Combining AI and Cyberinfrastructure to Understand Solar Astronomy and Forecast Space Weather

Jason Wang, Ph.D., Computer Science

Joint Work with Prof. Haimin Wang, Physics, BBSO and Space Weather

Work supported by NSF and NASA grants

Outline

- Solar Data Science: An Emerging Interdisciplinary Field
- AI/ML for Solar Data Analysis
- AI/ML for Space Weather Forecasting
- Roles of Generative AI (GenAI) and Explainable AI (XAI)
- Big Data-Enabled AI-Powered Community-Coordinated Cyberinfrastructure
- Conclusion

Solar Data Science

Solar Data Science = Using Artificial Intelligence (AI), Machine Learning (ML), Cyberinfrastructure (CI) and Big Data to Understand Solar Astronomy and Forecast Space Weather

Solar data science provides rich datasets with high complexity

AI/ML can learn patterns, relationships, associations from historical data and make predictions on unseen future testing data (we are mainly interested in end-to-end deep learning)

Challenges for AI/ML in Solar Data Science

Although solar astronomy and space weather are filled with big data, we often lack **labeled training data (ground truth data with labels)** that can be used to train an AI/ML model. This is different from AI/ML in computer science in which there are a lot of labeled training data (e.g., ImageNet).

Our approach: adopt the expert-in-the-loop paradigm and perform AI/ML via data integration

Why AI/ML?

We already have computational tools (e.g., to detect fibrils), why is AI/ML needed?

Two reasons:

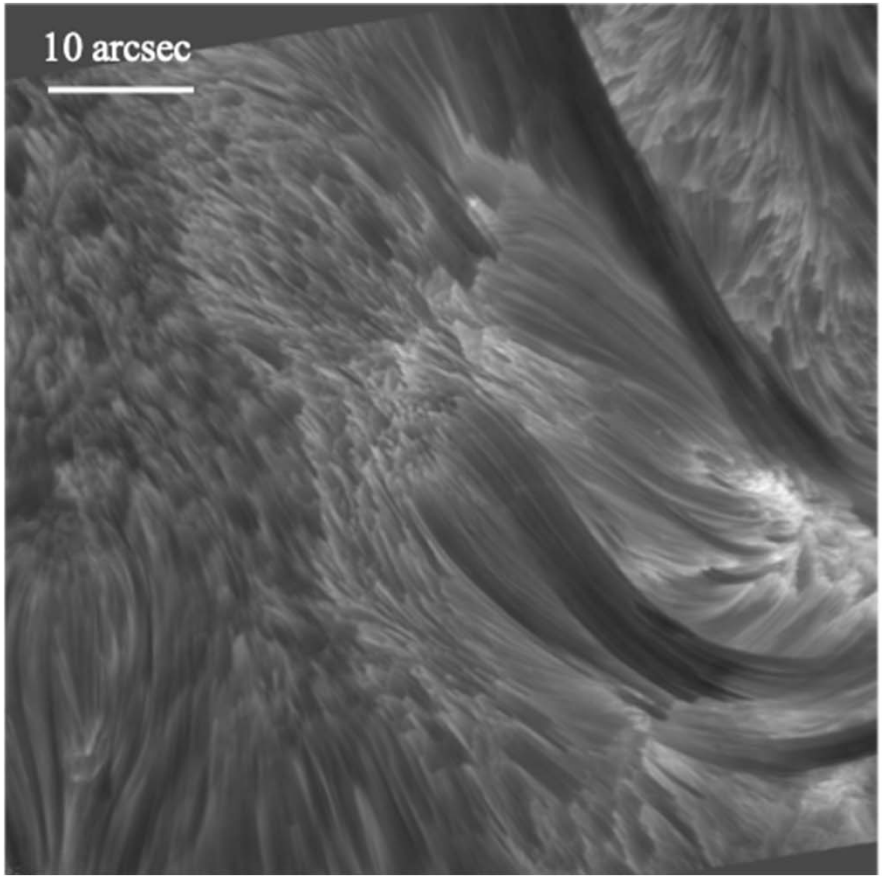
1. AI/ML detects fibrils through making predictions, and hence is often faster than computational tools which involve time-consuming computation. We aim at building near real-time operational and forecasting systems, so AI/ML better meets our need.
2. AI/ML models have inference and generalization capabilities and often produce better results than computational tools.

