



Detecting and Classifying Flares in High-resolution Solar Spectra with Supervised Machine Learning

Nicole Hao¹ , Laura Flagg^{1,2} , and Ray Jayawardhana²

¹ Cornell University, Ithaca, NY 14853, USA; laura.s.flagg@gmail.com

² Department of Physics and Astronomy, Johns Hopkins University, 3400 N. Charles Street, Baltimore, MD 21218, USA

Received 2024 March 4; revised 2024 June 24; accepted 2024 June 24; published 2024 September 24

Abstract

Flares are a well-studied aspect of the Sun’s magnetic activity. Detecting and classifying solar flares can inform the analysis of contamination caused by stellar flares in exoplanet transmission spectra. In this paper, we present a standardized procedure to classify solar flares with the aid of supervised machine learning. Using flare data from the RHESSI mission and solar spectra from the HARPS-N instrument, we trained several supervised machine-learning models, and found that the best-performing algorithm is C-Support Vector Classification (SVC) with nonlinear kernels, specifically radial basis functions (RBF). The best-trained model, SVC with RBF kernels, achieves an average aggregate accuracy score of 0.65, and categorical accuracy scores of over 0.70 for the no-flare and weak-flare classes, respectively. In comparison, a blind classification algorithm would have an accuracy score of 0.33. Testing showed that the model is able to detect and classify solar flares in entirely new data with different characteristics and distributions from those of the training set. Future efforts could focus on enhancing classification accuracy, investigating the efficacy of alternative models, particularly deep learning models, and incorporating more data sets to extend the application of this framework to stars that host exoplanets.

Unified Astronomy Thesaurus concepts: [The Sun \(1693\)](#); [High resolution spectroscopy \(2096\)](#); [Red dwarf flare stars \(1367\)](#); [Solar flares \(1496\)](#); [Exoplanet atmospheres \(487\)](#); [Stellar activity \(1580\)](#); [Stellar astronomy \(1583\)](#)

1. Introduction

Transmission spectroscopy is highly useful and widely used for characterizing exoplanets since it can yield valuable constraints on the nature and composition of planetary atmospheres. Yet, due to the inhomogeneity and time variability of the stellar photo- and chromospheres, this method is intrinsically impacted by stellar spectral contamination (Rackham et al. 2023). Often, the stellar contamination will rival or even exceed the planetary spectral features, making it very difficult to disentangle the exoplanet atmospheric signals from stellar contamination (Rackham et al. 2023).

Stellar flares are another source of stellar contamination, resulting in increased brightness in exoplanet transmission spectra. They are wavelength dependent, which leads to additional noise and distortions of the observed spectra across a range of wavelengths. Such contamination poses a challenge for measuring an exoplanet’s transit depth accurately. To consider the impact of stellar flares, it is helpful to begin with an investigation of solar flares, given the abundance of flare data for the Sun. Efficient detection and classification methods for solar flares in transmission spectra could help astronomers correct for stellar contamination in exoplanet transmission spectra with greater accuracy.

This study has been designed with two primary objectives. First, it aims to address the challenge of detecting flare events in high-resolution solar spectra. To achieve this, we correlated solar spectra with solar flare events based on their start and end times, labeled solar spectra, and fed labeled solar spectra into supervised machine-learning models, which enabled accurate

detection of flares based on their energy levels. Second, the project strives to develop a robust machine-learning model that can classify flares in solar spectra as accurately as possible. Identifying low-energy flares in high-resolution solar spectra may present a greater difficulty compared to their high-energy counterparts. However, the impact of low-energy flares on spectra should not be disregarded. Thus, we aim to detect and classify all flares in solar spectra, regardless of their energy levels.

In this study, we employed supervised learning algorithms, specifically Support Vector Classification (SVC), to detect and classify solar flares in high-resolution solar spectra. Our methodology encompasses a standardized procedure for classifying flares, primarily based on their energy levels. To assess the performance of the classification models, we utilized multilabel metrics such as accuracy scores. By leveraging these evaluation measures, we aim to provide a framework for detecting and categorizing solar flares in the context of high-resolution spectra.

The paper is organized as follows. Section 2 considers the data sets used, followed in Section 3 by a description of our methods, including data selection, analysis, and model testing. Section 4 addresses the selection of supervised learning algorithms and the validation of results using multilabel metrics, and includes a discussion on the optimization of model performance through means such as performing a grid search and tweaking hyperparameters.

