Interpretable Solar Flare Prediction with Sliding Window Multivariate Time Series Forests

Anli Ji and Berkay Aydin
Department of Computer Science, Georgia State University
Atlanta, GA
{aji1, baydin2}@gsu.edu

***Abstract*—Recently, the synergy of physics-based feature engineering and data-intensive methods, including machine learning and deep learning, has ushered in a new era in the analysis and prediction of space weather forecasting, specifically for solar flare prediction. These sophisticated approaches play a pivotal role in understanding the complex mechanisms leading to solar flares, with a primary focus on forecasting these events and mitigating potential risks they pose to our planet. While current methodologies have made substantial advancements, they are not without limitations, and one particularly glaring limitation is the neglect of temporal evolution characteristics within the active regions from which solar flares originate. This oversight impairs the capacity of these methods to capture the intricate relationships among high-dimensional features of these active regions, thereby constraining their practical utility. Our study focuses on two key objectives: the development of interpretable classifiers for multivariate time series data and the introduction of an innovative feature ranking method using sliding window-based sub-interval ranking. The central contribution of our work lies in bridging the gap between complex, less interpretable "black-box" models typically employed for high-dimensional data and the exploration of pertinent sub-intervals within multivariate time series data, with a specific emphasis on solar flare forecasting. Our findings underscore the efficacy of our sliding-window time series forest classifier in solar flare prediction, achieving a True Skill Statistic of over 85%. Our approach is capable of pinpointing the most critical features and sub-intervals relevant to any given learning task. These results indicate significant progress toward improving the interpretability and accuracy of flare prediction models, further advancing our understanding of these impactful events.**

***Index Terms*—Multivariate Time Series Classification, Solar Flare Prediction, Interval-based Classification**

I. Introduction

Solar flares are intense localized eruptions of electromagnetic energy from the Sun's atmosphere that could last from minutes to hours. The sudden outburst of energy caused by solar flares travels approximately at the speed of light and can be accompanied by other solar phenomena such as coronal mass ejections (CMEs), which can potentially lead to devastating geomagnetic storms, massive radio collapse on the sunlit side of Earth, and disruption of sensitive space devices close to Earth. Current physics-based and data-driven (i.e., machine learning-based) methods [16] analyze these events as predictive learning tasks, with the most commonly used data products [18] derived from solar magnetograms.

In recent times, the field of flare prediction has witnessed an influx of methodologies and corresponding models, primarily approaching it as a classification task based on point-in-time measurements. However, these approaches often overlook the intrinsic time-dependent nature of the data, failing to account for its temporal evolution characteristics (as discussed in [8]). They tend to treat multiple physical observations as isolated entities, generating predictions solely based on instantaneous values [6]. Such an approach disregards the dynamic essence of solar flares, which exhibit characteristics intricately tied to the evolving behavior of solar active regions [5] [13] [14].

By analyzing the temporal characteristics of time series intervals, we have the potential to unveil implicit relationships and capture pertinent patterns. In our earlier work [10], we applied ensembles of interval-based time series classification models to multivariate time series data, leveraging a set of interval features. However, these models lack the capacity to handle a systematic evaluation of the underlying temporal characteristics for identifying critical predictive features due to two reasons: (1) random sub-interval sampling and (2) late fusion of derived features. Hence, the primary focus of this study is on interpreting statistical features extracted from multivariate time series. We aim to identify sequences of features and associated patterns using multi-scale sliding windows with varying interval sizes and step sizes, thus enabling a more comprehensive understanding of the temporal dynamics inherent in solar flare prediction.

The rest of the paper is structured as follows: Section II provides background information on existing time series classification models pertinent to flare prediction. In Section III, we provide our problem formulation and introduce our multivariate time series classification model and feature ranking method used for extracting relevant feature intervals from provided time series data. Our experimental setup and evaluation framework are presented in Section IV. Finally, Section V provides conclusions drawn from our study and discusses potential avenues for future research.