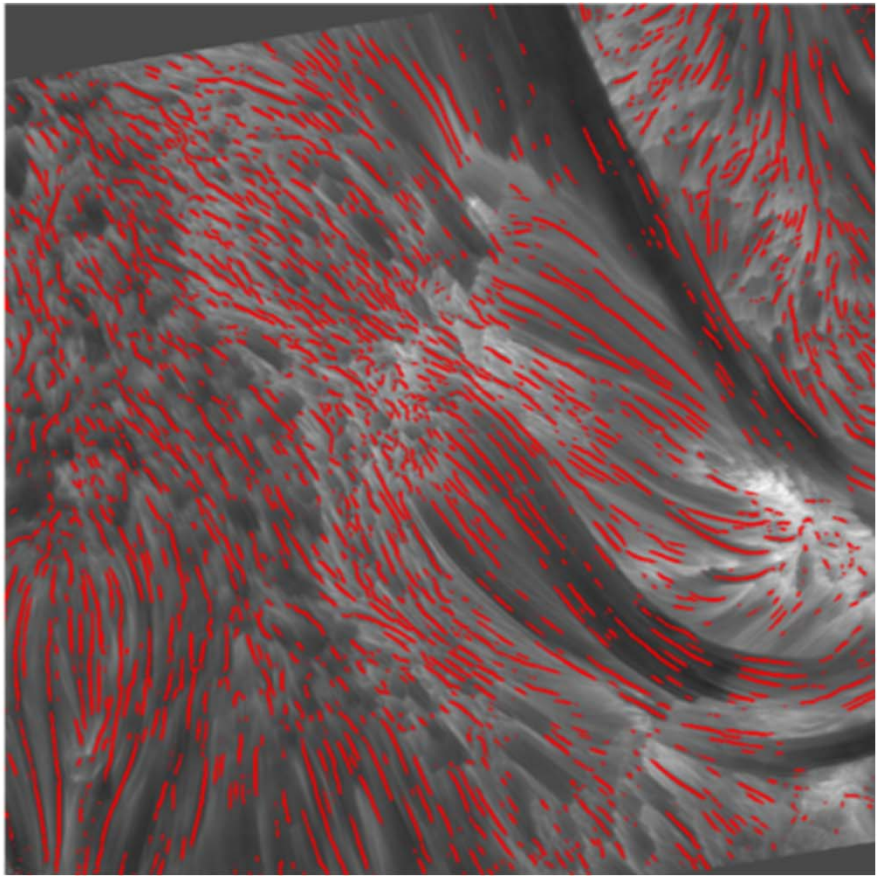
Outline

- Solar Data Science: An Emerging Interdisciplinary Field
- AI/ML for Solar Data Analysis
- AI/ML for Space Weather Forecasting
- Roles of Generative AI (GenAI) and Explainable AI (XAI)
- Big Data-Enabled AI-Powered Community-Coordinated Cyberinfrastructure
- Conclusion