2. Data

Solar spectra, ranging from 2015 July to 2018 April, were drawn from the Data & Analysis Center for Exoplanets (2020). The spectra were collected with HARPS-N, an optical spectrograph installed at the Italian Telescopio Nazionale Galileo, designed specifically for the precise measurement of



Original content from this work may be used under the terms of the [Creative Commons Attribution 4.0 licence](#). Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

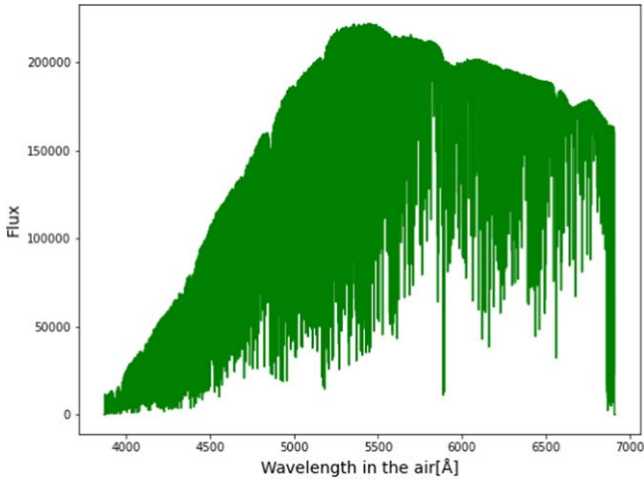


Figure 1. Solar spectrum, a single observation.

radial velocities in the search for exoplanets (Cosentino et al. 2012). The instrument offers a high resolution of $\sim 120,000$, allowing for detailed spectral analysis, and covers a range of wavelengths 378–691 nm. Figure 1 shows the flux from one observation against wavelength in the air. One observation here refers to the single set of data collected by the HARPS-N over 5 minutes, capturing the intensity of light across different wavelengths from the Sun (Dumusque et al. 2021).

Solar flare data used in this investigation were collected by RHESSI (NASA Goddard Space Flight Center 2003). RHESSI was a NASA satellite mission designed to study the Sun at high-energy X-ray and gamma-ray wavelengths, with a focus on investigating the particle acceleration and energy release processes in solar flares (Lin et al. 2003). The RHESSI flare list covers a time period of approximately 17 yr from 2002 to 2019. Since our study includes solar spectra only for the 2015–2018 period, we used solar flare data for the same date range. The RHESSI data contain the following information for each observation: start times and end times, solar peak times, energy levels, duration, total peak counts, X-position and Y-position of the event on the solar disk, radial distance of a solar flare event from the center of the Sun, active region, and flags. We removed data points from the data set that contained zero information. For example, if the X-position and Y-position of one observation are both zero, the Spectroscopic Imager on the satellite failed to collect any information; therefore, the data point has to be removed to prevent it from introducing errors into the machine-learning results.

3. Methods

3.1. Selection of Wavelength Range, Normalization, and Principal Component Analysis

For this study, we used solar spectra spanning 6400 Å–6700 Å as our training data. We selected this particular wavelength range because it contains $H\alpha$, which exhibits increased emission during a solar flare (Ichimoto & Kurokawa 1984).

We conducted a series of preprocessing steps on the data: First, we plotted the solar spectra for all distinct observations and identified the region exhibiting the least fluctuation in flux. Next, we calculated the 98th percentile within that region and divided the flux values of all data points by this value, generating a set of normalized flux values.

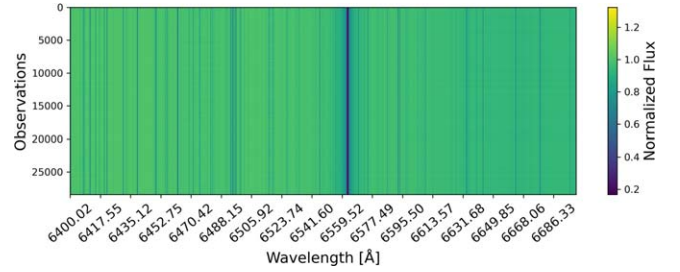


Figure 2. Plot of normalized flux against wavelengths.

To reduce the effects of high-noise data on our training results, specifically outliers that appear as spikes on the plot of a solar spectrum observation, we replaced all of the normalized flux values that are five standard deviations or more from the mean of the normalized spectra with the average of their neighboring values. See Figure 2 for a plot of noise-filtered normalized spectra. Then, we applied principal component analysis (PCA) to the processed data. PCA is a mathematical technique that reduces the dimensionality of training data by projecting data points onto a lower dimensional space that best captures the variance of the original data, therefore enhancing the efficiency of the subsequent machine-learning process (Abdi & Williams 2010). Our processed data initially consisted of 28,415 observations and 16,747 wavelength bins. After PCA, we reduced the dimension of the normalized data to (28,415, 1000) with 1000 principal components. Figure 3 shows the first 10 principal components of the normalized solar spectra.

3.2. Data Correlation and Labeling

Supervised learning algorithms take in a collection of labeled data. The labeled solar spectra represent the input values and their labels represent the output values. We approximated the true relationship between the input value and their labels using machine learning by labeling each normalized solar spectrum observation after PCA as one of the following three labels: no flares, weak flares, and strong flares. We extracted flare events from the RHESSI data and their corresponding energy band. In the RHESSI data, each observation of a potential flare is considered a “flare event,” and each has a corresponding range of energy that was observed in the duration of the flare, referred to as the event’s “energy band.” The flare events were divided into three label classes: spectra without any flare events are categorized as “no flares”; events with the energy band 6–12 keV are categorized as “weak flares”; all solar events with energy bands greater than 6–12 keV are categorized as “strong flares.” Visually, there are no differences between the spectra despite some spectra having flares and others not.

