Contrastive Representation Learning for Predicting Solar Flares from Extremely Imbalanced Multivariate Time Series Data

Onur Vural, Shah Muhammad Hamdi, Soukaina Filali Boubrahimi
Department of Computer Science, Utah State University, Logan, UT 84322, USA
Emails: {onur.vural, s.hamdi, soukaina.boubrahimi}@usu.edu
ORCID: 0009-0004-4950-7520, 0000-0002-9303-7835, 0000-0001-5693-6383

arXiv:2410.00312v1 [astro-ph.SR] 1 Oct 2024

***Abstract*—Major solar flares are abrupt surges in the Sun's magnetic flux, presenting significant risks to technological infrastructure. In view of this, effectively predicting major flares from solar active region magnetic field data through machine learning methods becomes highly important in space weather research. Magnetic field data can be represented in multivariate time series modality where the data displays an extreme class imbalance due to the rarity of major flare events. In time series classification-based flare prediction, the use of contrastive representation learning methods has been relatively limited. In this paper, we introduce CONTREX, a novel contrastive representation learning approach for multivariate time series data, addressing challenges of temporal dependencies and extreme class imbalance. Our method involves extracting dynamic features from the multivariate time series instances, deriving two extremes from positive and negative class feature vectors that provide maximum separation capability, and training a sequence representation embedding module with the original multivariate time series data guided by our novel contrastive reconstruction loss to generate embeddings aligned with the extreme points. These embeddings capture essential time series characteristics and enhance discriminative power. Our approach shows promising solar flare prediction results on the Space Weather Analytics for Solar Flares (SWAN-SF) multivariate time series benchmark dataset against baseline methods.**

***Index Terms*—time series representation learning, contrastive representation learning, multivariate time series analysis, solar flare prediction, deep learning**

I. Introduction

In the domain of solar activity, a solar flare appears as a sudden surge in magnetic flux, originating from the Sun's surface and extending into the solar corona and heliosphere. Classified logarithmically based on the peak soft X-ray flux in the 1–8 Å wavelength range, flares are denoted by categories A, B, C, M, and X where M and X classes suggest intense flare activity [1]. Such solar phenomena may emit gamma-ray, x-ray, and extreme ultraviolet radiation, posing radiation-induced risks to astronauts, technological infrastructure, electronic devices, navigation, and communication systems [2]. The 1859 Carrington Event demonstrates the potential magnitude of a solar superstorm's impact, with a recurrence potentially causing prolonged blackouts and massive economic losses in our technology-dependent society [3]. Consequently, the heliophysics community underscores the importance of meticulous data analysis and the exploration of diverse methods for robust predictions of major flares from solar active region magnetic fields and flare data.

In 2020, the Space Weather Analytics for Solar Flares (SWAN-SF) dataset [3] was introduced as a pivotal resource in solar flare research. Derived from Space-weather HMI Active Region Patch (SHARP) solar photospheric vector magnetograms, SWAN-SF includes multivariate time series (MVTS) data from May 2010 to December 2018. It features 24 flare-predictive magnetic field parameters and over 10,000 flare reports, reformulating solar flare prediction as an MVTS classification task. Recent models built upon SWAN-SF MVTS data have shown enhanced efficacy in classifying flaring activities compared to earlier single timestamp models [3]. Due to major flares occurring rarely, the extreme class imbalance is a remarkable challenge in SWAN-SF. Accordingly, an effective learning methodology is representation learning, the process of learning meaningful fixed-dimension embeddings as data representations from the raw input data domain that keep their inherent features and have better transferability in downstream tasks. However, in the time series domain and particularly in solar flare tasks, the use of contrastive representation learning methods has been relatively limited. In contrastive methods, positive samples are contrasted with negative samples such that similar examples are mapped closer in the new feature space while the distance between dissimilar examples is maximized. Contrastive learning methods have been widely adopted in various domains for their soaring performance in representation learning, including vision, language, and graph-structured data [4]. Consequently, it becomes imperative to explore the impact of using contrastive approaches with the ultimate goal of enhancing flare prediction performance.

