

On The Effectiveness of Imaging of Time Series for Flare Forecasting Problem

Yang Chen^{*†}, Anli Ji[†], Pavan Ajit Babajiyavar^{*†}
Azim Ahmadzadeh^{*‡}, and Rafal A. Angryk^{*‡}

^{*}Department of Compute Science, Georgia State University, Atlanta, GA, USA

Email: [†]{ychen113,aji1,pbabajiyavar1}@student.gsu.edu,

[‡]{aahmadzadeh1,angryk}@cs.gsu.edu,

Abstract—Forecasting the occurrence of solar flares is a typical 21st century rare-event classification task. Over the past two decades, many studies have implemented various techniques and approaches for classification of strong and weak solar flares. The release of the recent flare forecasting benchmark dataset, named SWAN-SF, has opened the door for taking advantage of multivariate time series (MVTs) of pre-flare magnetic fields' activity in order to potentially achieve higher performance and increase the robustness of the new forecasting models. In this study, we take a new approach and explore the effectiveness of imaging algorithms on the time series. We convert MVTs data into multi-channel image data using Gramian Angular Fields (GAF) and Markov Transition Fields (MTF) to explore the proven strength of deep neural networks in the Image Processing domain, on flares' MVTs data. Inspired by GAF, we propose another imaging matrix that our experiments show that it significantly improves the performance of the CNN, by 120% in terms of TSS and 440% in terms of HSS2. Finally, we juxtapose two approaches to tackle the flare forecasting problem: one, to utilize a time-series specific Support Vector Classifier for classification of flares, and the other, to train a Convolutional Neural Network (CNN) on the derived images using GAF, MTF, and our modified GAF.

Index Terms—multivariate, time series, forecasting, imaging, flare

I. INTRODUCTION

Solar flares are sudden bursts of electromagnetic radiations from the Sun when the magnetic fields generated by constantly moving electrically charged gases on the surface tangle, stretch and twist. Since 1974, X-ray flares have been automatically detected and classified by the National Oceanic and Atmospheric Administration's (NOAA) constellation of GOES satellites in the 1-8 Angstrom wavelength range. Based on the peak soft X-ray flux in this range, flares are logarithmically classified into the five classes of A, B, C, M, and X, from weaker to stronger. Strong solar flares occasionally releases large expulsions of plasma and magnetic field, called coronal mass ejections (CME), which can reach the Earth in as little as 15-18 hours. Strong CMEs can potentially cause geomagnetic storm with direct impact on the well-being of astronauts during spacewalk missions, the GPS system, and consequently the GPS-based positioning industries, and the power grid system, and consequently the electronic infrastructures [1].

(Y. Chen, A. Ji, and P. A. Babajiyavar are co-first authors.)(Corresponding author: Yang Chen)

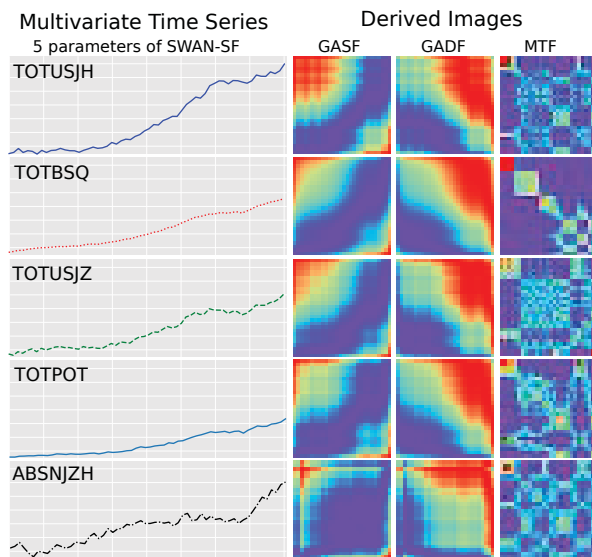


Fig. 1. An example of a multivariate time series from SWAN-SF and the derived images using three different imaging algorithms.

The damage was estimated to be \$1-\$2 trillion (for the United States) during the first year alone [2]. It was estimated in 2012 that the probability of such an extreme event to occur within the next decade is $\sim 12\%$ [3]. The severity of such events and the fact that simple remedies (e.g., rescheduling a spacewalk, or injecting reserve power in the power grid system) can significantly reduce the damage, has made the solar flare forecasting an outstanding interdisciplinary research topic.

II. BACKGROUND

Researchers over the past two decades have implemented various techniques and approaches for classification of strong and weak solar flares. From a sample of such studies, [4] used the classical approach of logistic regression to build a classifier that distinguishes between the X-, M-, and C-class flares using only 3 predictive parameters. Using the Poisson statistics technique, [5] calculated the probabilities for the same three flare magnitudes. Taking into account the conditional independence among the magnetic field parameters of active regions, [6] built their short-term prediction model

using a Bayesian network. Support Vector Machines have also been used as a more advanced classifier in [7] on 25 different predictive parameters, and in [8] combined with k-nearest neighbors algorithm, using only 2 predictive parameters. More advanced approaches have also been carried out. For instance, [9] developed a deep neural network, named Deep Flare Net, to learn the patterns from the 79 extracted (magnetic-field) parameters. In contrast, [10] employed the convolutional neural networks to automatically learn the flaring patterns from the magnetograms, instead of manually calculating them.