The process of data correlation and labeling must also consider the temporal differences between the observations from RHESSI and $H\alpha$ line data. RHESSI operates in the high-energy regime from soft X-rays to gamma rays within the coronal region, whereas the $H\alpha$ line originates from the chromosphere. Despite documented temporal differences between these regions, as supported by relevant literature, these differences are not consistently predictable (Gontikakis et al. 2023). Given the variability and our lack of precise knowledge about these time discrepancies, our investigation does not explicitly account for them. Additionally, solar flares, which are the focus of our study, persist over extended periods

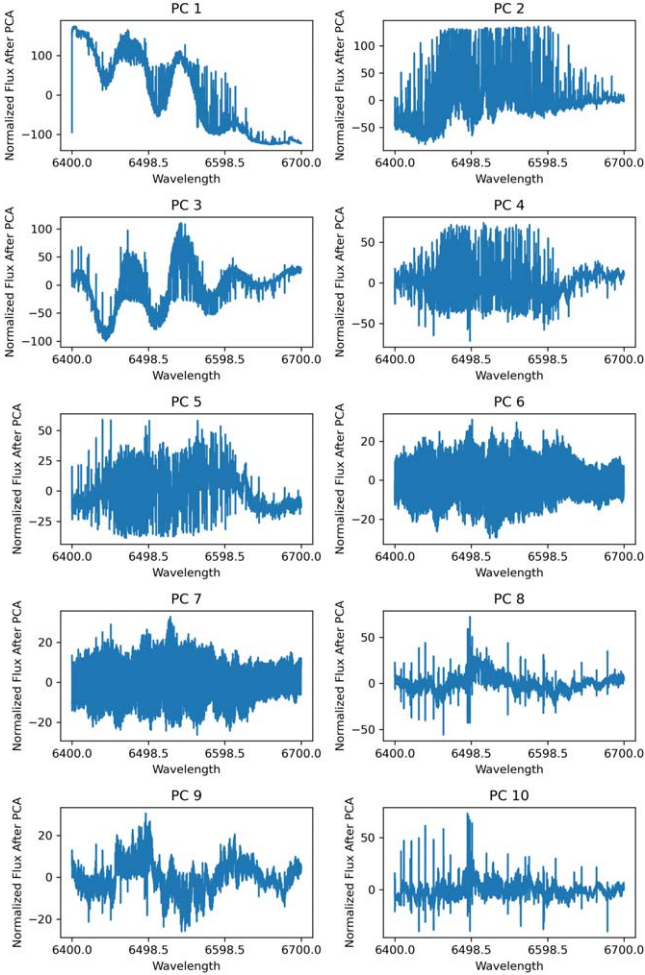


Figure 3. First 10 principal components of normalized solar spectra.

rather than occurring as instantaneous events. Our approach assumes the effectiveness of our current method, and the performance of our final machine-learning model supports this assumption, showing satisfactory results.

3.3. Imbalanced Data and Undersampling

Undersampling is a strategy that resolves the issue of imbalanced data by reducing the size of the overrepresented class (He & Garcia 2009; Mosley 2013). In our study, the dominant class was the no-flare class. Initially, we detected 28,415 flare events from the RHESSI data set corresponding to solar spectra observations. Notably, the strong flare class exhibited the lowest representation, comprising only 467 instances. To preserve the model’s ability to classify strong flares when undersampling, we randomly selected 467 no-flare spectra and 467 weak flare spectra from the PCA 1000 components data using the *make_imbalance* function from the *imbalanced-learn* module. The ratios of data before and after undersampling are represented in Figure 4. For each class, there are 467 spectra, which makes the undersampled set 1401 spectra.

3.4. Testing Supervised Learning Models

Since we divided the flares into three classes, we selected models that are best fit for multiclass training and compared their performance on two tasks: the average confusion matrices

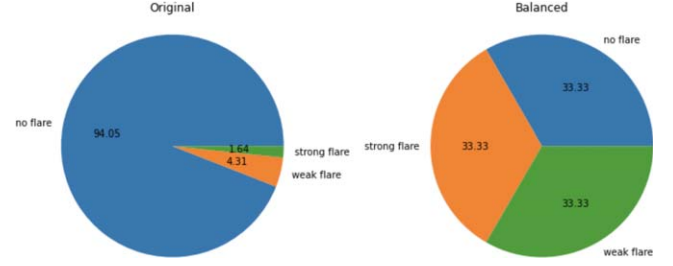


Figure 4. Original solar flare data proportion (left); balanced solar flare data (right).

over 10 trials and categorical accuracy scores (see Section 4.1 Equation (3) for the definition of categorical accuracy scores). The models we selected are random forest, SVC with stochastic gradient descent, and C-SVC, all implemented with *sklearn* (Pedregosa et al. 2011). We selected these models because they are well suited for supervised multilabel classification tasks and consistently outperformed other models designed for similar problems (Pedregosa et al. 2011).