In this paper, we introduce CONTREX, a novel contrastive representation learning approach designed specifically for extremely imbalanced time series characteristics of data points. Our method consists of four main parts: first, we extract features from MVTS instances that capture their essential dynamical properties. Next, we derive two contrastive representations from the feature vectors of positive and negative instances that provide maximum separation capability. Then, we train a sequence representation embedding module with the original MVTS instances to generate embeddings that encapsulate time series characteristics, guided by our cus-

tom contrastive reconstruction loss. Finally, based on these embeddings, we employ a downstream classifier for binary classification. This comprehensive method enables effective representation learning and classification in time series analysis tasks. The contributions made by this paper are listed as:

- Introducing a novel contrastive learning framework tailored specifically for time series data that can be applied to both univariate and multivariate settings to address the challenges posed by temporal dependencies and extreme class imbalance.
- Defining a custom contrastive reconstruction loss during training of the sequence representation embedding module to guide the embeddings towards the positive and negative extremes, enhancing the discriminative power of the learned representations.
- Experimentally demonstrating the effectiveness of the proposed method regarding a performance metric evaluation reflecting the nature of the benchmark dataset.

II. Related Works

A. Solar Flare Prediction

One of the earliest endeavors in solar flare prediction history was THEO, an expert system that relied on human entries, officially used by The Space Environment Center (SEC) of the National Oceanic and Atmospheric Administration (NOAA) in 1987 [5]. Subsequently, as space-based and ground-based observatories amassed a wealth of magnetic field data, flare prediction transitioned into a data science task with models emerging based on line-of-sight and vector magnetograms, delineating solar active region photospheric magnetic field parameters. Since 2010, NASA's Solar Dynamics Observatory (SDO) has been continuously mapping the full-disk vector magnetic field every 12 minutes through the Helioseismic and Magnetic Imager (HMI) instrument, leading to reliance on this continuous vector magnetogram data in current literature [6]. Nonlinear statistical models, particularly machine learning classifiers have gained prominence in solar flare prediction such as logistic regression [7], C4.5 decision tree [8], fully connected neural network [9], support vector machine [10], and relevance vector machine [11]. The introduction of temporal window-based flare prediction [3] led to the creation of the SWAN-SF dataset [12], which records magnetic field data over time. Following SWAN-SF, various MVTS classification methods emerged, including kNN training with statistical summarization [13], MVTS decision trees with clustering [14], LSTM-based sequence modeling [15], and functional network embedding [16]. These advancements signify a shift from traditional linear models to sophisticated machine learning techniques in predicting solar flare activities.

B. Time Series Contrastive Representation Learning

In contrastive representation learning, unlike learning a mapping to labels as in discriminative models or reconstructing input samples as in generative models, data representations are learned through comparison between data points in the input space by mapping similar data points close while increasing the distance between dissimilar data points in the embedding space [4]. Although contrastive learning methods are less explored in the time series domain than vision, language, and graph domains, there is growing interest. Accordingly, sampling positive and negative instances and time pieces from the anchor to learn inter-sample and intra-temporal relations [17], temporal and contextual contrasting by creating two views for each sample with strong and weak augmentations [18], introducing fidelity and variety criteria, and creating a meta-learner for selecting feasible data augmentations [19], improving representation quality by instance-wise and temporal contrastive loss with soft assignments [20], employing a siamese structure and convolutional encoder to learn representations without negative pairs [21] were such works that highlight the potential of contrastive learning in enhancing time series representation learning.

III. Methodology

A. Extracting Dynamical Time Series Features

To capture the dynamical properties of time series, catch22 feature extraction method [22] is selected. The 22 features obtained by catch22 method give a low dimensional summary to represent the diverse and interpretable characteristics of time series including linear and non-linear autocorrelation, successive differences, value distributions and outliers, and fluctuation scaling properties. The catch22 method has shown highly discriminative and low redundancy feature representation power for various benchmark time series datasets [23]. In our approach, we extract catch22 features for each univariate time series in MVTS instances. Accordingly, for each MVTS data instance $M^{(\mathrm{k})} \in \mathbb{R}^{\tau \times N}$ having N parameters as univariate time series with a length of τ, we extract a fixed-dimensional multi-catch22 vector $V^{(\mathrm{k})} \in \mathbb{R}^{D}$ where D, the length of fixed-dimensional multi-catch22 vector, is equal to $22N$.