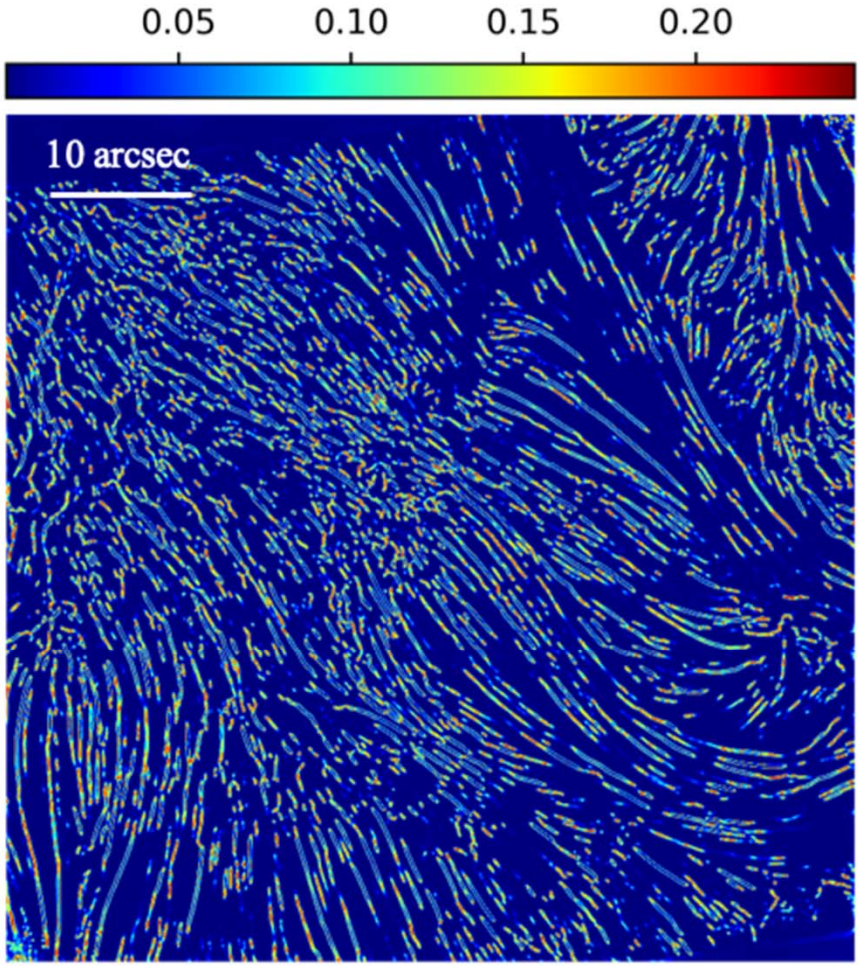
AI/ML for Fibril Detection

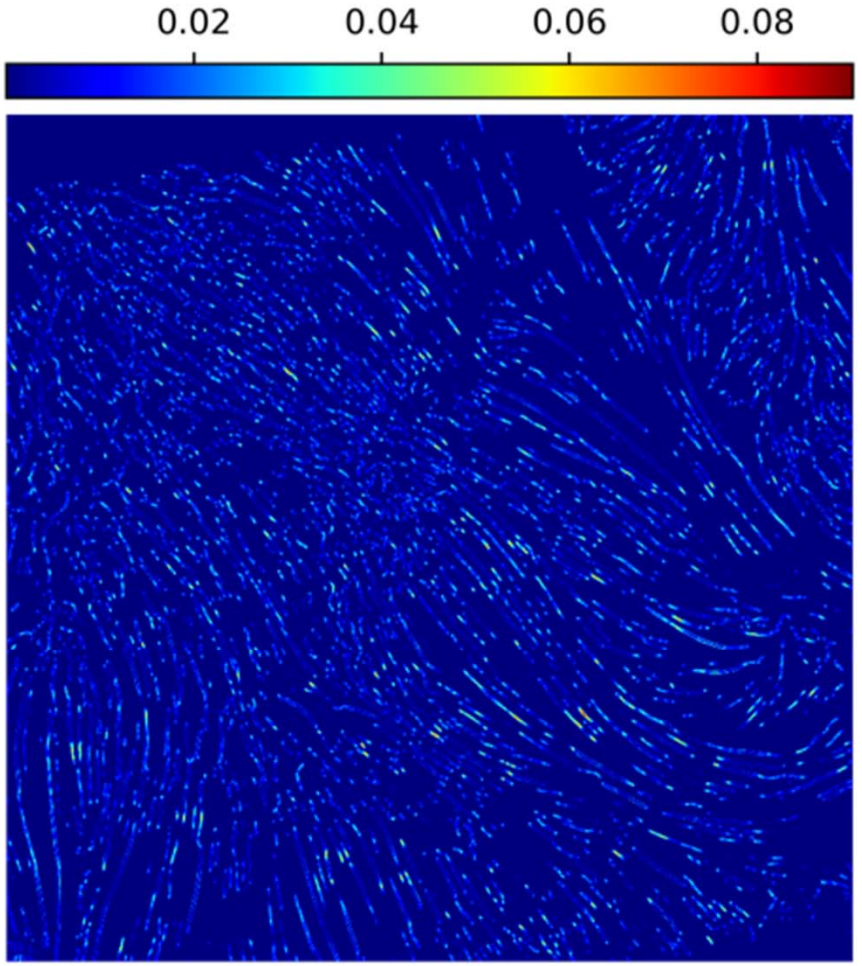
[Jiang et al. ApJS 2021]