While there seems to be an array of interesting avenues proposed for tackling the flare forecasting challenge, unfortunately, comparing the reported forecasting performances of those models is not feasible. The reasons as identified in [11] can be summarized as follows: (1) employing different sampling strategies, (2) the evaluation of models are carried out on the datasets which are not publicly available nor their exact creation and integration methodologies are well documented, (3) the minor or major differences in the problem formulation (e.g., probabilistic versus categorical classification, or multi-class versus binary task), and (4) the use of different preprocessing strategies on the data. Each of these decisions changes the difficulty of the task in hand, which in turn renders the numerical comparison of the reported scores meaningless. The benchmark dataset we employ in this study (see Section III) is intended to remedy this very problem. In the absence of a shared experimental setting and comparable results, based on our rigorous analysis of many such studies we conclude that a reliable method that takes on the binary flare-forecasting problem with a single model (i.e., no ensemble) achieves the true skill statistic (TSS) of ≈ 0.6 while keeping a similar Heidke skill score ($HSS2$) (e.g., in [7], [12]). That said, the reader should note that even that does not set a fair bar for our experiments to be compared against. Because (1) those models have not been evaluated on the same dataset as ours (SWAN-SF), and (2) their classifier, i.e., SVM, targeted single-point data and not time series data as ours do.

Our study approaches the problem from a new angle in which the (multivariate) time series of flaring and non-flaring active regions are transformed to (multi-channel) images. This transformation in fact, takes our problem to the Image Processing domain where we can benefit from the widely accepted superiority of deep neural networks [13] to the classical machine learning algorithms in many applications. While one could expect data loss when raw data in the form of real values are transformed to unsigned 8 bit integers, experiments in this study were found to not have had a significant effect from this transformation. The two main imaging techniques we put to the test are the Gramian Angular Fields (GAF) and Markov Transition Fields (MTF) transformations [14].

The intention behind converting the multivariate time series of SWAN-SF data is to experiment whether a state-of-the-art CNN can discover patterns in the data that classical machine learning techniques may have missed. A cursory examination of the dataset would reveal, even to a casual observer, that the data is imbalanced and largely belong to the weak classes

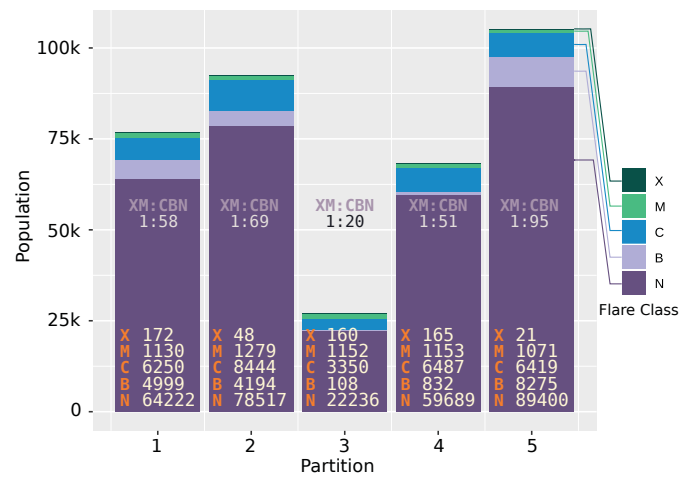


Fig. 2. Counts of the five flare classes across the five partitions. (The numbers correspond to v0.7 of SWAN-SF data, and may vary in the most recent version.)

of flares (C, B, and Q) and this research is set to explore whether a CNN can better identify those fewer flaring classes with greater flux in observations that could potentially trigger coronal mass ejections. [15] is one such research, although with a different objective, in which the time series data were subjected to GAF transformation and the authors showed that deep neural networks have surpassed the performance of the classical machine learning models in identifying the minority classes. Similar results were obtained by [16] and [17] wherein the time series data were transformed into GAF images before subjecting it to a CNN model for classification. [18] claims an accuracy of 90.7% outperforming Long Short-Term Memory (LSTM) model when the time series data of currency exchange rate was remodeled as GAF images and fed to a neural network based on LeNet-5. Power load consumption time series data of individual customers that exhibits fluctuations quite frequently, when transformed into Recurrence Plots [19] and further subjected to a CNN have shown promising results in [20]. Yet another successful experimentation on imaging data for classification has been demonstrated by [21] which has implemented a Recurrent Neural Network (RNN) integrated with MTF proving better results than the baseline classical machine learning model. Results observed in aforementioned works and the fact that this technique has not been tried before on solar data are the encouraging factors for the authors to adopt a non-traditional approach to this problem.

III. DATASET

A. SWAN-SF Benchmark Dataset

Solar flares are understood to be dynamical phenomenon which have clear pre-flare and post-flare phases [22]. However, many studies utilize point-in-time measurements, e.g., [7], [12], [23], to predict flares' behavior in the future. To expand the horizon of flare forecasting, and for the reasons discussed in Section II, a recent study generated a large collection of multivariate time series data of flares' magnetic fields

extracted from solar photospheric vector magnetograms in Space weather HMI Active Region Patch (SHARP) series [24]. The benchmark dataset, named Space Weather ANalytics for Solar Flares (SWAN-SF), is meant to serve as a testbed for flare forecasting models [25]. It is hoped that using this dataset some significant improvements in the performance and robustness of the forecasting models be achieved. This benchmark dataset is made openly available on Harvard Dataverse repository [26].