B. Obtaining Contrastive Extremes

After obtaining K fixed-dimensional multi-catch22 vectors $V^{(\mathrm{k})} \in \mathbb{R}^{D}$, K being total number of MVTS instances, the second step in our proposed method is to obtain two extreme points as overarching representations for positive and negative classes to enhance the contrastive power, effectively drawing positive data points closer to the positive extreme and negative data points closer to the negative extreme. Positive and negative extremes $E_P \in \mathbb{R}^{D}$ and $E_N \in \mathbb{R}^{D}$ are selected as multi-catch22 vectors that yield the complete linkage, representing the data points that yield the greatest distance between clusters:

$$D(P, N) = \max_{V_P \in P, V_N \in N} d(V_P, V_N) \tag{1}$$

In (1), $D(P, N)$ is the distance between clusters of positive and negative classes, and $d(V_P, V_N)$ is the Euclidian distance between positive multi-catch22 vector $V_P \in \mathbb{R}^{D}$ and negative multi-catch22 vector $V_N \in \mathbb{R}^{D}$.

C. Framework

CONTREX is composed of two modules integrated in an end-to-end manner. The initial module derives fixed-dimensional embeddings from MVTS data points, while the subsequent module leverages these embeddings to execute the binary classification task. Fig. 1 provides an overview of our architecture.

1) Obtaining Representation Embeddings: In the sequence representation embedding module, the sequence modeling of MVTS data points is done via a long short-term memory (LSTM) layer where each MVTS instance $M^{(\mathrm{k})} \in \mathbb{R}^{\tau \times N}$ is regarded as total τ timestamp vectors $x^{<t>} \in \mathbb{R}^{N}$, which are sequentially processed by the LSTM cells. The input size corresponds to N parameters, and the final hidden state representation $h^{<\tau>}$ outputs an internal embedding vector. Subsequently, this vector undergoes projection to a D-dimensional vector, mirroring the size of multi-catch22 vectors, facilitated by a multilayer perceptron (MLP) layer. Our network incorporates a single dropout layer to enhance robustness. After the training is complete, for the kth MVTS instance, the embedding vector $X_{label}^{(\mathrm{k})} \in \mathbb{R}^{D}$ can be extracted from the last layer.

2) Loss Function: Here, we propose a novel loss function, the contrastive reconstruction loss to guide the training of our sequence representation embedding module such that it will learn similar representations to the extremes in a supervised setting. Accordingly, for a positive class MVTS instance $X_{P}^{(\mathrm{k})} \in \mathbb{R}^{D}$, the mean squared error (MSE) loss is calculated against the positive extreme $E_P \in \mathbb{R}^{D}$ whereas for a negative class MVTS instance $X_{N}^{(\mathrm{k})} \in \mathbb{R}^{D}$, the MSE loss is calculated against the negative extreme $E_N \in \mathbb{R}^{D}$. This approach is designed to facilitate the contrastive learning of MVTS data points such that similar points that belong to the same class will be projected closer around their corresponding extreme points in the new representation space while data points that belong to different classes will be projected further apart. This personalized approach ensures that our embeddings are finely tuned to capture the distinctions between different class instances, facilitating accurate classification. In the end, $|P|$ and $|N|$ being the number of positive and negative class instances respectively, the final objective is to minimize this combined loss of:

$$\begin{aligned}\mathcal{L}_{\text{Extreme}} =& \frac{1}{|P|}\sum_{k=1}^{|P|}\sum_{i=1}^{D}(X_{P}^{(k)}[i] - E_P[i])^2 \\ &+ \frac{1}{|N|}\sum_{k=1}^{|N|}\sum_{i=1}^{D}(X_{N}^{(k)}[i] - E_N[i])^2\end{aligned} \tag{2}$$

3) Binary Classification of Representation Embeddings: In the second component of our framework, we use extracted representation embedding vectors $X_{label}^{(\mathrm{k})} \in \mathbb{R}^{D}$ as input instances to train a downstream classifier to provide a final binary class prediction.

IV. Experimental Evaluation

We present our experimental findings here. The source code for our model and experiments is on our GitHub repository. [1]

A. Benchmark Dataset

SWAN-SF was introduced as an MVTS dataset to facilitate unbiased flare forecasting, and drive advancements in solar flare prediction [3], [24], [25]. The data points in SWAN-SF are categorized into five distinct classes representing varying flare intensities. In binary solar flare prediction, we consider positive (P) examples as major flares responsible for health risks and infrastructural damages (i.e., M and X classes) and negative (N) examples as minor flares (i.e., B and C classes) and flare quiet events (i.e., FQ class) [13]. In SWAN-SF, a significant class imbalance exists due to the rarity of major flare events, with N examples greatly outnumbering P examples. This imbalance often results in a bias towards the majority class, yielding high true negative and low true positive rates, complicating the objective of accurately detecting solar flares [1]. Each data instance in the SWAN-SF dataset is an MVTS slice $M^{(\mathrm{k})} \in \mathbb{R}^{\tau \times N}$ representing a collection of univariate time series of 1-hour length τ, each having 24 magnetic field parameters [10] as N, extracted using a sliding window approach. Each time slice is labeled according to the most intense flare that occurs within the prediction window. The data is divided into five partitions (i.e., p1, p2, ..., p5), each covering a different observation period [3]. The distribution of classes is illustrated in detail in Fig. 2, underscoring the challenge posed to the primary objective of accurately detecting solar flares in flare-forecasting research.