II. Related Work

With the proliferation of available time series data sets [17] and a wide spectrum of machine learning-based techniques proposed for time series classification, commonly used algorithms make predictions by measuring the similarity between training/testing instances [4], [3], [15] or by utilizing temporal features extracted from full time series or subsequences within these time series instances to capture associations between target variables and time series instances. For instance, in [12], basic statistical features (e.g., mean and standard deviation) were extracted from global time series and used as input for a multi-layer perceptron network, yet neglecting localized informative characteristics and properties.

Feature-based classification methods face challenges when dealing with multivariate time series data, as they require additional intricate information across features. Generating discriminating features in high-dimensional spaces becomes difficult due to the unknown interrelations among input parameters of the multivariate time series, therefore, adding complexity to model construction. Various techniques focus on extracting relevant features in multivariate aspects and then applying traditional machine learning algorithms for classification. Methods such as Shapelet-based decision trees [19] can effectively capture local patterns in multivariate time series data but also can be computationally expensive and hard to identify relevant shapelets (especially in high dimensional spaces) that are both informative and applicable across dimensions. Another problem is that the shapelets extracted from one multivariate time series dataset might not generalize well to other datasets with different dimensionality, characteristics, and patterns. To mitigate these issues, Time Series Forest (TSF) [7] incorporates subseries, but instead of relying on distance measurement from learned subsequences, it derives summarized statistical features within randomly selected time series. It treats each time step as a separate component and constructs decision trees in each feature dimension to capture temporal relationships and reduce the high-dimensional feature space. However, important interrelationships between different components of the time series cannot be fully captured by treating them as separate features, which can lead to a loss of inter-channel relationships and dependencies that are often crucial in multivariate time series data. Understanding the combined effects of multiple trees on multivariate time series data might be less intuitive compared to single decision trees.

Many of these methods focus solely on understanding how each feature behaves on its own, without considering how different features might interact. There are instances where a particular relationship within a single time series parameter might be significant for a specific feature, but not necessarily for others. When trying to generate features that effectively differentiate classes, it is more advantageous to select the most relevant time intervals to create a more robust model. However, identifying these relevant intervals is not an easy task as they typically cannot be directly determined and require an expensive search across the entire series. Being able to extract the underlying mutual information present in these relevant intervals can enhance our understanding of the predictive process and accelerate the transition from research to operations in flare forecasting models. Our objective in this work is to establish a framework that can recognize these characteristics and offer deeper insights into the behavior of classifiers during prediction tasks.

III. Methodology

In this section, we will outline our approach, including the statistical features derived from time intervals, the introduction of the sliding-window time series forest, and the technique we employ to rank our features.

1) Problem Formulation: The sliding-window multivariate time series forest is an early fusion, interval-based ensemble classification method. We present an illustration of the overview of how we generate features with a sliding window operation in Fig. 1. It employs a multitude of preferably short decision trees, similar to random forests, that make use of interval-based features extracted from all univariate time series using multi-scale sliding windows. By combining the features from univariate time series early, we aim to understand the relationships among these features, utilizing an embedded feature ranking based on mutual information. Formal definitions and explanations for processing the multivariate time series and extraction of vectorized features are provided from our previous research [9].

2) Interval Features: To generate well-structured and relevant intervals, we calculate statistical characteristics for intervals, including mean, standard deviation, and slope. In addition, we derive additional transformed features (i.e., maximum, minimum, and mean) comprising an additional localized pooling procedure, which is used on the individual interval features extracted from a set of consecutive intervals obtained after sliding window operation. In this process, all potential interval sets originating from the same time series are collected, and pooling functions are applied for consolidation. Essentially, during this stage, we consider the highest, lowest, and average values of statistical properties from each parameter of each subseries obtained through sliding window operations. A detailed description of the extraction and transformation process can be found in our previous research [9].