We evaluated the performance of each model using its confusion matrix averaged over 10 trials. For each trial, we split the data into two sets: 80% went into the training set, while 20% went into the test set. A confusion matrix is a structured arrangement used to visualize the effectiveness of an algorithm, often used within the context of supervised learning (Stehman 1997). The matrix expands to an $n \times n$ format, where n is the number of categories or classes. In our case, we have a 3×3 matrix. Each row represents the actual classes, while each column represents the predicted classes. Each entry in the matrix shows the number of observations from the actual class (row) that were predicted to be in a specific class (column) (Stehman 1997). Therefore, visually, the brighter the diagonal of the heat map generated from the confusion matrix is, the more instances from each class are classified correctly.

We created the average confusion matrices by summing the confusion matrices obtained from individual trials entry by entry and then dividing them by the number of trials. We chose to run 10 trials because unlike reinforcement learning or deep learning models, there is less randomness in the training process for supervised learning; therefore, it is not necessary to run the models for a large number of trials to get their average performance.

4. Results and Discussion

4.1. Model Evaluation and Selection

We trained four models and analyzed their performance to identify the most promising candidates for further tuning and optimization. Table 1 shows the descriptions of the models and their hyperparameters. Hyperparameters are parameters of the model that are not determined before the training process, which can be tuned to minimize the generalization error or underfitting (Probst et al. 2018). None of the models that we used in the study use the default hyperparameters from *sklearn*. We selected these specific hyperparameters to optimize the training results based on experiments we ran previously.

Kernels are mathematical functions used to transform the nonlinearly separable data into a higher-dimensional space, where the data can be separated by a linear equation (Vert et al. 2004). We tried three kernels: polynomial, sigmoid, and radial basis function (RBF). RBF proved to have the best results on

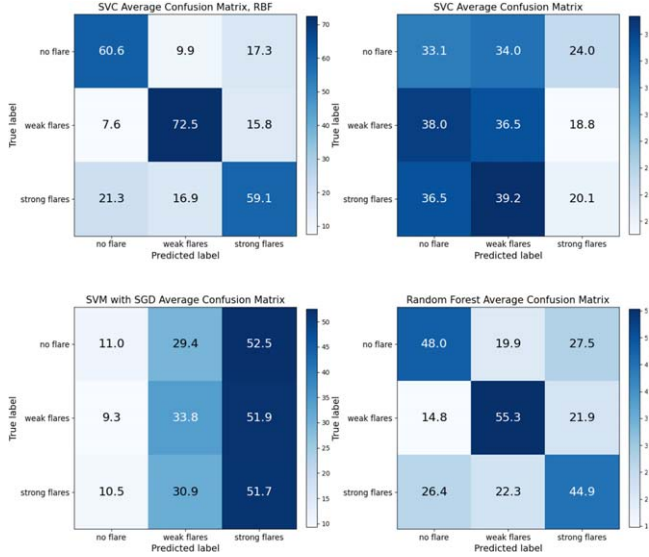


Figure 5. Average confusion matrices for SVC with an RBF kernel (top left), SVC with polynomial kernel (top right), SVM with stochastic gradient optimizer (bottom left), and random forest (bottom right).

average. RBF is described by the following equation:

$$K(\mathbf{x}, \mathbf{x}') = \exp\left(-\frac{\|\mathbf{x} - \mathbf{x}'\|^2}{2\sigma^2}\right), \quad (1)$$

where \mathbf{x} and \mathbf{x}' are two distinct data samples from the data set, and σ is a free parameter (Vert et al. 2004).

Figure 5 shows the average confusion matrices for the four models we trained. The average confusion matrices of SVC with an RBF kernel and random forest have the brightest diagonals, which shows that the models can predict most of the data points in each class correctly, especially when compared to the confusion matrices of SVC with polynomial kernel and SVM with stochastic gradient descent optimizer. This confirms that our models have learned to correctly detect the presence of solar flare events and classify some of them correctly using solar spectra. We further evaluated the performance of random forest and SVC with an RBF kernel by comparing their accuracy scores and categorical accuracy scores. Accuracy score, or aggregated accuracy score as seen in the following Equation (2), is a commonly used metric that calculates the overall accuracy of a classifier. It measures the proportion of correctly predicted instances out of the total instances in the data set (Mosley 2013).

$$\text{Aggregate accuracy} = \frac{\text{number of correct predictions}}{\text{total number of predictions}}. \quad (2)$$

Since aggregate accuracy is “blind” to specific classes, we introduced another metric called the categorical accuracy score to evaluate the performance of the model for each class. Equation (3) shows the formula for the categorical accuracy score, which is a variant of the aggregate accuracy score specifically designed for multiclass classification problems, like the problem in this study where there are three classes instead of only two (Mosley 2013). It calculates the accuracy considering each class separately. In other words, it calculates the proportion of correctly predicted instances for each class