B. Performance Evaluation Metrics

Due to the substantial P and N class imbalance in SWAN-SF, accuracy alone, which reports the number of correct predictions without going into class specifics, is inadequate. Therefore, we use several additional metrics that are commonly used in previous works of solar flare research for assessing binary solar flare classification: F1 score, receiver operating characteristic area under the curve (ROC AUC) that measures the classifier's ability to distinguish between classes, Heidke Skill Score 2 (HSS2) that evaluates improvement over random predictions, Gilbert Skill Score (GS) that assesses the likelihood of obtaining true positives by chance, and True Skill Statistic (TSS) [6], [10], [13], [26]. Among these, TSS is particularly robust against class imbalance, expressing the difference between true positive and false positive rates ranging from -1 (all incorrect predictions) to 1 (all correct predictions), with 0 indicating random predictions. TSS is recommended as the primary measure for evaluating solar flare prediction models [10]. The evaluation of classification results, involving true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN), can be expressed as:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{3}$$

[1] https://github.com/OnurVural/contrex

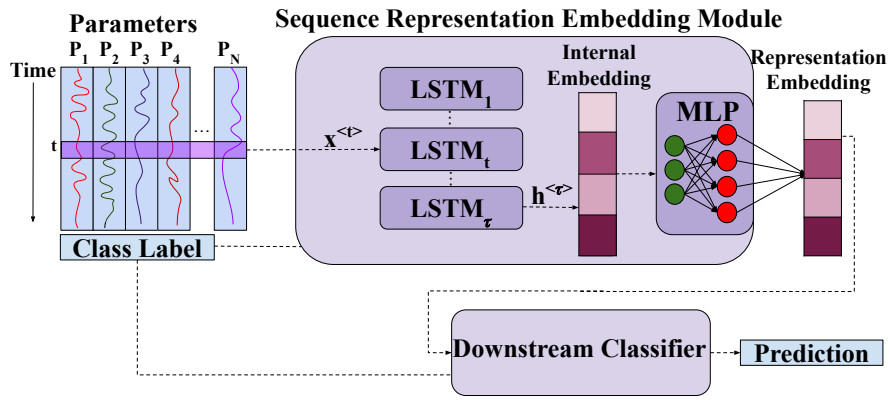

Fig. 1: CONTREX is composed of (1) a sequence representation embedding module to derive fixed-dimensional embeddings from MVTS data points, and (2) a downstream classifier that utilizes the representation embeddings for binary prediction.

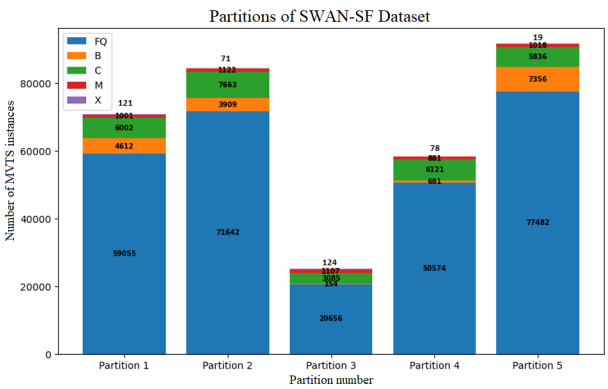

Fig. 2: Flare class distribution of each division in stacked bar plot format. For each partition, the count of five flare classes is shown in the plot.