3) ***Sli****ding Window* ***M****ultivariate* ***T****ime* ***S****eries* ***F****orest:* After extracting interval features from subsequences obtained after sliding window operation and applying secondary transformation to these statistical attributes, we merge these two groups of derived features into an input vector. This vector serves as the foundation for creating a versatile time series classifier we refer to as *Slim-TSF*. Among the wide array of supervised learning models available for making predictions, we have chosen random forest classifiers for two reasons: (1) their effectiveness and resilience when dealing with noisy, high-dimensional data and (2) due to their ability to select the most relevant features from a given dataset w.r.t. a target feature.

It is important to highlight that depending on the chosen parameter settings, such as using smaller window and step sizes, the interval feature vectors' data space can expand considerably. Additionally, the process of vectorization, based on the sliding window approach, may generate data points that exhibit some degree of correlation and potential noise. Consequently, it is important to systematically identify and remove these features. This is achieved through the application of information-theoretic relevance metrics (e.g., Gini index or

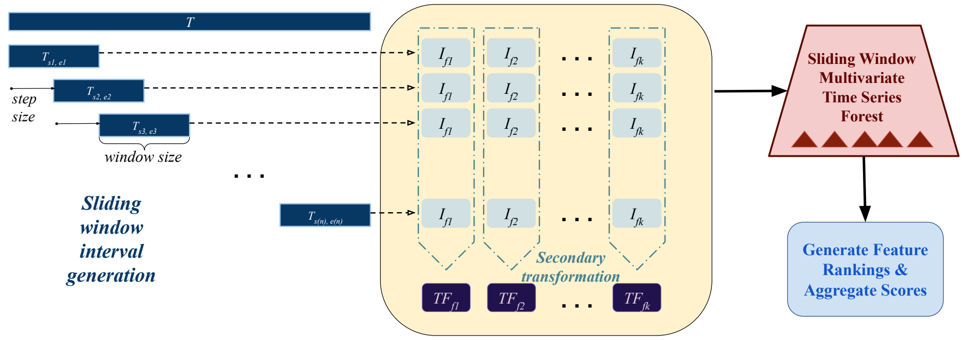


Fig. 1: The overview of the sliding window-based statistical feature generation. We first generate subsequences (intervals) with a fixed-size sliding window and step size. Then, we create vectorized features from these intervals where these features can be used as input for the sliding window multivariate time series forest (a random forest built on multivariate time series features) and features are ranked with aggregated relevance scores.

entropy). This meticulous feature selection process ensures the efficacy of our approach by retaining only the most informative attributes while discarding redundant ones.

4) Feature Ranking: An important aspect of our proposed Slim-TSF model is its ability to provide a synchronized and integrated feature selection and learning process. For ranking the features, we make use of the feature importance scores, which quantify the relevance and contribution of individual features derived from univariate intervals. Various feature importance scores can be used (such as the mean decrease in impurity (i.e., the impact of a feature on reducing classification uncertainty within the model) or feature permutation (i.e., the extent to which randomly shuffling features' values affects the model's performance)). In essence, for ranking the features, the sorted feature importance scores from a series of trained Slim-TSF models are aggregated. Note that the scores are given for vectorized input data for random forests; however, we are interested in the ranking of specific intervals of parameters.

Our ranking methodology involves a systematic process conducted through multiple experiments, denoted by the total number N, each executed with distinct experimental configurations. This is analogous to a grid search process. These experiments yield individual feature rankings, denoted as exp_j, where j signifies a specific experiment. The ranking, denoted as r, is a mapping that assigns a rank i to each feature, reflecting its position within the ranking. In each experiment, the features are ranked to create a specific ordering, denoted as $r_{f,i}$, which designates that feature f has achieved the i^{th} rank in that particular experiment. Subsequently, the top-k features selected for inclusion in the selected feature set, denoted as SFS_j, within each individual experiment j are determined from the ranking r (i.e., include features whose rank is less than or equal to k). This selected feature set is represented as $\{r_{f,1}, r_{f,2}, ..., r_{f,k}\}$. In the end, the selected feature set across all experiments are aggregated by summing the sparse representation of top-k membership vectors ($\widehat{SFS_j}$) from each experiment (as in Eq. 1).

$$SFS = \sum_{j=1,N} \widehat{SFS_j} \tag{1}$$

This approach allows for a systematic and consistent method of selecting top features across multiple experiments, enhancing the robustness and reliability of the feature selection process. Furthermore, we create a counting vector per each interval of each parameter, denoted as ct_v to represent the value counts of individual intervals in the selected feature set SFS. This counting vector serves as a transformation function, indicating the frequency with which a given interval appears within the top-k selections of the feature set.