As illustrated in Fig. 2, SWAN-SF dataset is made up of five temporally non-overlapping partitions covering the period from May 2010 through August 2018. Each partition contains approximately an equal number of X- and M-class flares. The data points are slices of physical (magnetic field) parameters of multivariate time series, in a sliding fashion. That is, for a particular flare with a unique id, k equal-length multivariate time series are collected from a fixed period of time (24 hrs) throughout the history of that flare. SWAN-SF originally contains a collection of 82 physical parameters derived from the vector magnetic field data. The feature ranking process carried out by [7] ranked the top 5 features (physical parameters) to be ‘TOTUSJH’, ‘TOTBSQ’, ‘TOTPOT’, ‘TOTUSJZ’, ‘ABSNJZH’ (for definition of the features see Table 1 in [25]). As this study focuses on the effectiveness of three imaging algorithms, we limit our experiments to only these five features.

B. Preprocessing

From the 5 partitions of the SWAN-SF dataset, the first is utilized for training the classifier and the second is used for evaluation of the models. Missing data are imputed using linear interpolation and to further ensure that no feature intrinsically influences the model, the dataset is re-scaled to the range $[-1, 1]$. As evident from Fig. 2, SWAN-SF is highly imbalanced and dominated by instances corresponding to the majority classes, i.e., B, C, and Q instances. The primary motivation in flare forecasting is to predict flares of stronger classes which are just a handful when compared. Such a significant imbalance in the dataset would affect any classifier by injecting a bias towards the majority classes. This class imbalance in SWAN-SF has been explored and analyzed before in [11]. Inspired by their observations, the technique of undersampling while preserving the climatology of the original dataset is applied on the training dataset. In this undersampling method, the number of instances of the majority classes and the minority classes (X and M) become equal while the ratio of the individual classes in turn are preserved from the original dataset to maintain the climatology of flares. We utilized the Python package named MVTS-Data Toolkit [27] in order to facilitate the preprocessing of the data.

C. Derived Image Dataset

From the multivariate time series instances of SWAN-SF, we create several image datasets, using six different imaging transformations (with abbreviations GASF, GADF, GASF_r, GADF_r, MTF, and RP). In the following, we briefly explain

these algorithms and how the image datasets are derived from the time series data.

Gramian Angular Fields (GAF) and Markov Transition Fields (MTF) are imaging algorithms that encode a univariate time series into an image with preserving temporal dependencies between observations [14]. To briefly review how GAF works, let $X = \{x_1, x_2, \dots, x_n\}$ be a given time series of length n . The first step in GAF algorithm is to normalize X by transforming its values into the interval $[-1, 1]$. This guarantees that the polar coordinate system can represent a time series in the Cartesian coordinate system with angular bounds $[0, \pi]$. Let $\tilde{X} = \{\tilde{x}_1, \tilde{x}_2, \dots, \tilde{x}_n\}$ denote the normalized time series. In the next step, $\tilde{X}$ is transformed into the polar coordinate system using Eq. 1. Encoding a time series into its polar coordinates has two primary advantages. First, the entire encoding process is bijective as $\cos(\phi)$ is monotonic when $\phi \in [0, \pi]$. Second, the polar coordinate system preserves temporal dependency by using the radial coordinate r .

$$\begin{cases} \phi_i = \arccos(\tilde{x}_i), & -1 \leq \tilde{x}_i \leq 1, \tilde{x}_i \in \tilde{X} \\ r_i = \frac{t_i}{n}, & t_i = 1, 2, \dots, n \end{cases} \quad (1)$$

There are two different Gramian Angular Matrices, as defined in Eq. 2: the Gramian Angular Summation Field (GASF) and the Gramian Angular Difference Field (GADF). These matrices preserve absolute temporal relationships between the observations since time increases with the direction from upper-left to bottom-right. The main diagonal contains the original value/angular information.

$$\begin{aligned} GASF &= \begin{bmatrix} \cos(\phi_1 + \phi_1) & \cos(\phi_1 + \phi_2) & \dots & \cos(\phi_1 + \phi_n) \\ \cos(\phi_2 + \phi_1) & \cos(\phi_2 + \phi_2) & \dots & \cos(\phi_2 + \phi_n) \\ \dots & \dots & \dots & \dots \\ \cos(\phi_n + \phi_1) & \cos(\phi_n + \phi_2) & \dots & \cos(\phi_n + \phi_n) \end{bmatrix} \\ GADF &= \begin{bmatrix} \sin(\phi_1 - \phi_1) & \sin(\phi_1 - \phi_2) & \dots & \sin(\phi_1 - \phi_n) \\ \sin(\phi_2 - \phi_1) & \sin(\phi_2 - \phi_2) & \dots & \sin(\phi_2 - \phi_n) \\ \dots & \dots & \dots & \dots \\ \sin(\phi_n - \phi_1) & \sin(\phi_n - \phi_2) & \dots & \sin(\phi_n - \phi_n) \end{bmatrix} \end{aligned} \quad (2)$$

To help the reader better picture these transformations, we chose three MVTS instances of X, C, and Q flare classes, and illustrated their corresponding derived GASF and GADF images for the top-five features. These time series and the images are depicted in Fig. 3.

Inspired by GAF’s definition, and noting that neither of GASF or GADF utilize the radial coordinate and preserve the temporal information solely by encoding the angular coordinate, we propose two other matrices. Based on the assumption that the time series observations closer to one another have stronger correlations than those temporally farther apart, we derive the new matrices by multiplying (element-wise) GASF/GADF matrices by $W \in \mathbb{R}^2$, the matrix of weights, as shown in Eq. 3. In this equation, i and j are indices of time series observations, and n is the length of the time series. This weighting approach is consistent with our above-mentioned assumption as it gives higher weights to closer observations.