$$F1 = \frac{TP}{TP + \frac{1}{2}(FP + FN)} \tag{4}$$

$$\text{HSS}_2 = \frac{2\times[(TP\times TN)-(FN\times FP)]}{[(TP+FN)\times(FN+TN)]+[(TN+FP)\times(TP+FP)]} \tag{5}$$

$$\text{GS} = \frac{TP - C}{TP + FP + FN - C} \tag{6}$$

$$\text{where } C = \frac{(TP + FP) \times (TP + FN)}{P + N} \tag{7}$$

$$\text{TSS} = \frac{TP}{TP + FN} - \frac{FP}{TN + FP} \tag{8}$$

C. Preprocessing and Training Settings

For preprocessing the MVTS instances of SWAN-SF, we performed KNN imputation [27] to fill the missing values followed by instancewise normalization of data points across individual time series features. For training and testing purposes, consequent SWAN-SF partitions are used respectively (e.g., p1 for obtaining the extremes and training our framework, p2 for testing). Our framework is trained with the following hyperparameter settings: LSTM with input dimension 24 (N as # time series in MVTS) and hidden state dimension 128, dropout probability: 0.5, MLP with input size 128 and output size 528 (D-dimension), # training epochs: 20, Adam learning rate: 10^{-2} from the search space of [10^{-1}, 10^{-2}, 10^{-3}, 10^{-4}].

D. Study of Components in CONTREX

We conducted several experiments to evaluate the effects of different components within the CONTREX framework. First, LSTM-based sequence modeling of MVTS data points was compared with other deep learning sequence models, namely recurrent neural network (RNN) and gated recurrent unit (GRU). This experiment also varied the hidden space dimensions to assess their impact on model underfitting or overfitting. Fig. 3 shows that the preferred model LSTM exhibits the best average performance in three consecutive partitions as train-test pairs (i.e., p1-p2, p2-p3, and p3-p4) under varying dimensionality in hidden space. Following that, the downstream classifier is selected as logistic regression showing the best average performance in the same three consecutive partitions as train-test pairs as reflected in Table I after experimenting with support vector machine (SVM), k-neighbors classifier (KNC), decision tree (DT), multilayer perception (MLP), and fully-connected network (FC).

E. Representation Embedding Analysis

To assess the extent of the contrastive abilities of our embeddings in separating classes, t-SNE visualization [28] is performed to map the extracted embeddings $X^{(k)}_{label} \in \mathbb{R}^D$ into two-dimensional space. Fig. 4 gives insight into benchmark dataset's data distribution and suggests that our contrastive learning model effectively separates P and N classes.

F. Baselines

To evaluate the performance of our proposed framework, we use the following baselines from current solar flare research where each method obtains a different data representation out of MVTS instances. When a downstream classifier is needed, we use logistic regression to be consistent with our experiments.

TABLE I: Binary Solar Flare Prediction Performance Results of CONTREX with Various Downstream Classifiers

Classifier	Accuracy	TSS	HSS2	F1	GS	ROC AUC
LR	**0.7306 ±0.09662**	**0.7098 ±0.09779**	**0.1189 ±0.04965**	**0.1579 ±0.04766**	0.02303 ±0.007004	**0.8549 ±0.04889**
SVM	0.6819 ±0.1344	0.6645 ±0.1362	0.1026 ±0.055	0.1428 ±0.051	0.02318 ±0.007065	0.8322 ±0.06812
KNC	0.7163 ±0.125	0.693 ±0.1248	0.1182 ±0.06089	0.1574 ±0.0564	0.02297 ±0.007	0.8465 ±0.06242
DT	0.7216 ±0.1354	0.6318 ±0.1176	0.1113 ±0.05374	0.1507 ±0.04774	0.02068 ±0.006054	0.8159 ±0.05879
MLP	0.6805 ± 0.08599	0.6671 ± 0.09098	0.09438 ±0.03769	0.1348 ±0.03946	**0.02322 ±0.006865**	0.8336 ±0.04549
FC	0.6824 ±0.09883	0.6662 ±0.1018	0.09238 ±0.02994	0.1331 ±0.02902	0.0232 ±0.007036	0.8331 ±0.0509

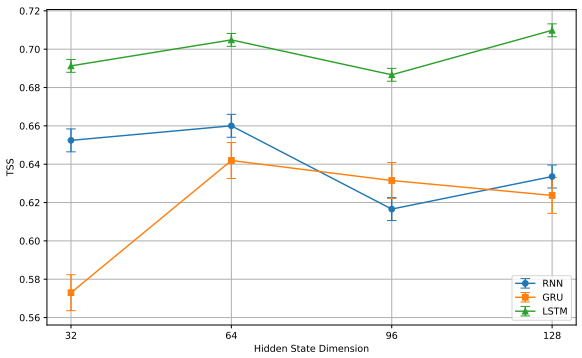

Fig. 3: Averaged TSS results of RNN, GRU, and LSTM within the CONTREX framework for binary solar flare prediction with varying hidden dimensions.