IV. Experimental Evaluations

The experiments conducted in this study are designed with two primary objectives. Firstly, they aim to demonstrate the effectiveness of time series classifiers developed using distinct interval features and to perform a comprehensive performance comparison among them. Another key objective is to identify the intervals within the time series that hold the greatest relevance to the initial time series. This effort is primarily designed to offer interpretable insights into our model. It involves pinpointing the specific segments of the time series that exert significant influence on predictions and understanding the aggregation strategies that can lead to more accurate outcomes.

A. Data Collection

For the solar flare prediction task, we utilized the SWAN-SF dataset, an open-source multivariate time series dataset introduced in [2]. This dataset provides comprehensive space weather-related physical parameters derived from solar magnetograms, integrating data from various solar active regions and flare observations. Our experiments (both classification and feature ranking) include the entire 24 active region parameters available. These parameters were chosen as they are widely recognized as highly representative features of solar activity.

Notably, flare intensity is determined by the logarithmic classification of peak X-ray flux, categorized into major flaring classes (X, M, C, B, or A). For our analysis, we consider instances labeled with M- and X-class flares as flaring (i.e., positive class), while relatively weaker C- and B-class flares and flare-quiet regions as non-flaring (i.e., negative class). This binary classification approach allows us to model the flare forecasting problem as a binary multivariate time series classification task. Throughout our studies, we employ Partitions 1, 2, 3, and 5 for training purposes, reserving Partition 4 for testing. Partition 4 was chosen due to its relatively balanced distribution of flaring and non-flaring instances, facilitating a more rigorous evaluation of classifier performance.

B. Experimental Settings

In assessing the performance of our model, we adopted a binary 2×2 contingency matrix, complemented by a set of evaluation metrics suitable for assessing the forecast accuracy. In this context, True Positives (TPs) and True Negatives represent instances where the model correctly predicts a flaring or non-flaring event, respectively. False Positives (FPs) represent false alarms (incorrect flaring predictions) and False Negatives represent misses (incorrect non-flaring predictions).

Within our study, we employ two widely recognized skill scores for rigorous evaluation: the True Skill Statistic score (TSS, in Eq. 2) and the Heidke Skill Score (HSS, in Eq. 3). The TSS score quantifies the disparity between the Probability of Detection (i.e., recall for the positive class) and the Probability of False Detection (POFD). The HSS score measures the forecast's improvement over a climatology-aware random prediction.

$$TSS = \frac{TP}{TP+FN} - \frac{FP}{FP+TN} \tag{2}$$

$$HSS = \frac{2 \cdot ((TP \cdot TN) - (FN \cdot FP))}{P \cdot (FN+TN) + N \cdot (TP+FP)} \tag{3}$$

C. Experiment 1: Window Size Comparison

This experiment aims to investigate the relationship within our multivariate time series dataset under different window size configurations so that we can reveal interval features that hold greater significance for a predictive task, thereby facilitating the interpretation of the model's feature selection process. To pinpoint intervals of interest, we employ our fixed-size sliding window approach with three specific settings: window sizes of 8, 15, and 30, each corresponding to step sizes of 4, 8, and 15, respectively. This methodology effectively partitions the original time series into additional intervals based on these pre-defined window sizes. Subsequently, we associate these candidate intervals with statistical functions, encompassing measures such as mean, standard deviation, and slope, to generate descriptive interval features. The primary objective is to assess the predictive capabilities of these derived features. To achieve this, we train our Slim-TSF classifier individually for each window size setting.