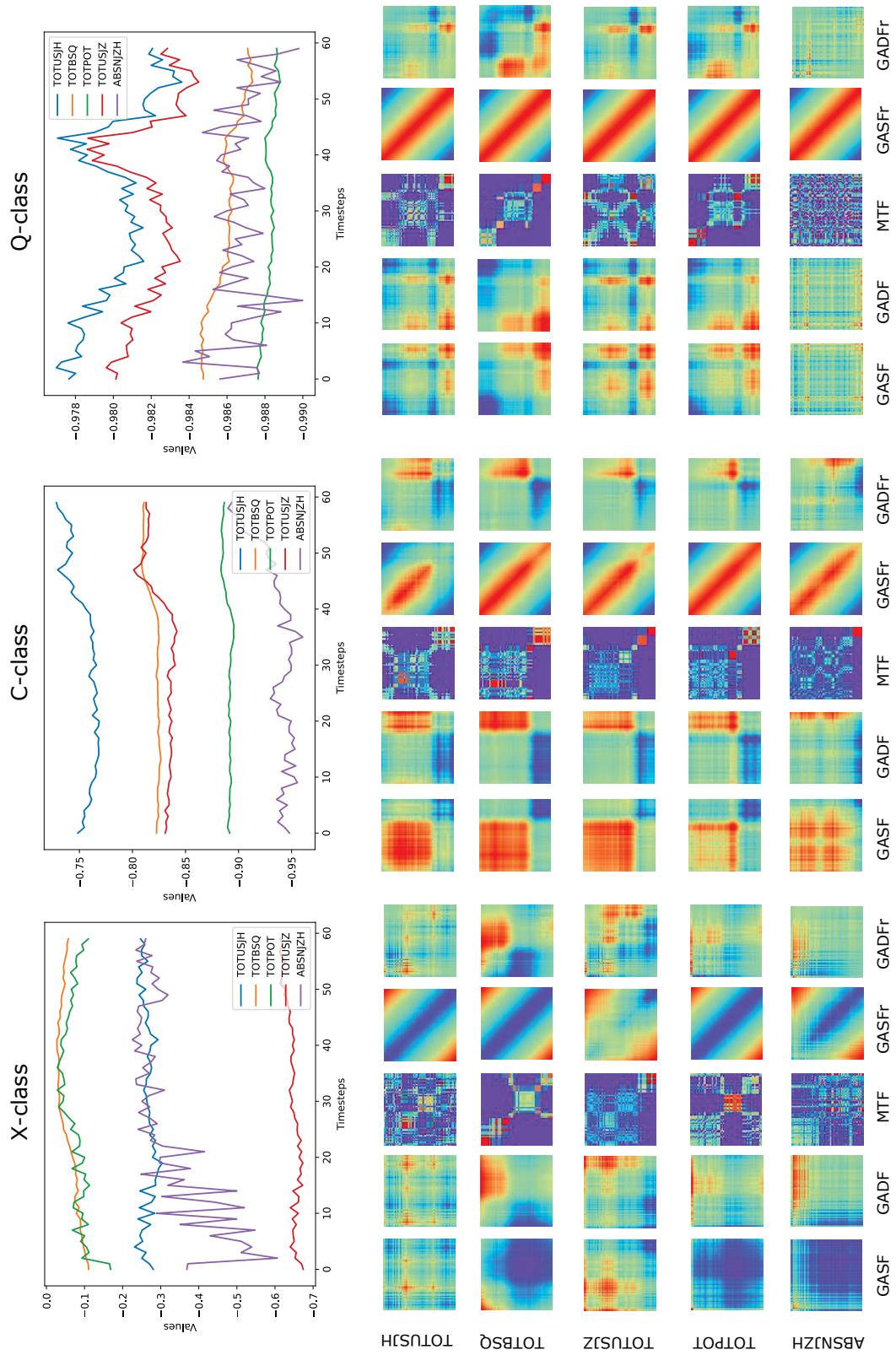


Fig. 3. Three examples of multivariate time series with top five parameters and derived images using five imaging algorithms, including GASF, GADF, MTF, GASFr and GADFr. The three multivariate time series instances are flares of X-class, C-class and Q-class.

We denote the modified matrices as $GASF_r$ and $GADF_r$ where the subscript r highlights the additional use of the radial coordinate. Images generated by $GASF_r$ and $GADF_r$ with three instances are illustrated in Fig. 3.

$$W = [(1 - |\frac{i-j}{n}|)], 1 \leq i, j \leq n \quad (3)$$

The second imaging algorithm, MTF, as formulated by Eq. 4, is similar to the Stochastic Matrix [28]. This is used to capture the transitions of a Markov Chain describing a sequence of possible states that a time series goes through. In such transitions, the probability of each state solely depends on its previous state.

$$M = \begin{bmatrix} w_{i,j}|_{x_1 \in q_i, x_1 \in q_j} & \cdots & w_{i,j}|_{x_1 \in q_i, x_n \in q_j} \\ w_{i,j}|_{x_2 \in q_i, x_1 \in q_j} & \cdots & w_{i,j}|_{x_2 \in q_i, x_n \in q_j} \\ \vdots & \ddots & \vdots \\ w_{i,j}|_{x_n \in q_i, x_1 \in q_j} & \cdots & w_{i,j}|_{x_n \in q_i, x_n \in q_j} \end{bmatrix} \quad (4)$$

Given a time series X , and q as the number of bins, M is a $q \times q$ matrix computed by discretizing the time series data into q quantile bins. Along the temporal axis, for every i and j time stamps, the corresponding data are presented in q_i and q_j bins. In this matrix, each element $M_{i,j}$ denotes the probability of transition $q_i \rightarrow q_j$. As MTF captures the probability of time series transitioning between different states during its life cycle, the periodicity of the signal emerges as patterns in an MTF matrix. To show an example, we illustrated the five features of three different MVTs samples of X, C, and Q flare classes. The MTF matrices are illustrated in Fig. 3. In the case of TOTBSQ of the chosen M-class flare (red dotted time series in the second row), the feature appears to be monotonically decreasing over time which has given rise to the patterns of squares along the diagonal in the corresponding MTF plot. If the time series were to be periodic, the pattern would have been found to be repeated across the MTF image. On the other hand, ABSNJZH appears to be noisy varying between a small range over time, which is captured in the corresponding MTF image by smaller square shapes.