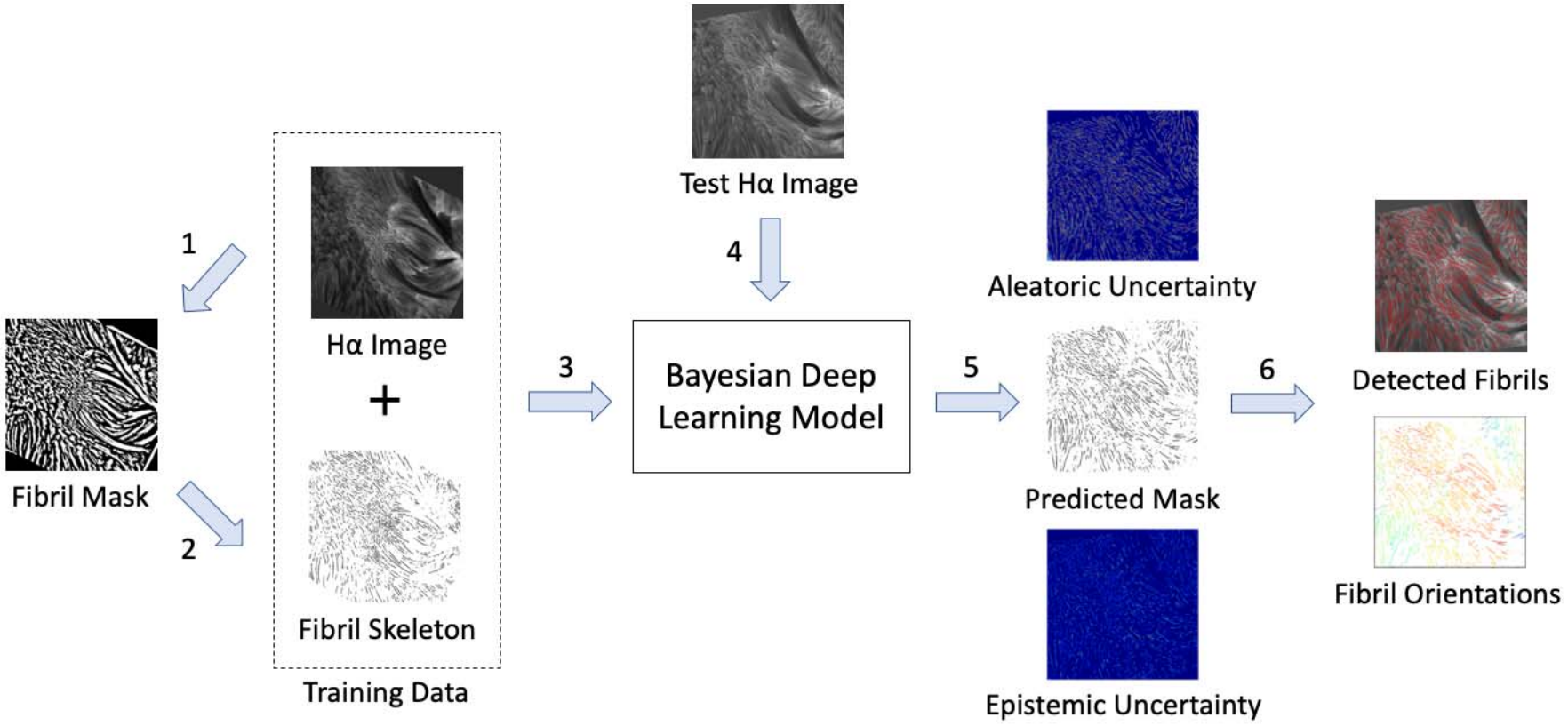


AI/ML for Fibril Detection

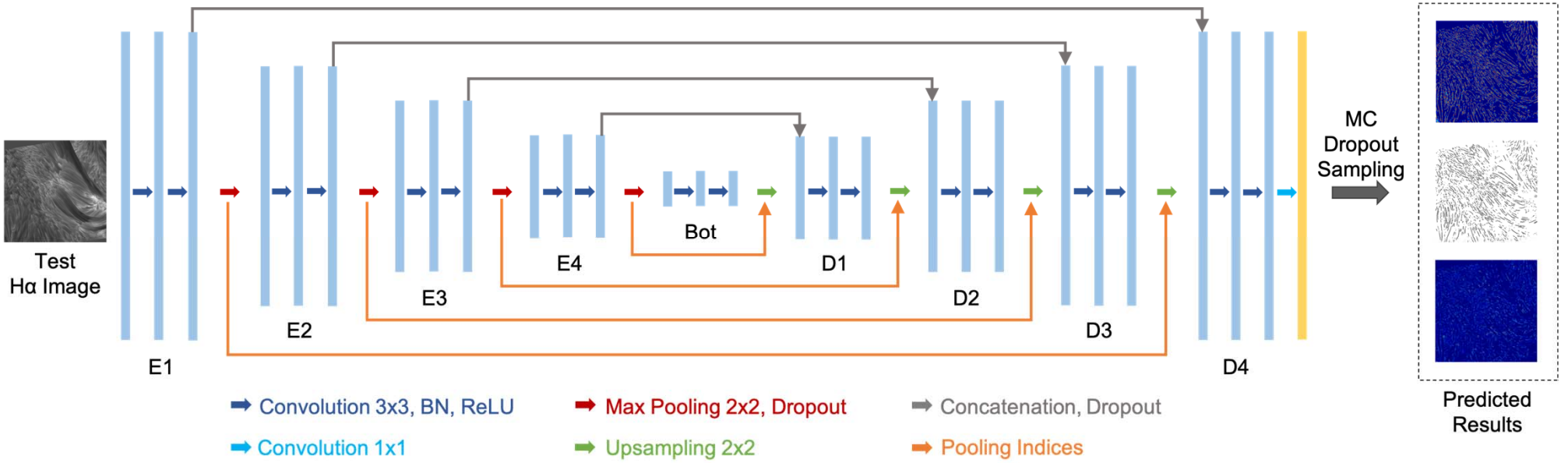




AI/ML for Fibril Detection

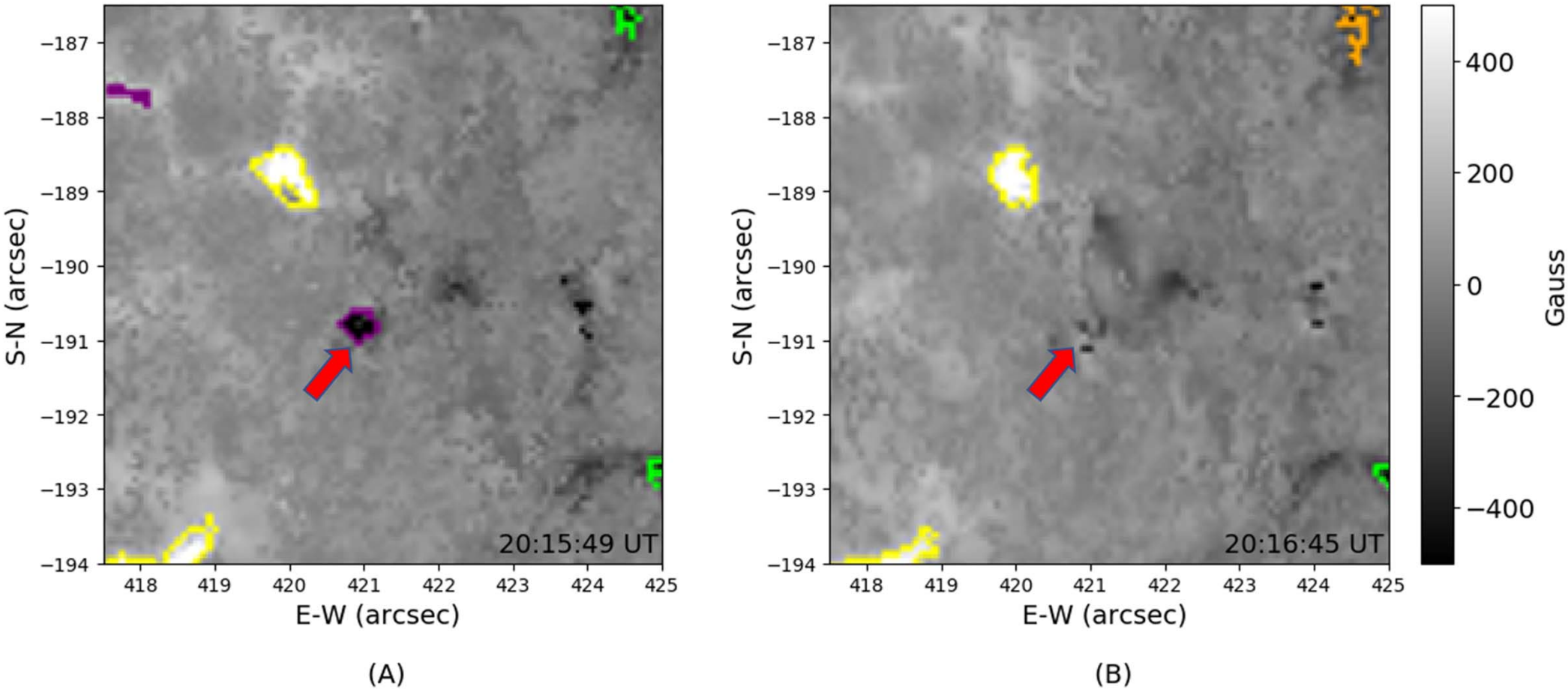