It is worth noting that these models are trained on a notably imbalanced class distribution, with a ratio of approximately 1:50 between our positive (flaring) and negative (non-flaring) classes, following the original SWAN-SF benchmark dataset. To address this imbalance, we conduct a set of additional experiments involving class weight adjustments. Three distinct sub-experiments are conducted, each revealing unique insights into the use of only the sliding window slices, secondary transformations, and the combination of both.

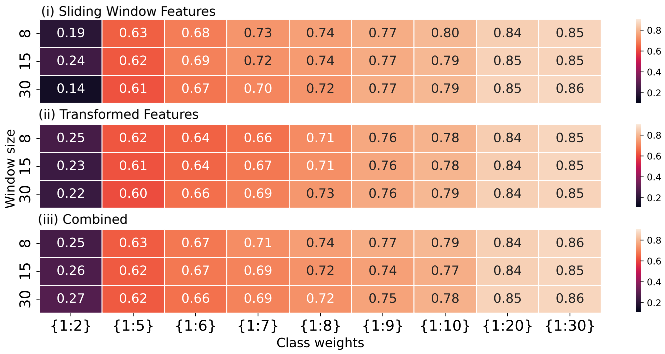
The outcomes of our experiment are presented in Fig. 2a and 2b, illustrating forecast skill scores measured by TSS and HSS as labeled heatmaps. This evaluation encompasses a range of class weight settings and pre-defined window sizes. Examining the results, it is evident that Slim-TSF models developed under various window size configurations are fairly consistent across different window size settings with respect to TSS and HSS. Overall the best results are achieved with features with window size=8 (on average TSS $\sim$68.9% and HSS $\sim$ 38.6%), but the difference between different window sizes is not substantial. We note that by using larger window sizes, one can improve the data and computational efficiency of the models, trading off the accuracy.

Furthermore, as we fine-tune class weights (from 1:5 to 1:10), we observe a well-known trend [1]. The TSS performance of our Slim-TSF model experiences a modest increment with higher class weight allocation for the XM class. However, this improvement comes at a trade-off, as it corresponds to a reduction in HSS. It is important to note that the overall effectiveness of a forecasting system can be compromised if all predictions are disproportionately assigned to the CBN classes, potentially leading to lower skill scores (closer to no-skill baselines). Although the TSS of the Slim-TSF model exhibits a marginal increase with higher XM class weight, it results in a relatively diminished HSS skill score and model robustness.

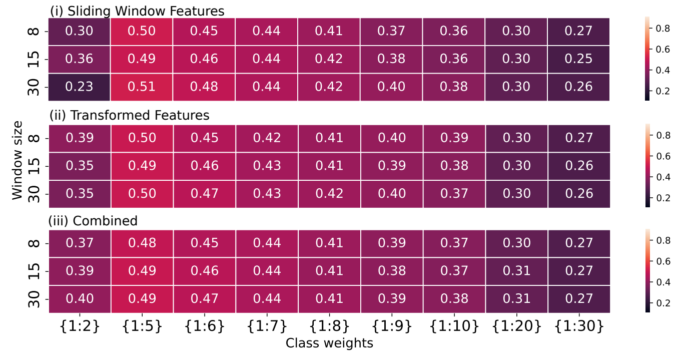
D. Experiment 2: Top-5 Selected Parameters

Building upon the results of the prior experiments, this part of our experiments focuses on assessing model performance with a specific focus on the most relevant parameter selection. Here, we repeat Experiment 1 with only the five most relevant parameters selected after determining the ranking of our derived features. In this experiment, our goal is to compare the predictive capabilities of Slim-TSF trained on these most representative parameters. We demonstrate the results of Experiment 2 in Fig. 2c and 2d.