Another popular imaging algorithm is the Recurrence Plots which captures the recurrences of observations in the time series [19]. But as it is not the nature of the time series of SWAN-SF to have a clear periodic pattern, it may not be effective for our classification purpose. Therefore, in this study we do not consider this algorithm in our experiments.

All the above-mentioned imaging algorithms transform univariate time series into single-channel images. In SWAN-SF however, the data points are multivariate time series. For a multivariate time series with k parameters, we first derive the k images, one per feature, and then convert them into one k -channel image. The width and height of the generated images is equal to the fixed length of the time series, i.e., 60. Since we only work with the top 5 features, each MVTs of SWAN-SF turn into a 3D tensor of shape (5, 60, 60).

For all these transformations we utilize the Pyts Python package [29].

IV. METHODOLOGY

To properly evaluate the effectiveness of using imaging algorithms on the time series for the purpose of flare forecasting we draw the reader's attention to a few points: (1) A set of baseline experiments are needed in which the classification task should be carried out on the time series data. The reported forecasting performances should be then compared with that of CNN on the derived images. (2) Since CNN automatically extracts the features (from images), in order to have a fair comparison, the baseline classifier may not depend on any (manual) feature extraction step. That is, it should be able to take in the multivariate time series in their raw form, and not their potentially important features, such as their statistical moments. Otherwise, the strength or weakness of the baseline model could always be attributed to how well those features were chosen or optimized, and therefore makes the comparison depend on more variables than just the data format (time series versus images). Lastly, (3) the baseline model's algorithm should be somewhat comparable with that of CNN in terms of its learning process.

To address these three concerns we choose our baseline classifier to be the time-series specific support vector classifier (T-SVC) [30]. T-SVC is an implementation of SVM that utilizes the dynamic time warping (DTW) distance measure as the positive definite kernels for time series, and therefore, it does not depend on the extracted features in the tabular form. Regarding the third concern, CNN algorithms, and deep neural networks in general, with the help of their several transformations separate the feature space using multiple hyperplanes. The SVM family of classifiers, when equipped with the kernel trick, works similar to a 2-layer neural network. And a 1-layer neural network with the Hinge loss is simply a linear SVM. Therefore, T-SVC and CNN have comparable learning processes.

A. Models

Support Vector Machine (SVM) in general, and T-SVC in particular, fall into the general category of kernel methods. The biggest advantage of using SVM is that we can generate non-linear decision boundaries by using methods that are designed for linear classifiers. As for the SVM model, there exists a basic set of hyperparameters: the soft margin constant, C , and the parameters on which the kernel function depends (including width of a Gaussian kernel and the degree of a polynomial kernel). The first hyperparameter, C , acts as a regularization parameter in our SVM classifier. It controls the influence of each individual support vector and allows some instances to be misclassified to improve the generalization of the model and consequently the robustness. Larger values of C encourage smaller margins if the decision function is better at correctly classifying all the training samples. Another important hyperparameter in the tuning process is γ which is the free parameter of the Gaussian Radial Basis Function (RBF) kernel. The γ parameter defines how far the influence of a single training example reaches. Larger values of γ have

smaller variances which imply that the support vectors do not have wide-spread influence.

Convolutional Neural Network (CNN) is a class of deep neural networks that have become dominant in various computer vision tasks. The key characteristic of CNN is that it automatically learns the spatial features from images by stacking an array of layers, such as the convolutional layers, ReLu layers, pooling layers, and fully connected layers. In this study, we use LeNet-5 [31] which is a CNN-based architecture that comprises of seven layers including three convolutional layers, two pooling layers, two fully connected layers, and a Rectified Linear Unit (ReLU) as the activation function. The cross entropy loss function is utilized for calculating the difference between the estimated and the actual labels of the training data. The computed loss accompanied by a back-propagation algorithm facilitates the weight adjustments within the network. Supporting factors, such as learning rate and the choice of an optimizer, help boost the performance of the neural network.

B. Metrics

Our approach toward the flare forecasting problem is in fact a rare-event classification task. A proper evaluation of models' performance in the presence of scarce (imbalanced) data requires special treatment [32]. Choosing the right evaluation metric is certainly one of them. From years of exploration, domain experts have come to agree on the effectiveness of two metrics, namely the *true skill statistic* (TSS) [33] and the updated *Heidke skill score* ($HSS2$) [34], [35]. As shown in Eq. 5, TSS is simply the difference between the probability of detection, i.e., true-positive rate, and the probability of false alarm, i.e., false-positive rate. There, $P = tp + fn$ and $N = fp + tn$ are the numbers of the positive and negative instances, respectively. This measure ranges from -1 to $+1$, where -1 indicates that the model has misclassified all the instances, 0 means that the model has no forecasting skill, and $+1$ represents a perfect model that correctly assigns labels to all instances.

$$TSS = \frac{tp}{P} - \frac{fp}{N} \quad (5)$$

$HSS2$, as formulated in Eq. 6, quantifies the model's performance by comparing it to a model that classifies instances randomly. Similar to TSS , $HSS2$ ranges within the interval $[-1, 1]$, with 0 indicating that there is no difference between the model's performance and a random guess. The positive values indicate how better than random the model performs. Decreasing negative values reflect a higher similarity to a model that misclassifies all instances.

$$HSS2 = \frac{2 \cdot ((tp \cdot tn) - (fn \cdot fp))}{P \cdot (fn + tn) + N \cdot (tp + fp)} \quad (6)$$

For the evaluation of our models, since we preserve the balance ratio of the data in all experiments, we report both TSS and $HSS2$.

