Dan Aguirre, Jordan Crews, Jay Daie, Katherine Davis, Chris Kimberly, Sanjay Lindsay Team 43

Solar Flares Analysis Project - CSE 6242 Project Proposal

Project Definition

This project will culminate in an interactive web-based dashboard providing amateur astronomers, students, teachers, and citizen scientists with a resource that can be used to visualize and understand solar flare data. The tool will display key solar activity data, such as classifying solar flares by intensity or energy (A, B, C, M, or X), utilizing interactive animations, such as a heliographic coordinate visualization of the locations where flares have occurred over the sun's surface, and striving to predict future larger flares occurrences.

How is it done today?

Lopez-Urias et al. highlights the influence of solar activity on the ionosphere [as] a critical area of investigation due to its relevance to the Sun-Earth relationship (Lopez-Urias et al., 2023). Modern research in the field has linked solar activity to several vital phenomena in our own Earth's atmosphere, including important implications on radio-wave propagation (Hays L. A. et al., 2021), satellite applications, operations, and critical systems support (Fagundes P.R. et al. 2024), advancing our ability to diagnose rapid changes in geospace weather (Liu J. et al., 2021), and even impacts to the mesosphere, where many jet aircraft fly, the troposphere where weather occurs, and the ozone layer, which absorbs harmful rays from the Sun (Kumar P. et al., 2015). Today, many government agencies (NASA, the National Weather Service, National Oceanic and Atmospheric Administration, etc.) and academic research institutions use solar activity data to monitor and protect against many of the above-mentioned impacted areas. These sources are primarily hyper-focused on one specific issue and highly technical in nature (not generally consumable by the public).

Our Approach

Many institutions study solar flares, but most are academic. Miesch et al. from NOAA noted that the sun cycle forecast hasn't been updated since 2019. They are developing a new forecast (Miesch et al. 2023), emphasizing the need for modeling techniques. Our project combines insights from academic articles with interactive and engaging visualizations to create machine-learning models that classify and predict solar flares.

Who Cares?

Our project can help anticipate and prepare for solar events that can cause critical disruption to technology, as explained by Tom Berger, Director of NOAA (Berger et al., 2015), but also preventing human harm to astronauts or exposing flight passengers to radiation, through predictive models that can help forecast and prepare accordingly for these events. Additionally, we will provide more compelling and attractive visualization tools to query and consult solar flare information, by enhancing instruments currently available or expanding upon existing ones (Nishizuka et al., 2021), providing amateur astronomers, students, teachers, and citizen scientists with resources for exploring solar flare data.

What does success look like?

If successful, our models will measure the accuracy of predictions and classifications, as Juan Curto et al. explains accuracy keeps remaining as one of the challenges (Curto et al., 2020), by comparing our model results against prior observed events, and render such results in the visualization map of solar flare occurrences to easily show these forecasts.

Methodology, risks, and payoffs

Solar flares occurrences are highly correlated with repetitive solar cycles (that repeat every eleven years) and their intensities. Siami-Namini, Tavakili and Namin et al. study of LSTM [Long Short-Term Memory] presents evidence that they can outperform common time series models like ARIMA by a high margin (Siami-Namini et al., 2018). More specifically, LSTM have been proven to improve accuracy in solar flare predictions in prior studies (Wang et al., 2020; Liu et al., 2019), making them an ideal candidate for feature engineering in this project.

Traditionally, models primarily rely on the use of x-ray brightness measurements to classify the intensity of solar flares (Reep J. et al., 2019). While this approach is effective, relying solely on the x ray brightness maximum intensity disregards key solar information (Hong J. et al., 2023; Sinha, S. et al., 2022). We believe that incorporating these additional features can aid in classifying solar flares more accurately, while giving a granular prediction of solar flare events.

Our goal is to build a robust hybrid classification model that utilizes the LSTM for feature engineering and random forest for classification. The LTSM network will analyze time-series data to aid in the capture of temporal patterns and interactions between these solar phenomena, LSTM (and related methods like LSTM-A) have also been found in other studies to be very effective (Zheng, Y. et al., 2023). The outputs will then be fed into the random forest model for classification. Leveraging random forest will allow us to classify and predict solar flare events more accurately.