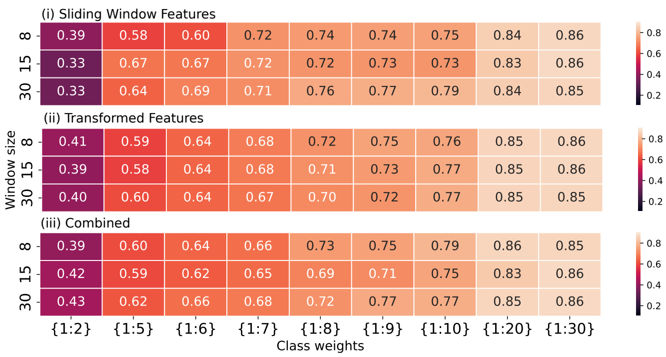
The results show that overall the performance of models trained with selected parameters is similar to those using all parameters in Experiment 1. We see significant improvements in models trained with class weight $\{1{:}2\}$. This effect can be attributed to the reduction in feature redundancy for more balanced class weight settings (leading to a more streamlined feature set). In other class weight settings, we observe up to 8% decrease in TSS and up to 5% decrease in HSS. Nevertheless, the best results (i.e., more balanced skill scores (e.g., TSS $\sim$0.77, HSS $\sim$0.40 at WS=30 and class weight=$\{1{:}9\}$) have fairly similar skill scores. This result demonstrates the overall effectiveness of our parameter ranking method and its applicability to reducing the computational cost.



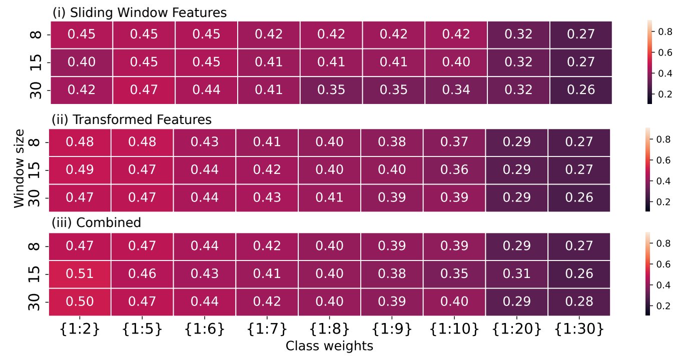
(a) TSS with 24 parameters



(b) HSS with 24 parameters



(c) TSS with top-5 parameters



(d) HSS with top-5 parameters

Fig. 2: Heatmap representation of sliding window multivariate time series forest evaluation with 24 parameters and evaluation with top-5 selected univariate parameters. The parameter selection process is performed post-hoc and models are retrained with five highly ranked parameters. The x-axis corresponds to different class weight settings while the y-axis corresponds to the resulting TSS (a) and HSS (b) scores for each experiment. Three sub-experiments are conducted: (i) using only the sliding window slices (ii) only the secondary transformation, and (iii) combining both.

E. Experiment 3: Ex-ante Feature Selection

In our last set of experiments, we investigate the effectiveness of early embedded future selection, this time applied directly to all statistical features. The rankings of the individual features are obtained after training a set of Slim-TSF classifiers from some considerably smaller random subsets of all training data. The subsets are obtained after stratified sampling with replacement (similar to bootstrapping). The aggregated rankings from classifiers are used to determine the Top-5 most relevant statistical interval features. Then, a Slim-TSF classifier is trained only with the Top-5 most relevant features. By doing so, we create a much leaner training dataset (using Top-5 features instead of hundreds) for our Slim-TSF models. This procedure is repeated with different class weight and window size settings similar to Experiments 1 and 2, and results are demonstrated in Fig. 3. Note that in this experiment, we only use the combined features (both interval and transformed features).

Our results show that Top-5 most relevant features found by ex-ante feature selection (early embedded selection with bootstrapping) have particularly high skill scores. We obtain our overall best performances from window size=15 setting (with class weights {1:6}, {1:7}, and {1:8}): TSS in 0.81 to 0.86 and HSS in 0.43 to 0.46. These scores are higher than both Exp 1 and Exp 2, suggesting that highly relevant statistical features, with hyperparameter optimization, can provide the best results. We note that the experiments with window size=30 settings show relatively lower performance, which can be attributed to a lack of fine-grained temporal information.