V. EXPERIMENTS AND RESULTS

We now put the main thesis of this study to the test. That is, to evaluate the effectiveness of imaging algorithms for classification of multivariate time series. We conduct a series of experiments and compare the results, as explained below.

We first train T-SVC on partition 1 of SWAN-SF and LeNet-5 on the images derived from the same partition using each of the imaging algorithms GASF, GADF, and MTF. To remedy the class-imbalance issue, as discussed in Section III, we undersample the training dataset. Then we test each model on partition 2 of the dataset and compare their performances in terms of TSS and $HSS2$. Note that although the class frequencies in the training data is made balanced, we leave the imbalance ratio in partition 2 untouched so that the results are comparable with an operational forecasting model. We repeat this comparison on each of the top 5 features individually followed by an experiment involving all those features.

Another group of experiments are conducted to evaluate the effectiveness of our proposed imaging matrices, $GASF_r$ and $GADF_r$. Similar to the first group of experiments, we conduct this for each of the top 5 features and for a combination of them as well.

In each scenario, we tune the hyperparameters of the classifiers using the grid-search approach and find the optimal settings based on the highest and the most consistent TSS and TSS values of the estimated labels. As a result, T-SVC model is built with its two hyperparameters C and γ set to 1000 and 0.0001, respectively, equipped with 'sigmoid' kernel. Similarly for LeNet-5, we conclude the learning rate to be initialized at $10e-4$, and use Stochastic Gradient Descent as the optimizer. We halt the training process at 500 epochs.

From the results shown in Fig. 4, performance of T-SVC clearly stands out with TSS at close to 0.8, compared to LeNet-5 with GASF, GADF, or MTF. Looking at the reported $HSS2$ of the same experiments in Fig. 5, although T-SVC barely reaches 0.3, it still outperforms LeNet-5 in all 6 cases.

Fig. 4 and Fig. 5 also show that LeNet-5 performs best on images derived by the MTF imaging algorithm. One reason for this superiority is that MTF is a measure of probability of transition between states of a time series. Therefore, it may capture the patterns and fluctuations more distinctly than GAF. This can be visually spotted in Fig. 1 as the MTF images are more distinct than those generated using GAF.

Finally, we compare the forecasting performance of LeNet-5 on GAF- and GAF_r -derived images. Fig. 6 shows that incorporating the radial coordinate in GASF made a significant improvement. In fact, $GASF_r$ resulted in an over 120% (on average TSS) improvement in LeNet-5's classification performance (an increase from 0.21 to 0.48). An even more dramatic performance boost is measured by $HSS2$, shown in Fig. 7, with an over 440% increase (from 0.07 to 0.38). $GADF_r$ has also made some improvements, as shown in the same plots, but not consistently and not nearly as significantly.

Comparing Figs. 4 and 5 with 6 and 7 also reveals some interesting points. While in terms of TSS , $GASF_r$ does not

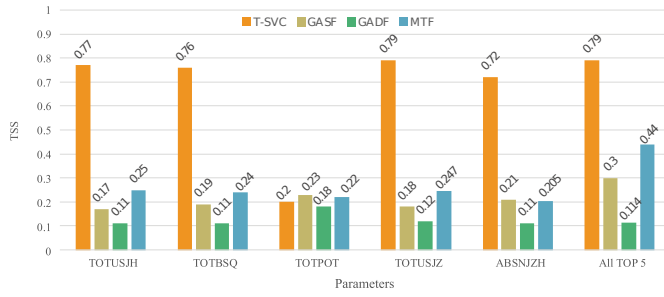


Fig. 4. TSS comparison of T-SVC and LeNet-5 on images generated by GASf, GADF and MTF.

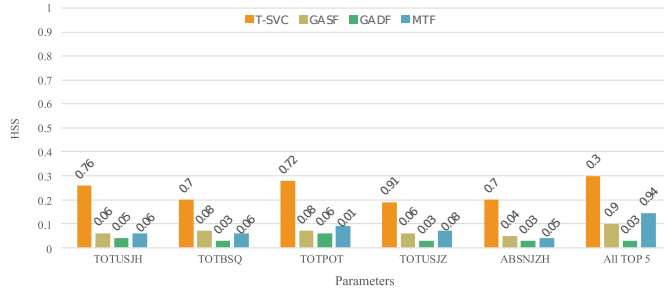


Fig. 5. HSS comparison of T-SVC and LeNet-5 on images generated by GASf, GADF and MTF.

seem to have strong enough discriminating features for LeNet-5 that the raw time series provide for T-SVC, it comes on top when HSS_2 is the measure of comparison. T-SVC performs on average 0.38% better than LeNet-5, in terms of TSS , when $GASf_r$ is used, but LeNet-5 outperforms that by $\approx 60\%$ boost on average, in terms of HSS_2 .

All in all, the results indicate that the imaging algorithm can capture the important characteristics of time series data while allowing a powerful family of classifiers, i.e., CNNs, to be tried and evaluated for the outstanding task of flare forecasting.