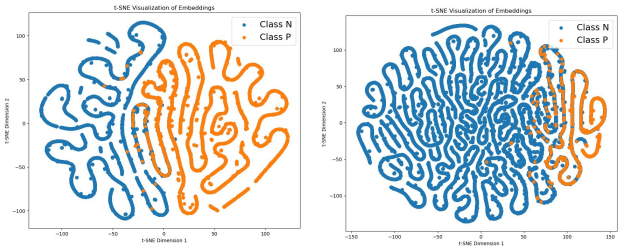

Fig. 4: t-SNE visualization of partition 3 and partition 4 train-test pair embeddings.

- **Vector MVTS (VMVTS)**: Flattens each MVTS instance to train a downstream classifier [13].
- **Vector of last timestamp (LTV)**: Extracts magnetic field parameters from only the last timestamp of the MVTS being temporally nearest to the flaring event to train a downstream classifier [10].
- **Long-short term memory (LSTM):** MVTS instances are fed into LSTM as a timestamp vector at each time step, and the last hidden representation is extracted as the representation [15]. We maintain the same hyperparameter settings: a hidden state dimension of 128, and a stochastic learning rate of 10^{-2}.
- **Random convolutional kernel transform (ROCKET):** Leverages the strengths of convolutional neural networks for time series classification by using randomly generated convolutional kernels, achieving state-of-the-art accuracy on 26 MVTS datasets of the UEA archive [23], [29].

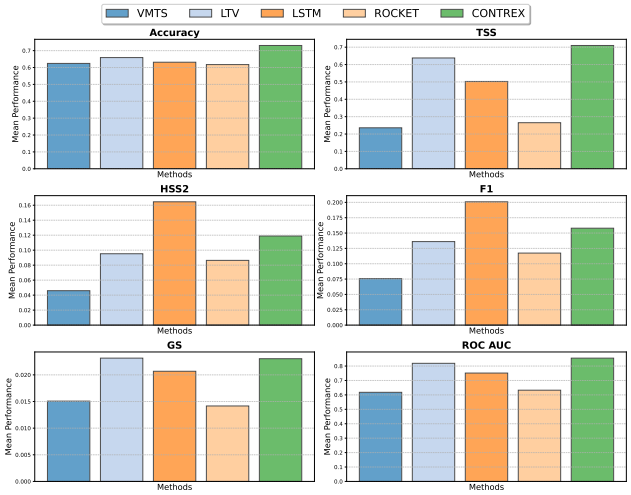

Fig. 5: Binary solar flare classification performance of baselines.

G. Binary Classification Performance

Following our experiments with three consecutive partitions as train-test pairs (i.e., p1-p2, p2-p3, and p3-p4), the averaged performance results are displayed in Table II and Fig. 5, demonstrating that the contrastive ability of CONTREX manages to show promising performance. CONTREX emerges as the top performer with 0.7306 accuracy, 0.7098 TSS and 0.8549 ROC AUC, outperforming LTV by 5.1%, 7.2% and 3.6%, respectively. However, with 0.1189 HSS2 and 0.1579 F1 score, CONTREX has the second position, lagging behind the leading performer, LSTM, by 4.6% and 4.3%, respectively. With 0.02303 GS, CONTREX shares the top position with LTV, with only a marginal 0.011% difference between the two. Overall, our experimental findings underscore CONTREX's competitiveness against state-of-the-art methods.

V. Conclusion

In this paper, a novel contrastive approach for time series is introduced and evaluated by the task of binary solar flare prediction of the SWAN-SF data instances. Our methodology comprised extracting dynamic attributes from each MVTS instance, computing contrastive extreme points from feature vectors, obtaining sequence representation embeddings for MVTS data instances guided by our custom reconstruction loss that leveraged the idea of generating embeddings that encapsulate the distinctive class characteristics, and training a downstream classifier with embeddings to binary classify solar

TABLE II: Binary Solar Flare Prediction Performance Results of Data Representation Methods