F. Remarks

Our experimental evaluations show primarily two things: (1) flare prediction models trained using the Slim-TSF method can effectively predict flares and (2) the embedded feature ranking method can convincingly be utilized such that it can help to identify the most relevant parameters and intervals within those parameters and/or significantly reduce data size while maintaining similar skill scores. Specifically, the ex-ante feature selection experiments (Exp 3) showed that highly skilled classifiers (with more balanced TSS and HSS) can be trained using much leaner datasets and much smaller forests.

We note that transformed (pooled) features have less impact and models solely trained on those features have lower skill scores. Across different window size settings, we generally observe better accuracy with lower window size values, meaning

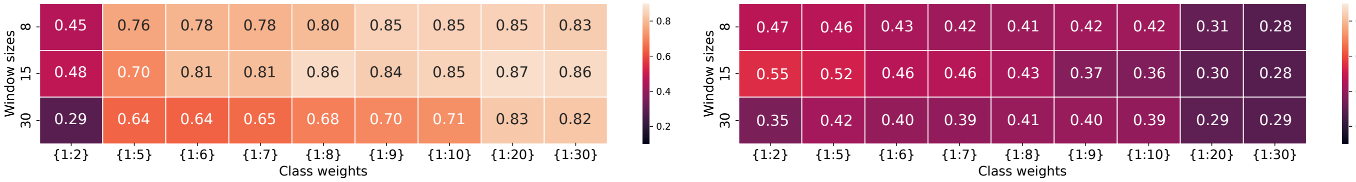


Fig. 3: Heatmap representation of sliding window multivariate time series forest evaluation with ex-ante feature selection. Top-5 most relevant features are selected (per each model trained) across different class weights and window size settings. The TSS (a) and HSS (b) scores for each experiment are shown. All features from sliding window intervals and transformed features are used.

the temporal granularity of sliced intervals plays an important role or in other words statistical features derived from higher window sizes lack the details important for predictive targets. Overall, our results show better performance in solar flare forecasting when compared to traditional time series forest models [10] and deep learning models [11].

V. Conclusions and Future Work

This study presents a novel multivariate time series classifier using interval-based features generated using sliding window operations that can also be used for ranking important features, intervals, and transformed pooling features. Our primary goal in this work is to improve the interpretability of high-dimensional multivariate time series classifiers. By combining interval-based feature ranking with random forest classification, our approach improves the interpretability of the trained model and enhances its predictive power and resilience to overfitting, thus making it suitable for a wide range of classification tasks with potentially noisy high-dimensional data. An important contribution of our work lies in improving the understandability of high-dimensional classification processes and the investigation of critical sub-intervals within multivariate time series, particularly their relevance in the context of solar flare prediction. As for future work, we plan to expand the temporal granularities of window size/step size combinations, explore additional ranking metrics, and explore other data fusion approaches.

Acknowledgment

This work is supported in part under two grants from NSF (Award #2104004) and NASA (SWR2O2R Grant #80NSSC22K0272).

References

[1] Ahmadzadeh, A., Aydin, B., Georgoulis, M.K., Kempton, D.J., Mahajan, S.S., Angryk, R.A.: How to train your flare prediction model: Revisiting robust sampling of rare events. The Astrophysical Journal Supplement Series **254**(2), 23 (2021)

[2] Angryk, R.A., Martens, P.C., Aydin, B., Kempton, D., Mahajan, S.S., Basodi, S., Ahmadzadeh, A., Cai, X., Boubrahimi, S.F., Hamdi, S.M., Schuh, M.A., Georgoulis, M.K.: Multivariate time series dataset for space weather data analytics. Scientific Data **7**(1) (Jul 2020). https://doi.org/10.1038/s41597-020-0548-x

[3] Bagnall, A., Lines, J., Bostrom, A., Large, J., Keogh, E.: The great time series classification bake off: a review and experimental evaluation of recent algorithmic advances. Data Mining and Knowledge Discovery **31**(3), 606–660 (Nov 2016). https://doi.org/10.1007/s10618-016-0483-9

[4] Baydogan, M.G., Runger, G., Tuv, E.: A bag-of-features framework to classify time series. IEEE Transactions on Pattern Analysis and Machine Intelligence **35**(11), 2796–2802 (2013). https://doi.org/10.1109/TPAMI.2013.72