VI. CONCLUSION AND FUTURE WORK

In order to approach the flare forecasting problem from a new angle, we converted the multivariate time series of SWAN-SF dataset into multi-channel images using several different imaging algorithms. To evaluate the effectiveness of such transformations of data on our forecasting problem, we employed T-SVC classifier for classification of time series and LeNet-5 for classification of the derived images. The image classifier was intentionally chosen to be rather simple so that it remains comparable with the SVM-based classifier we employed. The analyses of the results showed no improvements in classification of flaring and non-flaring instances when derived images were used.

Looking into the lossy nature of these imaging algorithms, we identified the main reason as to why the derived image data were not nearly as promising as the multivariate time series data were. As seen in III-C, since radial information is not encoded into the imaging algorithms, MVTs instances cannot be reconstructed from the image without data loss.

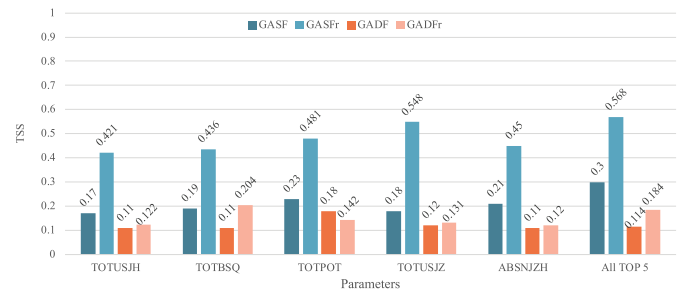


Fig. 6. TSS comparison of LeNet-5 on images generated by GASf and GASf_r.

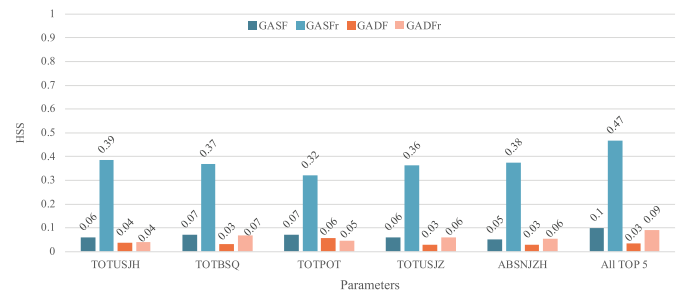


Fig. 7. HSS comparison of LeNet-5 on images generated by GADF and GADF_r.

Based on our findings, we modified one of those algorithms and achieved a significant improvement in the classification of flares, compared to the existing imaging algorithms. Although the performance is not comparable to what T-SVC achieved on the time series data, we consider the significant improvement seen as a proof of concept that we would like to further explore in our future work. Our weighting mechanism that resulted in the improvement was mathematically simple with no interpretable or data-driven justification. This implies that a more appropriate weighting function may result in an imaging algorithm more effective at capturing the time series features. Exploiting the advantages of complex CNNs is one of the avenues we would like to explore. In that case, we must replace T-SVC with a deep neural network that takes in multivariate time series so that the model still remain comparable in complexity, and the superiority of one model over the other does not influence the results.

We hope that our new approach opens new doors to more effective ways of tackling the outstanding challenge of flare forecasting.

ACKNOWLEDGMENT

This project has been supported in part by funding from the Division of Advanced Cyberinfrastructure within the Directorate for Computer and Information Science and Engineering, the Division of Astronomical Sciences within the Directorate for Mathematical and Physical Sciences, and the Division of Atmospheric and Geospace Sciences within the Directorate for Geosciences, under NSF award #1931555. It was also sup-

ported in part by funding from the Heliophysics Living With a Star Science Program, under NASA award #NNX15AF39G.