The random forest model will combine predictions from multiple decision trees (Boulesteix, A. et al., 2020), which will enable greater accuracy, and reduce overfitting/over classification. Within this ensemble approach, each decision tree will independently evaluate the different features of the solar data, and their output will be aggregated to produce the final result. A key advantage of this model is its ability assess and rank the importance of each features selected (Yuan X. et al., 2023). This will allow us to have deeper understanding of the factors influencing solar flare activities.

For the visualization component, we will build dashboards which consist of dynamic interactive plots that displays the importance of each feature. This will enable viewers to visually explore how different combinations of features influence prediction and classification of solar flares over time. We will also implement a 2D and 3D visualization of historical solar flare events that will allow users to filter flares by time period, class, and other criteria. Using D3, heliographic coordinates are converted into appropriate latitude and longitude markers on both the 2D map and the 3D rotatable sun model. Each flare is labeled with date, length of time, observed class of the flare, and predicted class from our classification model.

Resources Required

We plan to use free-tier cloud services for hosting and backend technologies (compute, storage, etc.). If costs exceed limits, we aim to keep expenses under $50-100 per month.

Timeline, midterm, and final checkpoints

Table 1: Table showing Plan of activity

Phase	Task	Description	Start Date	End Date	Assignee
Phase 1: Data Collection and Preprocessing	Data Collection	Connect to NASA and other space weather service api to retrieve data	9/30/2024	10/14/2024	Jordan & Sanjay
	Data Preprocessing	Handle of missing data/Normalization of features	10/7/2024	10/14/2024	Jordan & Sanjay
Phase 2: Model Development	Feature Engineering	Build LTSM network to analyze and derive new features for model	10/14/2024	10/21/2024	Jay & Dan
	Training/Test Data	Split data into training, test, and validation	10/21/2024	10/21/2024	Dan
	Build and Train Initial Random Forest Model	Train the model using the training data and the features identified	10/21/2024	10/28/2024	Jay & Sanjay
	Hyperparameter tuning	Using GridSearch/Randomized search to optimize parameters for the random forest	10/28/2024	10/31/2024	Chris
	Model Evaluation and Validation	Evaluate the model's accuracy, precision, recall, and other important criteria	10/31/2024	11/4/2024	Katherine & Sanjay
	Feature Analysis	Extract features from the model to determine which factors influence solar flare activity the most	11/4/2024	11/5/2024	Jordan & Dan
Phase 3: Visualization	Develop feature importance plot	Create dynamic interactive plots which showcases the ranking and influence of features from the model	11/5/2024	11/16/2024	Katherine
	Develop 2D/3D map	Visualization which shows historical solar flare events	11/5/2024	11/16/2024	Chris
	Interactive Dashboard Development	Tie all visualization together into a dashboard	11/5/2024	11/16/2024	Katherine & Jay
Deliverables	Final Report	This will be considered our final exam. All the task listed above should be completed	10/14/2024	11/17/2024	All
	Final Poster Presentation	Part of the final exam	11/1/2024	11/17/2024	Chris
	Progress Report	This will be considered our Midterm. We will complete optimization of parameters	10/14/2024	11/1/2024	All

All team members have contributed a similar amount of effort.

References

1. *Miesch, M. (2023, December 19). NOAA forecasts quicker, stronger peak solar activity. Space Weather Prediction Center.* https://www.swpc.noaa.gov/news/noaa-forecasts-quicker-stronger-peak-solar-activity
2. *Berger, T. (2015). Protecting Earth from solar storms.* https://eos.org/articles/protecting-earth-from-solar-storms