Bayesian Deep Learning Model for Fibril Detection



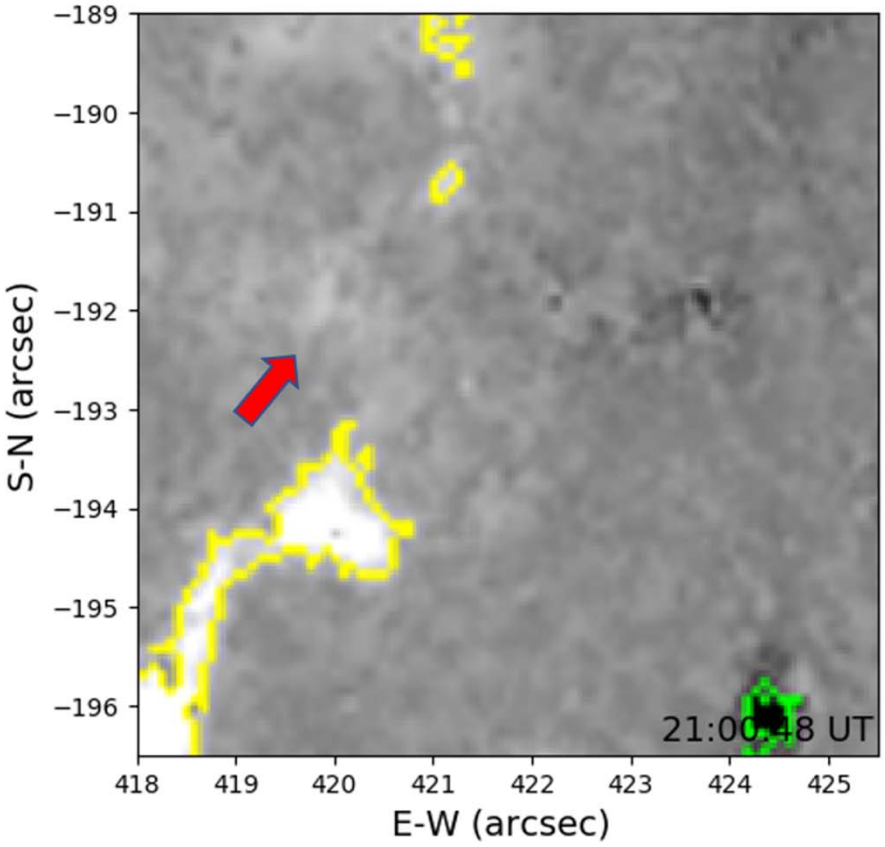
AI/ML for Magnetic Flux Tracking

flux element disappearance [Jiang et al. ApJS 2020]