Model	Accuracy	TSS	HSS2	F1	GS	ROC AUC
VMTS	0.6243 ±0.4204	0.2358 ±0.3335	0.04592 ±0.06494	0.07575 ±0.06675	0.01507 ±0.01333	0.6179 ±0.1667
LTV	0.6589 ±0.1728	0.6378 ±0.1748	0.09523 ±0.05173	0.1361 ±0.04493	**0.02314 ±0.007139**	0.8189 ±0.08738
LSTM	0.6314 ±0.2914	0.5027 ±0.1967	**0.1646 ±0.1766**	**0.2009 ±0.1626**	0.02068 ±0.008318	0.7514 ±0.09834
ROCKET	0.6174 ±0.2688	0.2652 ±0.162	0.08639 ±0.07387	0.1174 ±0.06113	0.01416 ±0.01185	0.6326 ±0.08102
CONTREX	**0.7306 ±0.09662**	**0.7098 ±0.09779**	0.1189 ±0.04965	0.1579 ±0.04766	0.02303 ±0.007004	**0.8549 ±0.04889**

flares. In future studies, we aim to enhance the performance of our framework. For this reason, bringing other elements of triplet loss to our loss function as commonly utilized in modern contrastive learning methods, experimenting with different sampling and extreme point calculation methodologies, and applying feature reduction to the extreme points will be possible study directions. Furthermore, a study is planned to test the abilities of our framework in its current state with different MVTS benchmark datasets having binary and multi-class conditions. This experimentation will help to understand whether utilizing the contrastive extremes can yield better separation results for other time series settings.

Acknowledgment

This project has been supported in part by funding from CISE and GEO Directorates under NSF awards #2204363, #2240022, #2301397, and #2305781.

References

[1] A. Ahmadzadeh, B. Aydin, M. K. Georgoulis, D. J. Kempton, S. S. Mahajan, and R. A. Angryk, "How to train your flare prediction model: Revisiting robust sampling of rare events," *The Astrophysical Journal Supplement Series*, vol. 254, no. 2, p. 23, 2021.

[2] J. Eastwood, E. Biffis, M. Hapgood, L. Green, M. Bisi, R. Bentley, R. Wicks, L.-A. McKinnell, M. Gibbs, and C. Burnett, "The economic impact of space weather: Where do we stand?" *Risk analysis*, vol. 37, no. 2, pp. 206–218, 2017.

[3] R. A. Angryk, P. C. Martens, B. Aydin, D. Kempton, S. S. Mahajan, S. Basodi, A. Ahmadzadeh, X. Cai, S. Filali Boubrahimi, S. M. Hamdi *et al.*, "Multivariate time series dataset for space weather data analytics," *Scientific data*, vol. 7, no. 1, p. 227, 2020.

[4] P. H. Le-Khac, G. Healy, and A. F. Smeaton, "Contrastive representation learning: A framework and review," *Ieee Access*, vol. 8, pp. 193 907–193 934, 2020.

[5] P. S. McIntosh, "The classification of sunspot groups," *Solar Physics*, vol. 125, pp. 251–267, 1990.

[6] J. P. Mason and J. Hoeksema, "Testing automated solar flare forecasting with 13 years of michelson doppler imager magnetograms," *The Astrophysical Journal*, vol. 723, no. 1, p. 634, 2010.

[7] H. Song, C. Tan, J. Jing, H. Wang, V. Yurchyshyn, and V. Abramenko, "Statistical assessment of photospheric magnetic features in imminent solar flare predictions," *Solar Physics*, vol. 254, pp. 101–125, 2009.

[8] D. Yu, X. Huang, H. Wang, and Y. Cui, "Short-term solar flare prediction using a sequential supervised learning method," *Solar Physics*, vol. 255, pp. 91–105, 2009.

[9] O. W. Ahmed, R. Qahwaji, T. Colak, P. A. Higgins, P. T. Gallagher, and D. S. Bloomfield, "Solar flare prediction using advanced feature extraction, machine learning, and feature selection," *Solar Physics*, vol. 283, pp. 157–175, 2013.

[10] M. G. Bobra and S. Couvidat, "Solar flare prediction using sdo/hmi vector magnetic field data with a machine-learning algorithm," *The Astrophysical Journal*, vol. 798, no. 2, p. 135, 2015.

[11] A. Al-Ghraibah, L. Boucheron, and R. McAteer, "An automated classification approach to ranking photospheric proxies of magnetic energy build-up," *Astronomy & Astrophysics*, vol. 579, p. A64, 2015.

[12] R. Angryk, P. Martens, B. Aydin, D. Kempton, S. Mahajan, S. Basodi, A. Ahmadzadeh, X. Cai, S. Filali Boubrahimi, S. M. Hamdi, M. Schuh, and M. Georgoulis, "SWAN-SF," 2020. [Online]. Available: https://doi.org/10.7910/DVN/EBCFKM

[13] S. M. Hamdi, D. Kempton, R. Ma, S. F. Boubrahimi, and R. A. Angryk, "A time series classification-based approach for solar flare prediction," in *2017 IEEE Int. Conf. on Big Data (Big Data), Boston, MA, USA*. IEEE, 2017, pp. 2543–2551.