[5] Benz, A.O.: Flare observations. Living Reviews in Solar Physics **5** (2008). https://doi.org/10.12942/lrsp-2008-1

[6] Boucheron, L.E., Al-Ghraibah, A., McAteer, R.T.J.: Prediction of solar flare size and time-to-flare using support vector machine regression. The Astrophysical Journal **812**(1), 51 (Oct 2015), https://doi.org/10.1088/0004-637x/812/1/51

[7] Deng, H., Runger, G., Tuv, E., Vladimir, M.: A time series forest for classification and feature extraction. Information Sciences **239**, 142–153 (Aug 2013). https://doi.org/10.1016/j.ins.2013.02.030

[8] Georgoulis, M.K.: On our ability to predict major solar flares. In: The Sun: New Challenges, pp. 93–104. Springer Berlin Heidelberg (2012). https://doi.org/10.1007/978-3-642-29417-4_9

[9] Ji, A., Aydin, B.: Active region-based flare forecasting with sliding window multivariate time series forest classifiers. In: The Fourth IEEE International Conference on Cognitive Machine Intelligence. IEEE (2023)

[10] Ji, A., Aydin, B., Georgoulis, M.K., Angryk, R.: All-clear flare prediction using interval-based time series classifiers. In: 2020 IEEE International Conference on Big Data (Big Data). pp. 4218–4225 (2020). https://doi.org/10.1109/BigData50022.2020.9377906

[11] Ji, A., Wen, J., Angryk, R., Aydin, B.: Solar flare forecasting with deep learning-based time series classifiers. In: 2022 26th International Conference on Pattern Recognition (ICPR). pp. 2907–2913. IEEE (2022)

[12] Nanopoulos, A., Alcock, R., Manolopoulos, Y.: Feature-Based Classification of Time-Series Data, p. 49–61. Nova Science Publishers, Inc., USA (2001)

[13] Pandey, C., Angryk, R.A., Aydin, B.: Solar flare forecasting with deep neural networks using compressed full-disk hmi magnetograms. In: 2021 IEEE International Conference on Big Data (Big Data). pp. 1725–1730 (2021). https://doi.org/10.1109/BigData52589.2021.9671322

[14] Pandey, C., Angryk, R.A., Aydin, B.: Explaining Full-Disk Deep Learning Model for Solar Flare Prediction Using Attribution Methods, p. 72–89. Springer Nature Switzerland (2023). https://doi.org/10.1007/978-3-031-43430-3_5

[15] Pandey, C., Ji, A., Nandakumar, T., Angryk, R.A., Aydin, B.: Exploring deep learning for full-disk solar flare prediction with empirical insights from guided grad-cam explanations. In: 2023 IEEE 10th International Conference on Data Science and Advanced Analytics (DSAA). pp. 1–10 (2023). https://doi.org/10.1109/DSAA60987.2023.10302639

[16] Shibata, K., Magara, T.: Solar flares: magnetohydrodynamic processes. Living Reviews in Solar Physics **8**(1), 6 (2011)

[17] Silva, D.F., Giusti, R., Keogh, E., Batista, G.E.A.P.A.: Speeding up similarity search under dynamic time warping by pruning unpromising alignments. Data Mining and Knowledge Discovery **32**(4), 988–1016 (Mar 2018). https://doi.org/10.1007/s10618-018-0557-y

[18] Song, H., Tan, C., Jing, J., Wang, H., Yurchyshyn, V., Abramenko, V.: Statistical assessment of photospheric magnetic features in imminent solar flare predictions. Solar Physics **254**(1), 101–125 (Nov 2008), https://doi.org/10.1007/s11207-008-9288-3

[19] Ye, L., Keogh, E.: Time series shapelets: a novel technique that allows accurate, interpretable and fast classification. Data Mining and Knowledge Discovery **22**(1-2), 149–182 (Jun 2010). https://doi.org/10.1007/s10618-010-0179-5