Table 1
Description of Hyperparameters for Models

Algorithm	Description of Hyperparameters
Random Forest	Number of Estimators: 100
SGDClassifier	Loss Function: Hinge
C-SVC	Kernel: Polynomial Degree: 2 C (Regularization Parameter): 1.0
C-SVC	Kernel: Sigmoid C (Regularization Parameter): 10 sigma: 0.0001

out of the total instances belonging to that class.

$$\text{Categorical accuracy} = \frac{\text{number of correct predictions in the class}}{\text{total number of instances in the class}}. \quad (3)$$

Other than accuracy scores, we also used metrics like precision and recall class-wise to further evaluate the performance of each model. The formulas for precision and recall are (Kelleher et al. 2015)

$$\text{precision}_{\text{Class A}} = \frac{\text{TP}_{\text{Class A}}}{\text{TP}_{\text{Class A}} + \text{FP}_{\text{Class A}}}, \quad (4)$$

$$\text{recall}_{\text{Class A}} = \frac{\text{TP}_{\text{Class A}}}{\text{TP}_{\text{Class A}} + \text{FN}_{\text{Class A}}}, \quad (5)$$

where “TP” is true positives, the number of instances that were positive in the data set and were correctly classified as positive by the model, “FN” is false negatives, the number of instances the model failed to identify as positive when they actually were, and “FP” represents false positives, the number of instances the model mistakenly classified as positive when they were actually negative. As seen in Equation (4), precision for a given class in multiclass classification is the fraction of instances correctly classified as belonging to a specific class out of all instances the model predicted to belong to that class (Kelleher et al. 2015). In other words, precision tells you the ratio of true positives to all instances that the model predicted as positive. Precision measures the accuracy of all positive predictions. Recall for a given class is the ratio between correctly classified positive instances and all positive instances in the data (Kelleher et al. 2015). It measures whether all positive instances are predicted correctly.

Another evaluation metric we used was the F1 score, which evaluates the performance of a model using both precision and recall. Equation (6) shows the formula for the F1 score.

$$F_1 = \frac{2}{\text{recall}^{-1} + \text{precision}^{-1}}. \quad (6)$$

A high F1 score indicates a balanced performance, meaning that the model is both accurate in predicting positive instances and is able to classify most of the positive instances correctly (Powers 2020).

4.2. SVC with an RBF Kernel

Figure 6 shows that in our preliminary experiments with different models SVC with an RBF kernel had the best performance in both aggregate and categorical accuracy scores

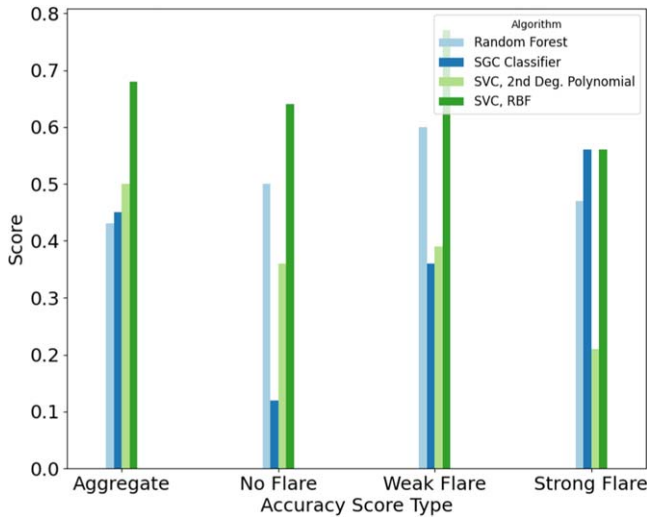


Figure 6. Categorical and aggregated accuracy score of each algorithm.

for all three classes. The average aggregate accuracy score over 10 trials is 0.68. The average categorical accuracy scores for each of the classes—no flare, weak flares, and strong flares—are 0.64, 0.77, and 0.56, respectively.

Figure 7 shows the average precision, recall, and F1 scores of each model. The random forest model demonstrates a balanced performance with both precision and recall at around 0.53. This suggests that it maintains a good equilibrium between correctly identifying positive cases and minimizing false positives. The Stochastic Gradient Descent Classifier (SGDClassifier), on the other hand, exhibits lower precision (0.28) and recall (0.35) compared to the random forest. This indicates that it might struggle with both correctly classifying positive instances and minimizing false positives. The F1 score of 0.21 underscores this performance gap, implying that this model may not be the best option in situations where precision and recall are critical. The SVC with a second-degree polynomial kernel shows similarly suboptimal performance with a low precision of 0.22 and recall of 0.33. Its F1 score of 0.22 suggests that it does not excel at either precision or recall. This model appears to underperform in comparison to the random forest. The SVC with an RBF kernel has the highest precision, recall, and F1 score, all at 0.67. This model performs well in both correctly classifying positive instances and minimizing false positives compared to the other models. Based on these results, We decided to further optimize the SVC model with an RBF kernel.