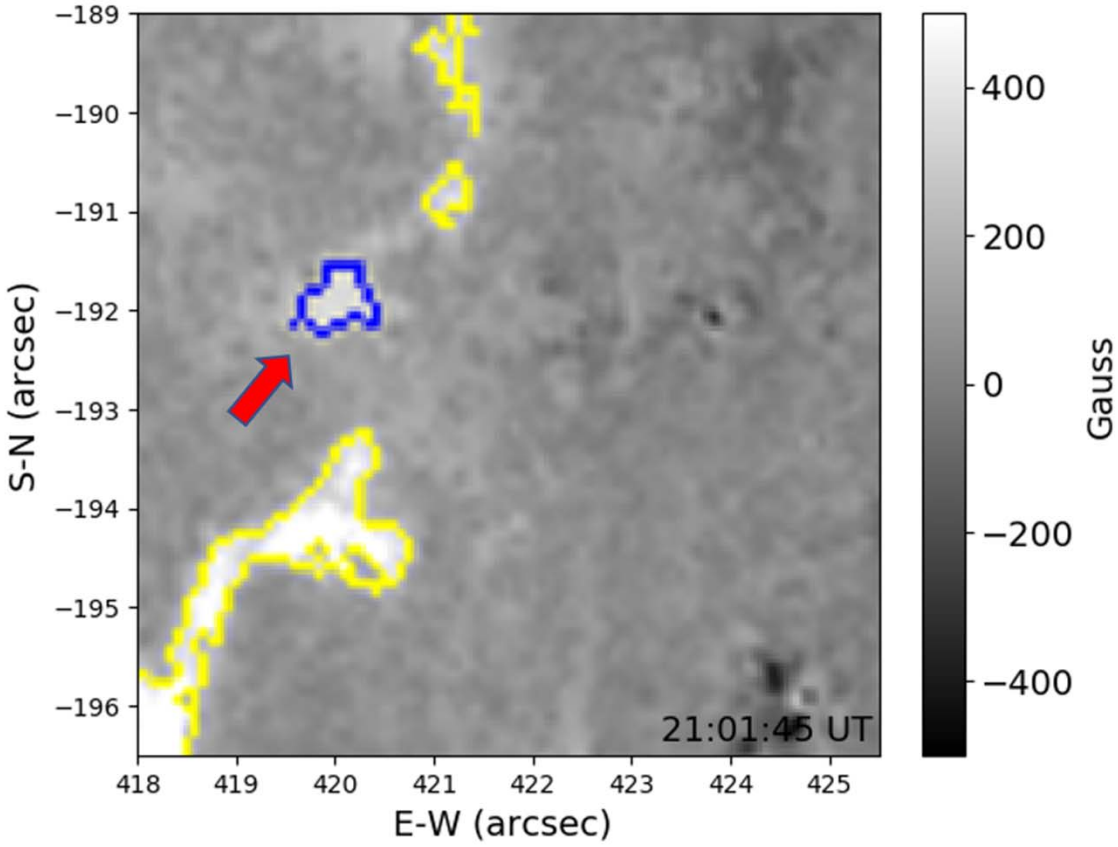


AI/ML for Magnetic Flux Tracking

flux element appearance

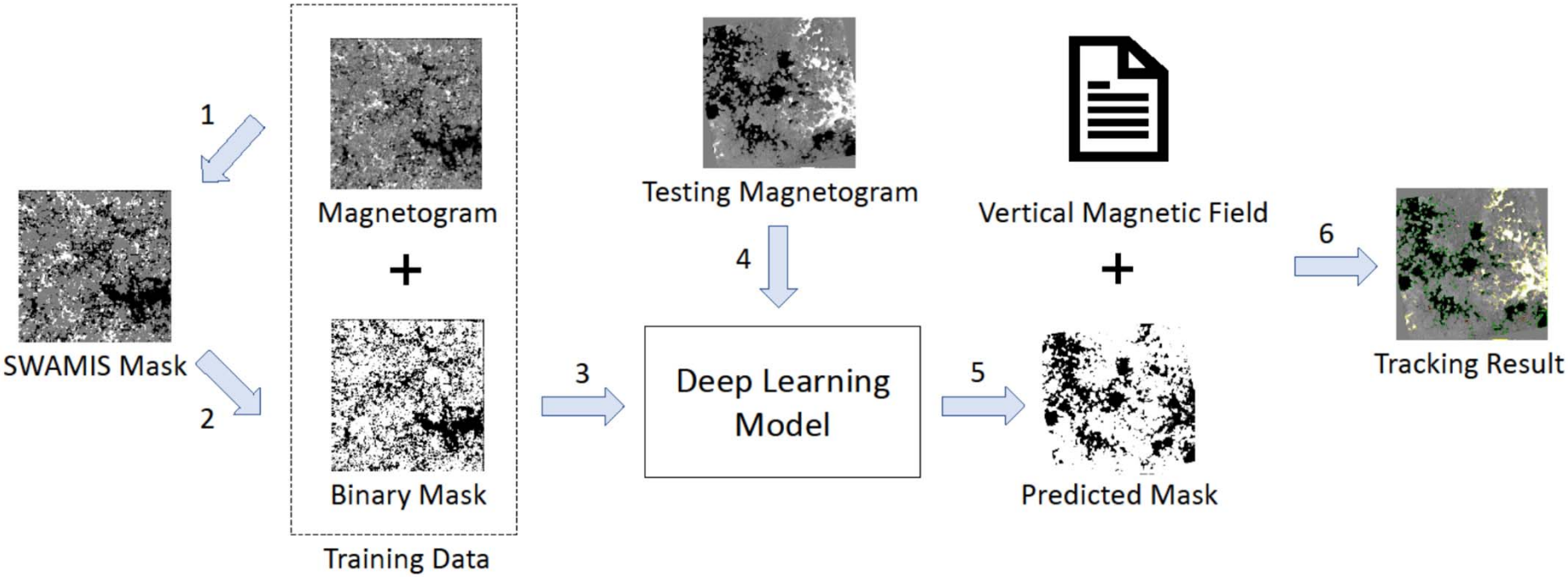


(A)

