational Intelligence) (Piscataway, NJ: IEEE), 1322

Horv, L., & Kokoszka, P. 2003, Bernoulli, 9, 201

Hosseinzadeh, P., Nassar, A., Boubrahimi, S. F., & Hamdi, S. M. 2023, Hydrology, 10, 29

Hostetter, M., & Angryk, R. A. 2020, in 2020 IEEE Int. Conf. on Big Data (Piscataway, NJ: IEEE), 4208

Inceoglu, F., Jeppesen, J. H., Kongstad, P., et al. 2018, ApJ, 861, 128

Iwana, B. K., & Uchida, S. 2021, PLoSO, 16, e0254841

Lario, D. 2005, AdSpR, 36, 2279

Laurenza, M., Cliver, E., Hewitt, J., et al. 2009, SpWea, 7, S04008

Li, P., Bahri, O., Boubrahimi, S. F., & Hamdi, S. M. 2022a, in 2022 IEEE Int. Conf. on Big Data (Piscataway, NJ: IEEE), 1564

Li, P., Boubrahimi, S. F., & Hamdi, S. M. 2021, in 2021 20th IEEE Int. Conf. on Machine Learning and Applications (ICMLA) (Piscataway, NJ: IEEE), 1373

Li, P., Boubrahimi, S. F., & Hamdi, S. M. 2022b, in Pattern Recognition, Computer Vision, and Image Processing. ICPR 2022 International Workshops and Challenges, ed. J. J. Rosseau & B. Kapralos (Cham: Springer), 203

Luhmann, J., Ledvina, S., Odstrcil, D., et al. 2010, AdSpR, 46, 1

Marcus, G. 2018, arXiv:1801.00631

McKenna-Lawlor, S., Gonçalves, P., Keating, A., Reitz, G., & Matthiä, D. 2012, P&SS, 63, 123

Mewaldt, R. A., Cohen, C. M. S., Haggerty, D. K., et al. 2007, in AIP Conf. Proc. 932, Turbulence and Nonlinear Processes in Astrophysical Plasmas, ed. D. Shaikh & G. P. Zank (Melville, NY: AIP), 277

Moniz, N., Branco, P., & Torgo, L. 2017, Int. J. Data Sci. Anal., 3, 161

Núñez, M., & Paul-Pena, D. 2020, Univ, 6, 161

Posner, A., & Strauss, R. 2020, SpWea, 18, e2019SW002354

Richardson, I., Mays, M., & Thompson, B. 2018, SpWea, 16, 1862

Rigatti, S. J. 2017, J. Insur. Med., 47, 31

Rotti, S., Aydin, B., Georgoulis, M., & Martens, P. 2022, GSEP Dataset, v5, Harvard Dataverse, doi:10.7910/DVN/DZYLHK

Ruiz, A. P., Flynn, M., Large, J., Middlehurst, M., & Bagnall, A. 2021, Data Min. Knowl. Discov., 35, 401

Sato, T., Kataoka, R., Shiota, D., et al. 2018, SpWea, 16, 924

Singh, H., & Ray, M. R. 2021, in Advances in Civil Engineering and Infrastructural Development, ed. L. M. Gupta, M. R. Ray, & P. K. Labhasetwar (Singapore: Springer), 255

Stumpo, M., Benella, S., Laurenza, M., et al. 2021, SpWea, 19, e2021SW002794

Talavera, E., Iglesias, G., González-Prieto, Á., Mozo, A., & Gómez-Canaval, S. 2022, arXiv:2206.13508

Tenishev, V., Shou, Y., Borovikov, D., et al. 2021, JGRA, 126, e2020JA028242

Van der Maaten, L., & Hinton, G. 2008, JMLR, 9, 2579

Wen, Q., Sun, L., Yang, F., et al. 2020, arXiv:2002.12478

Whitman, K., Egeland, R., Richardson, I. G., et al. 2023, AdSpR, 72, 5161

Yang, X., Zhang, Z., Cui, X., & Cui, R. 2021, in 2021 Int. Conf. on Computer Communication and Artificial Intelligence (CCAI) (Piscataway, NJ: IEEE), 116

Ye, L., & Keogh, E. 2009, in Proc. of the 15th ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining, ed. J. Elder et al. (New York: Association for Computing Machinery), 947