[14] R. Ma, S. F. Boubrahimi, S. M. Hamdi, and R. A. Angryk, "Solar flare prediction using multivariate time series decision trees," in *2017 IEEE Int. Conf. on Big Data (Big Data), Boston, MA, USA*. IEEE, 2017, pp. 2569–2578.

[15] A. A. M. Muzaheed, S. M. Hamdi, and S. F. Boubrahimi, "Sequence model-based end-to-end solar flare classification from multivariate time series data," in *2021 20th IEEE Int. Conf. on Machine Learning and Applications (ICMLA), Pasadena, CA, USA*. IEEE, 2021, pp. 435–440.

[16] S. M. Hamdi, A. F. Ahmad, and S. Filali Boubrahimi, "Multivariate time series-based solar flare prediction by functional network embedding and sequence modeling," in *CIKM workshop for Applied Machine Learning Methods for Time Series Forecasting (AMLTS), Atlanta, GA, USA*, 2022.

[17] H. Fan, F. Zhang, and Y. Gao, "Self-supervised time series representation learning by inter-intra relational reasoning," *arXiv preprint arXiv:2011.13548*, 2020.

[18] E. Eldele, M. Ragab, Z. Chen, M. Wu, C. K. Kwoh, X. Li, and C. Guan, "Time-series representation learning via temporal and contextual contrasting," *arXiv preprint arXiv:2106.14112*, 2021.

[19] D. Luo, W. Cheng, Y. Wang, D. Xu, J. Ni, W. Yu, X. Zhang, Y. Liu, Y. Chen, H. Chen *et al.*, "Time series contrastive learning with information-aware augmentations," in *Proceedings of the AAAI Conference on Artificial Intelligence*, 2023, pp. 4534–4542.

[20] S. Lee, T. Park, and K. Lee, "Soft contrastive learning for time series," *arXiv preprint arXiv:2312.16424*, 2023.

[21] X. Zheng, X. Chen, M. Schürch, A. Mollaysa, A. Allam, and M. Krauthammer, "Simts: rethinking contrastive representation learning for time series forecasting," *arXiv preprint arXiv:2303.18205*, 2023.

[22] C. H. Lubba, S. S. Sethi, P. Knaute, S. R. Schultz, B. D. Fulcher, and N. S. Jones, "catch22: Canonical time-series characteristics: Selected through highly comparative time-series analysis," *Data Mining and Knowledge Discovery*, vol. 33, no. 6, pp. 1821–1852, 2019.

[23] A. P. Ruiz, M. Flynn, J. Large, M. Middlehurst, and A. Bagnall, "The great multivariate time series classification bake off: a review and experimental evaluation of recent algorithmic advances," *Data Mining and Knowledge Discovery*, vol. 35, no. 2, pp. 401–449, 2021.

[24] M. EskandariNasab, S. M. Hamdi, and S. Filali Boubrahimi, "Swan-sf data preprocessing and sampling notebooks," Jun. 2024. [Online]. Available: https://doi.org/10.5281/zenodo.11564789

[25] M. EskandariNasab, S. Hamdi, and S. Boubrahimi, "Cleaned swansf dataset," Jun. 2024. [Online]. Available: https://doi.org/10.5281/zenodo.11566472

[26] G. Barnes and K. Leka, "Evaluating the performance of solar flare forecasting methods," *The Astrophysical Journal*, vol. 688, no. 2, p. L107, 2008.

[27] D. M. P. Murti, U. Pujianto, A. P. Wibawa, and M. I. Akbar, "K-nearest neighbor (k-nn) based missing data imputation," in *2019 5th Int. Conf. on Science in Information Technology (ICSITech), Yogyakarta, Indonesia*. IEEE, 2019, pp. 83–88.

[28] L. van der Maaten and G. Hinton, "Visualizing data using t-sne," *Journal of Machine Learning Research*, vol. 9, no. 86, pp. 2579–2605, 2008.

[29] A. Dempster, F. Petitjean, and G. I. Webb, "Rocket: exceptionally fast and accurate time series classification using random convolutional kernels," *Data Mining and Knowledge Discovery*, vol. 34, no. 5, pp. 1454–1495, 2020.