The SVC function from *scikit-learn* with an RBF kernel has two hyperparameters, C and γ , that allow us to find the right value of variance that optimizes the model performance with the RBF kernel function (Pedregosa et al. 2011). C value is the regularizer parameter that can be manipulated to control the extent of overfitting. A high C value may lead to overfitting, and a low C value may lead to underfitting, where the model is too generalized to identify the pattern in training data. γ defines how far the influence of a single training example reaches. A high γ value suggests “close” and may lead to overfitting because it requires the data points to be close to group them; a low γ value suggests “far” and may lead to underfitting (Pedregosa et al. 2011). We applied *GridSearchCV* from

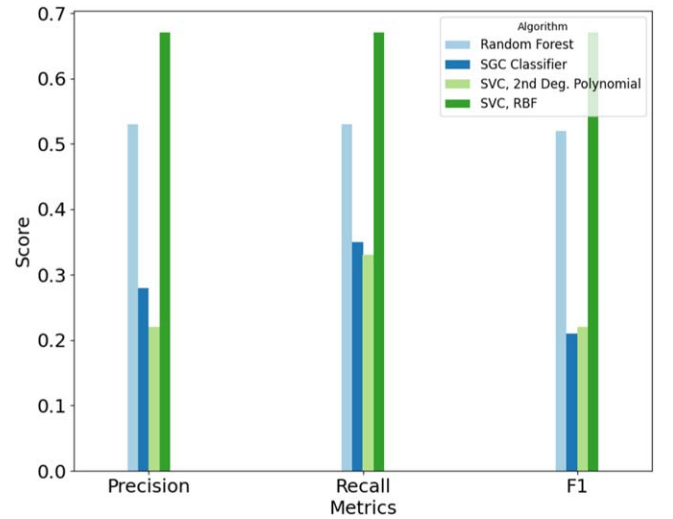


Figure 7. Average precision, recall, and F1 scores of each algorithm.

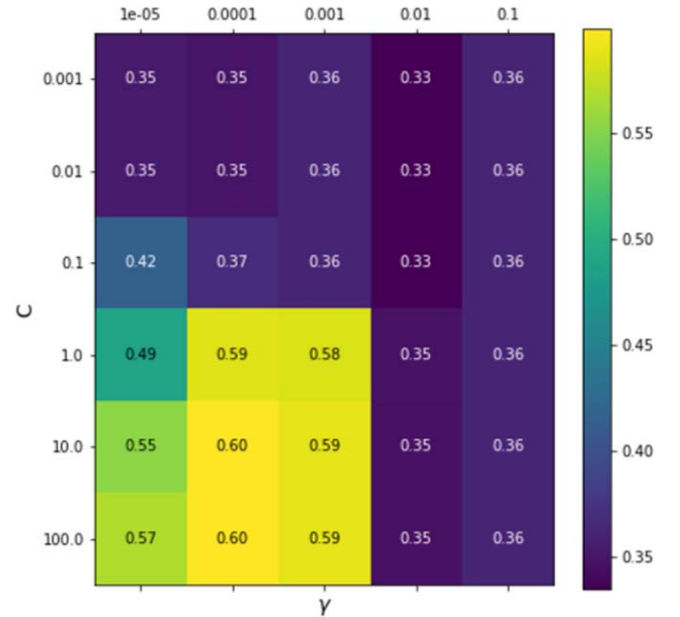


Figure 8. Grid search, aggregate accuracy, and corresponding C and γ values of SVC with an RBF kernel.

sklearn to determine which combination of C value and γ value results in the best aggregate accuracy score. The C value and γ value combinations we tested and their corresponding aggregate accuracy scores are in Figure 8.

To determine good values of the hyperparameters, it is important to search on the right scale. (Chih-Wei et al. 2016) We decided to test a range of C values, 0.001, 0.01, 0.1, 1.0, 10.0, and 100.0, aiming to encompass a broad spectrum, since the most common range of C values to test is 1–10 (Chih-Wei et al. 2016). Initially, our empirical testing focused on the common range of 0–1.0 for C . Notably, we observed a substantial increase in accuracy scores as C surpassed 1, particularly with γ values of 0.00001, 0.0001, and 0.001. However, for C values higher than 10.0, we detected no significant accuracy improvement; instead, there was a slight decrease. Moreover, C values exceeding 100 exhibited a higher likelihood of leading to overfitting. Consequently, we decided to stop further testing C values higher than 100.

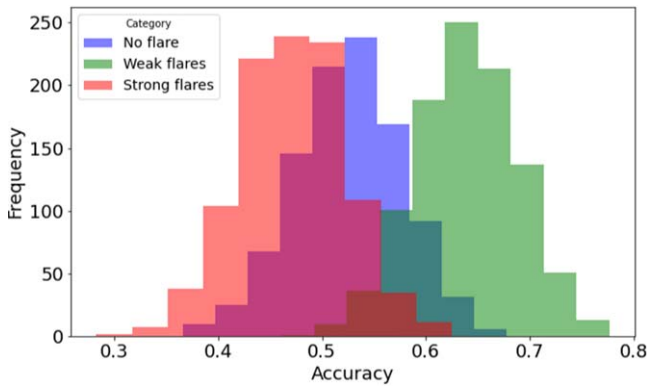


Figure 9. Accuracy score distribution across 1000 trials for no flare, weak flares, and strong flares.