REFERENCES

- [1] 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. [Online]. Available: <https://doi.org/10.1111/risa.12765>
- [2] N. R. Council, *Severe Space Weather Events—Understanding Societal and Economic Impacts: A Workshop Report*. Washington, DC: The National Academies Press, 2008. [Online]. Available: <https://doi.org/10.17226/12507>
- [3] P. Riley, "On the probability of occurrence of extreme space weather events," *Space Weather*, vol. 10, no. 2, 2012. [Online]. Available: <https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2011SW000734>
- [4] 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, no. 1, pp. 101–125, 2009.
- [5] D. S. Bloomfield *et al.*, "Toward reliable benchmarking of solar flare forecasting methods," *The Astrophysical Journal*, vol. 747, no. 2, p. L41, Feb 2012. [Online]. Available: <https://iopscience.iop.org/article/10.1088/2041-8205/747/2/L41>
- [6] D. Yu, X. Huang, H. Wang, Y. Cui, Q. Hu, and R. Zhou, "Short-term solar flare level prediction using a bayesian network approach," *The Astrophysical Journal*, vol. 710, no. 1, p. 869, 2010.
- [7] 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. [Online]. Available: doi.org/10.1088/0004-637X/798/2/135
- [8] R. Li, H.-N. Wang, H. He, Y.-M. Cui, and Z.-L. Du, "Support vector machine combined with k-nearest neighbors for solar flare forecasting," *Chinese Journal of Astronomy and Astrophysics*, vol. 7, no. 3, p. 441, 2007.
- [9] N. Nishizuka, K. Sugiura, Y. Kubo, M. Den, and M. Ishii, "Deep flare net (defn) model for solar flare prediction," *The Astrophysical Journal*, vol. 858, no. 2, p. 113, 2018.
- [10] X. Huang, H. Wang, L. Xu, J. Liu, R. Li, and X. Dai, "Deep learning based solar flare forecasting model. i. results for line-of-sight magnetograms," *The Astrophysical Journal*, vol. 856, no. 1, p. 7, 2018.
- [11] A. Ahmadzadeh, M. Hostetter, B. Aydin, M. K. Georgoulis, D. J. Kempton, S. S. Mahajan, and R. Angryk, "Challenges with extreme class-imbalance and temporal coherence: A study on solar flare data," in *2019 IEEE International Conference on Big Data (Big Data)*, Dec 2019, pp. 1423–1431.
- [12] G. Barnes *et al.*, "A comparison of flare forecasting methods. i. results from the "all-clear" workshop," *The Astrophysical Journal*, vol. 829, no. 2, p. 89, Sep 2016. [Online]. Available: <https://iopscience.iop.org/article/10.3847/0004-637X/829/2/89>
- [13] N. O'Mahony, S. Campbell, A. Carvalho, S. Harapanahalli, G. V. Hernandez, L. Krpalkova, D. Riordan, and J. Walsh, "Deep learning vs. traditional computer vision," in *Science and Information Conference*. Springer, 2019, pp. 128–144.
- [14] Z. Wang and T. Oates, "Imaging time-series to improve classification and imputation," *arXiv preprint arXiv:1506.00327*, 2015.
- [15] Y. Zhu, C. Liu, and K. Sun, "Image embedding of pmu data for deep learning towards transient disturbance classification," in *2018 IEEE International Conference on Energy Internet (ICEI)*. IEEE, 2018, pp. 169–174.
- [16] G. Krummenacher, C. S. Ong, S. Koller, S. Kobayashi, and J. M. Buhmann, "Wheel defect detection with machine learning," *IEEE Transactions on Intelligent Transportation Systems*, vol. 19, no. 4, pp. 1176–1187, 2017.
- [17] A. B. Said, A. Erradi, A. G. Neiat, and A. Bouguettaya, "A deep learning spatiotemporal prediction framework for mobile crowdsourced services," *Mobile Networks and Applications*, vol. 24, no. 3, pp. 1120–1133, 2019.
- [18] J.-H. Chen and Y.-C. Tsai, "Encoding candlesticks as images for pattern classification using convolutional neural networks," *Financial Innovation*, vol. 6, pp. 1–19, 2020.
- [19] J. Eckmann, S. O. Kamphorst, D. Ruelle *et al.*, "Recurrence plots of dynamical systems," *World Scientific Series on Nonlinear Science Series A*, vol. 16, pp. 441–446, 1995.
- [20] R. Rajabi and A. Estebarsari, "Deep learning based forecasting of individual residential loads using recurrence plots," in *2019 IEEE Milan PowerTech*. IEEE, 2019, pp. 1–5.
- [21] R. Zhang, F. Zheng, and W. Min, "Sequential behavioral data processing using deep learning and the markov transition field in online fraud detection," *arXiv preprint arXiv:1808.05329*, 2018.
- [22] A. Benz, "Flare observations," *Living Reviews in Solar Physics*, vol. 5, 02 2008. [Online]. Available: doi.org/10.12942/lrsp-2008-1
- [23] C. Liu, N. Deng, J. T. Wang, and H. Wang, "Predicting solar flares using sdo/hmi vector magnetic data products and the random forest algorithm," *The Astrophysical Journal*, vol. 843, no. 2, p. 104, 2017.
- [24] M. G. Bobra, X. Sun, J. T. Hoeksema, M. Turmon, Y. Liu, K. Hayashi, G. Barnes, and K. Leka, "The helioseismic and magnetic imager (hmi) vector magnetic field pipeline: Sharps-space-weather hmi active region patches," *Solar Physics*, vol. 289, no. 9, pp. 3549–3578, 2014.
- [25] 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, M. A. Schuh, and M. K. Georgoulis, "Multivariate time series dataset for space weather data analytics," *Scientific Data*, vol. 7, no. 1, p. 227, Jul. 2020.
- [26] 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>
- [27] A. Ahmadzadeh, K. Sinha, B. Aydin, and R. A. Angryk, "Mvts-data toolkit: A python package for preprocessing multivariate time series data," *SoftwareX*, vol. 12, p. 100518, 2020.
- [28] P. A. Gagniuc, *Markov chains: from theory to implementation and experimentation*. John Wiley & Sons, 2017, pp. 9–11.
- [29] J. Faouzi and H. Janati, "pyts: A python package for time series classification," *Journal of Machine Learning Research*, vol. 21, no. 46, pp. 1–6, 2020. [Online]. Available: <http://jmlr.org/papers/v21/19-763.html>
- [30] M. Cuturi, "Fast global alignment kernels," in *Proceedings of the 28th international conference on machine learning (ICML-11)*, 2011, pp. 929–936.
- [31] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.
- [32] B. Krawczyk, "Learning from imbalanced data: open challenges and future directions," *Progress in Artificial Intelligence*, vol. 5, no. 4, pp. 221–232, 2016.
- [33] A. Hanssen and W. Kuipers, *On the Relationship Between the Frequency of Rain and Various Meteorological Parameters (with Reference to the Problem of Objective Forecasting)*. Koninklijk Nederlands Meteorologisch Instituut, 1965.
- [34] P. Heidke, "Berechnung des erfolges und der güte der windstärkevorhersagen im sturmwarnungsdienst," *Geografiska Annaler*, vol. 8, no. 4, pp. 301–349, 1926. [Online]. Available: doi.org/10.1080/20014422.1926.11881138
- [35] C. C. Balch, "Updated verification of the space weather prediction center's solar energetic particle prediction model," *Space weather : the international journal of research & applications*, vol. 6, no. 1, 2008. [Online]. Available: doi.org/10.1029/2007SW000337