3. *Curto, J. J. (2020). Geomagnetic solar effects: A review. Journal of Solar Physics.* *https://doi.org/10.1051/swsc/2020027*
4. *Siami-Namini, S., Tavakoli, N. and Namin, A.S. (2018) A Comparison of ARIMA and LSTM in Forecasting Time Series. 2018 17th IEEE International Conference on Machine Learning and Applications, Orlando, 17-20 December 2018, 1394-1401.* *https://doi.org/10.1109/ICMLA.2018.00227*
5. *Wang, X., Chen, Y., Toth, G., Manchester, W. B., Gombosi, T. I., Hero, A. O., Jiao, Z., Sun, H., Jin, M., & Liu, Y. (2020). Predicting solar flares with machine learning: Investigating solar cycle dependence. The Astrophysical Journal, 1538(4357), Article ab89ac.* *https://doi.org/10.3847/1538-4357/ab89ac*
6. *Liu, H., Liu, C., Wang, J. T. L., & Wang, H. (2019). Predicting solar flares using a long short-term memory network. The Astrophysical Journal, 877(121), 1-14.* *https://doi.org/10.3847/1538-4357/ab1b3c*
7. *Kumar, Pramod & Bhatt, Yogesh & Jain, Rajmal. (2015). Solar flare and its interaction with the Earth atmosphere: An Introduction. 1.*
8. *Hong, J., Ji, A., Pandey, C., Aydin, B. (2023). Beyond Traditional Flare Forecasting: A Data-driven Labeling Approach for High-fidelity Predictions. In: Wrembel, R., Gamper, J., Kotsis, G., Tjoa, A.M., Khalil, I. (eds) Big Data Analytics and Knowledge Discovery. DaWaK 2023. Lecture Notes in Computer Science, vol 14148. Springer, Cham.* *https://doi.org/10.1007/978-3-031-39831-5_34*
9. *Boulesteix, A. L., Janitza, S., Kruppa, J., & König, I. R. (2020). The random forest algorithm for statistical learning. The American Statistician, 74(4), 332-345.* *https://doi.org/10.1177/1536867X20909688*
10. *Yuan, X.; Liu, S.; Feng, W.; Dauphin, G. Feature Importance Ranking of Random Forest-Based End-to-End Learning Algorithm. Remote Sens. 2023, 15, 5203.* *https://doi.org/10.3390/rs15215203*
11. *Liu, J., Wang, W., Qian, L., Lotko, W., Burns, A. G., Pham, K., Lu, G., Solomon, S. C., Liu, L., Wan, W., Anderson, B. J., Coster, A., & Wilder, F. (2021). Solar flare effects in the Earth's magnetosphere. Nature Physics, 17(7), 807–812.* https://doi.org/10.1038/s41567-021-01203-5
12. *Hayes, L. A., O'Hara, O. S. D., Murray, S. A., & Gallagher, P. T. (2021). Solar Flare Effects on the Earth's Lower Ionosphere. Solar Physics, 296(11).* https://doi.org/10.1007/s11207-021-01898-y
13. *Reep, J. W., & Knizhnik, K. J. (2019). What determines the xray intensity and duration of a solar flare? The Astrophysical Journal, 874(2), 174.* ***https://doi.org/10.3847/1538-4357/ab0ae7***
14. *López-Urias, C., Vazquez-Becerra, G. E., Nayak, K., & López-Montes, R. (2023). Analysis of Ionospheric Disturbances during X-Class Solar Flares (2021–2022) Using GNSS Data and Wavelet Analysis. Remote Sensing (Basel, Switzerland), 15(18), 4626-.* https://doi.org/10.3390/rs15184626
15. *Fagundes, P. R., Pillat, V. G., Tardelli, A., & Muella, M. T. A. H. (2024). Pole-To-Pole Ionospheric Disturbances Due To Solar Flares, During Low Solar Activity. Journal of Geophysical Research. Space Physics, 129(7).* https://doi.org/10.1029/2024JA032597
16. *Nishizuka, N., Kubo, Y., Sugiura, K., Den, M., & Ishii, M. (2021). Operational solar flare prediction model using Deep Flare Net. Earth, Planets and Space, 73(1), 64.* https://doi.org/10.1186/s40623-021-01381-91
17. *Zheng, Y., Qin, W., Li, X., Ling, Y., Huang, X., Li, X., Yan, P., & Yan, S. (2023). Comparative analysis of machine learning models for solar flare prediction. Astrophysics and Space Science, 368(1), 53.* https://doi.org/10.1007/s10509-023-04209-y

18. *Sinha, S., Gupta, O., Singh, V., Lekshmi, B., Nandy, D., Mitra, D., Chatterjee, S., Bhattacharya, S., Chatterjee, S., Srivastava, N., Brandenburg, A., & Pal, S. (2022). A comparative analysis of machine-learning models for solar flare forecasting: Identifying high-performing active region flare indicators. The Astrophysical Journal, 935(1), 7.* https://doi.org/10.3847/1538-4357/ac7955