Similarly, we tried different γ values to find one that balances the variance and the bias of our model. The γ value corresponds to the margin of the kernel function in the higher-dimensional space that training data were projected onto. In this higher-dimensional space, we partition the training data into three sections, corresponding to the number of classes in our multiclass classification problem. The boundaries at the intersection of these three spaces are called the decision boundaries (Kelleher et al. 2015). A high γ value means only the closest points to the decision boundary will carry the weight leading to a smoother boundary, which likely results in overfitting. Whereas a low γ value corresponds to a larger margin that contains more data points, which leads to underfitting. We tested a range of γ values to see their effects on the training results and whether decreasing the γ value would lead to high-accuracy model performance. The values chosen were 0.00001, 0.0001, 0.001, 0.01, and 0.1. We chose this range because, theoretically, choosing exponentially growing sequences of C and γ values is more efficient when determining good parameters using a grid search (Chih-Wei et al. 2016). In practice, it is good to try a γ value that is $6/k$, where k is the number of input data samples (Chapelle & Zien 2005). Here, we have 1401 data points, which theoretically makes $6/1401 = 0.002$ an ideal γ value. Nonetheless, using a moderately coarse grid helps to identify the optimal region within it. Then, we can empirically determine which region on the coarse grid results in better performance. This is why we chose a slightly wider range of C and γ values compared to common practice.

As seen in Figure 9, when $C = 10.0$ and $\gamma = 0.0001$, the SVC model with an RBF kernel achieves an aggregate accuracy score of 0.60, which is the highest out of all tested combinations. We identified that the model performance is optimal when C is between 10.0 and 100.0, and when γ is between 0.0001 and 0.001.

5. Conclusions

Based on the results presented above, we concluded that it is possible to detect and classify solar flares in optical high-resolution spectra using supervised learning algorithms. We used SVC with an RBF kernel to categorize solar flares into three classes. The model exhibited an overall accuracy score of 0.65, showcasing its ability to distinguish among these distinct flare categories. A blind classification algorithm would have an accuracy score of 0.33, so our algorithm is a significant

improvement. Nonetheless, there are some apparent limitations to our findings.

One limitation is the model’s comparatively low accuracy in classifying the “strong flare” category, as evident from a categorical accuracy of 0.56. A possible strategy for improvement is supplementing the training data with more actual data on strong flares; increasing the model’s exposure to this class should enhance its ability to classify strong flares accurately.

Another limitation is the overall performance of the SVC model with the RBF kernel. The aggregate accuracy shows that more than half of the time the model can classify most data points correctly, but there is still room for performance enhancement. The parameter choices, such as C (set at 10) and γ (0.0001), may not be optimally configured, necessitating a more comprehensive exploration of hyperparameter settings. To further improve the overall accuracy of our current SVC model, a multifaceted approach can be adopted. First, fine-tuning the hyperparameters, such as the regularization parameter (C) and the kernel-specific parameter (γ), through a more extensive grid search can lead to an optimized model configuration. This would enable us to strike the right balance between model complexity and generalization.

Also, while there are documented time differences between these regions, as supported by relevant literature, these differences are not consistently predictable. Given the variability and our lack of precise knowledge about these time discrepancies, our investigation does not explicitly account for them. Additionally, solar flares, which are the focus of our study, persist over extended periods rather than occurring as instantaneous events. Our approach assumes the effectiveness of our current method, and the performance of our final machine-learning model supports this assumption, showing satisfactory results.

Moreover, our findings reveal implications for the characteristics of the underlying data. Notably, the RBF kernel demonstrated significant performance in classifying weak flares, both before and after hyperparameter tuning. The model achieves an average categorical accuracy score of 0.77 before tuning, and an average categorical accuracy score of 0.80 after tuning. This suggests that weak flares may exhibit distinct and nonlinear patterns effectively captured by the RBF kernel. In other words, the data may not be homogeneous, and the classes may have varying degrees of complexity. One possible explanation could be that all of the weak flares are similar to each other, while the strong flares represent a larger range of flare energies. Therefore, when our model is tested on the testing data, it might be less accurate when predicting strong flares compared to weak flares.

In future work, one could explore other learning algorithms to determine if they achieve better performance. Ensemble learning methods, such as random forest and gradient boosting, harness the collective power of multiple models and potentially improve overall classification accuracy. Also, given the nonlinear nature of the underlying data, as evidenced by the performance of the RBF, it makes sense to consider the potential of employing deep learning techniques as a next step. Deep learning could potentially address the local cluster patterns within our high-dimensional data due to its capability to uncover hidden structures and nuances that may elude traditional machine-learning models. It is possible that a neural network with many neurons can preserve and identify the complex patterns in strong flare data that supervised learning models struggle to capture.