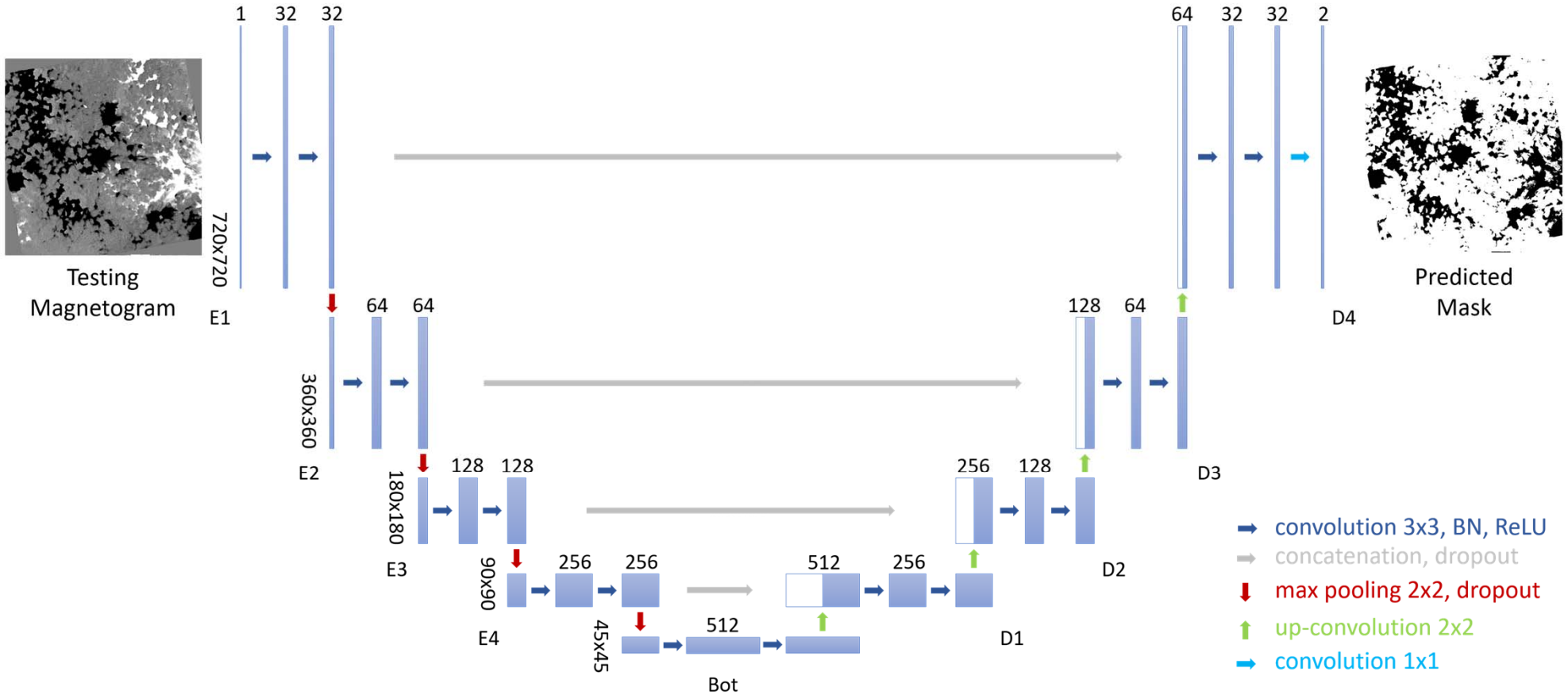


(B)