Our work reported here represents an initial investigation aimed at automating the detection and classification of flares in high-resolution solar spectra. While we have achieved the development of a model capable of classifying solar flares within this context, future efforts should focus on enhancing the prediction accuracy and exploring the potential of alternative models, including deep learning approaches, to further refine the classification capabilities.

The longer-term vision is to develop a robust framework for detecting and categorizing stellar flares not just in our solar system, but also in the broader context of exoplanetary systems. The latter would enable more accurate corrections for stellar contamination. Extending such an approach to the host stars of exoplanets involves adapting the SVC models to cater to the particular characteristics of these stars' spectra. Since the spectral signatures and flare activities of exoplanet host stars might differ from those of the Sun, the model would require recalibration and retraining with relevant data sets. The recalibration would involve adjusting the model to recognize flare signatures in different stellar environments, taking into account factors such as the star's size, age, and magnetic activity.


Acknowledgments

N.H. acknowledges support from the Nexus Scholars program in the College of Arts and Sciences at Cornell University. This publication makes use of The Data & Analysis Center for Exoplanets (DACE), which is a facility based at the University of Geneva (CH) dedicated to extrasolar planets data visualisation, exchange and analysis. DACE is a platform of the Swiss National Centre of Competence in Research (NCCR) PlanetS, federating the Swiss expertise in Exoplanet research. The DACE platform is available at <https://dace.unige.ch>.

Code Availability

The code used in this project can be accessed at https://github.com/nicolehao34/solar_flares_classification.

ORCID iDs

Nicole Hao  <https://orcid.org/0009-0009-8530-6183>
 Laura Flagg  <https://orcid.org/0000-0001-6362-0571>
 Ray Jayawardhana  <https://orcid.org/0000-0001-5349-6853>

References

- Abdi, H., & Williams, L. J. 2010, *Wiley Interdisciplinary Reviews: Computational Statistics*, 2, 433
- Chapelle, O., & Zien, A. 2005, International Conference on Artificial Intelligence and Statistics, R5, 57, <https://api.semanticscholar.org/CorpusID:14283441>
- Chih-Wei, H., Chih-Chung, C., & Chih-Jen, L. 2016, A Practical Guide to Support Vector Classification, Tech. rep., National Taiwan Univ. <http://www.csie.ntu.edu.tw/~cjlin>
- Cosentino, R., Lovis, C., Pepe, F., et al. 2012, *Proc. SPIE*, 8446, 84461V
- Department of Astronomy, University of Geneva 2020, HARPS-N Solar Spectroscopy, Data & Analysis Center for Exoplanets (DACE), <https://dace.unige.ch/observationSearch/?observationType=%22solarSpectroscopy%22>
- Dumusque, X., Cretignier, M., Sosnowska, D., et al. 2021, *A&A*, 648, A103
- Gontikakis, C., Antiochos, S. K., & Young, P. R. 2023, *ApJ*, 943, 120
- He, H., & Garcia, E. 2009, *IEEE Trans. Knowl. Data Eng.*, 21, 1263
- Ichimoto, K., & Kurokawa, H. 1984, *SoPh*, 93, 105
- Kelleher, J. D., MacNamee, B., & D'Arcy, A. 2015, *Fundamentals of machine learning for predictive data analytics: algorithms, worked examples, and case studies* (Cambridge, MA: MIT Press)
- Lin, R. P., Dennis, B. R., Hurford, G. J., et al. 2003, in *The Reuven Ramaty High-Energy Solar Spectroscopic Imager (RHESSI): Mission Description and Early Results*, ed. R. P. Lin, B. R. Dennis, & A. O. Benz (Dordrecht: Springer), 3
- Mosley, L. 2013, Doctor of Philosophy thesis, Iowa State Univ. doi:[10.31274/etd-180810-3375](https://doi.org/10.31274/etd-180810-3375)
- NASA Goddard Space Flight Center 2003, RHESSI Flare List, https://hesperia.gsfc.nasa.gov/hessidata/dbase/hessi_flare_list.txt
- Pedregosa, F., Varoquaux, G., Gramfort, A., et al. 2011, *JMLR*, 12, 2825
- Powers, D. 2020, arXiv:[2010.16061](https://arxiv.org/abs/2010.16061)
- Probst, P., Bischl, B., & Boulesteix, A.-L. 2018, *JMLR*, 20, 53, <http://jmlr.org/papers/v20/18-444.html>
- Rackham, B. V., Espinoza, N., Berdyugina, S. V., et al. 2023, *RASTI*, 2, 148
- Stehman, S. V. 1997, *RSEnv*, 62, 77
- Vert, J.-P., Tsuda, K., & Schölkopf, B. 2004, in *Kernel Methods in Computational Biology*, ed. B. Schölkopf, K. Tsuda, & J.-P. Vert (Cambridge, MA: MIT Press), 35