AI/ML for Magnetic Flux Tracking

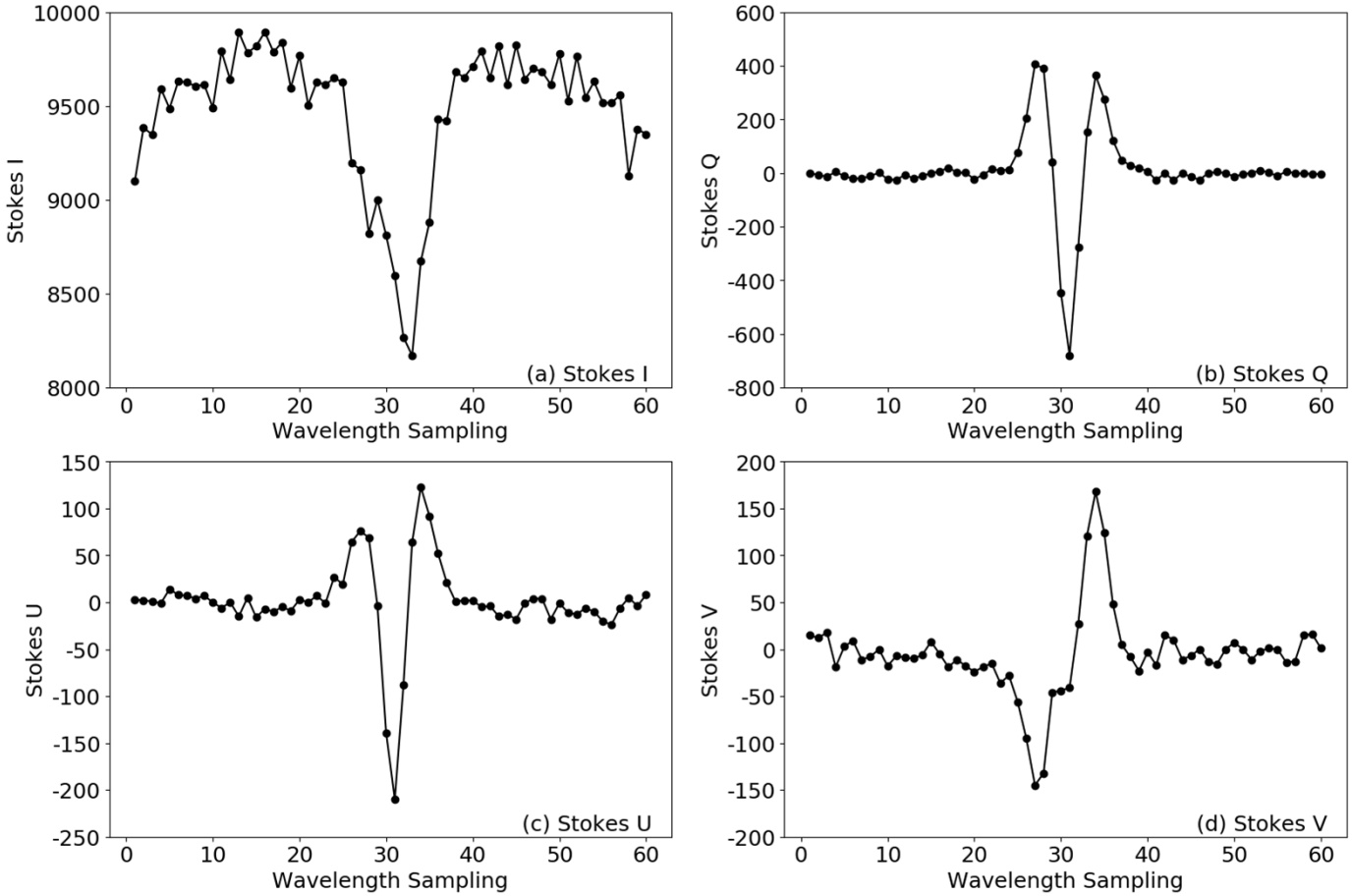


Deep Learning Model for Magnetic Flux Tracking

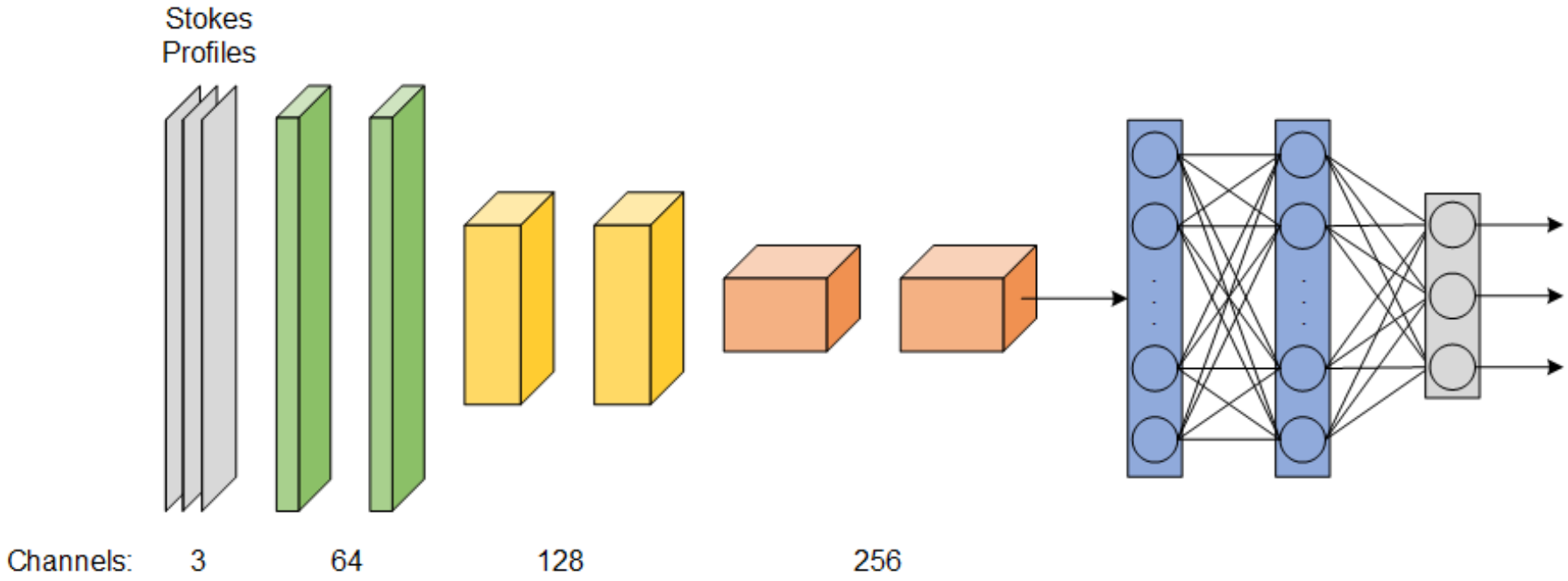


AI/ML for Stokes Inversion

[Liu et al. ApJ 2020, Jiang et al. 2022]

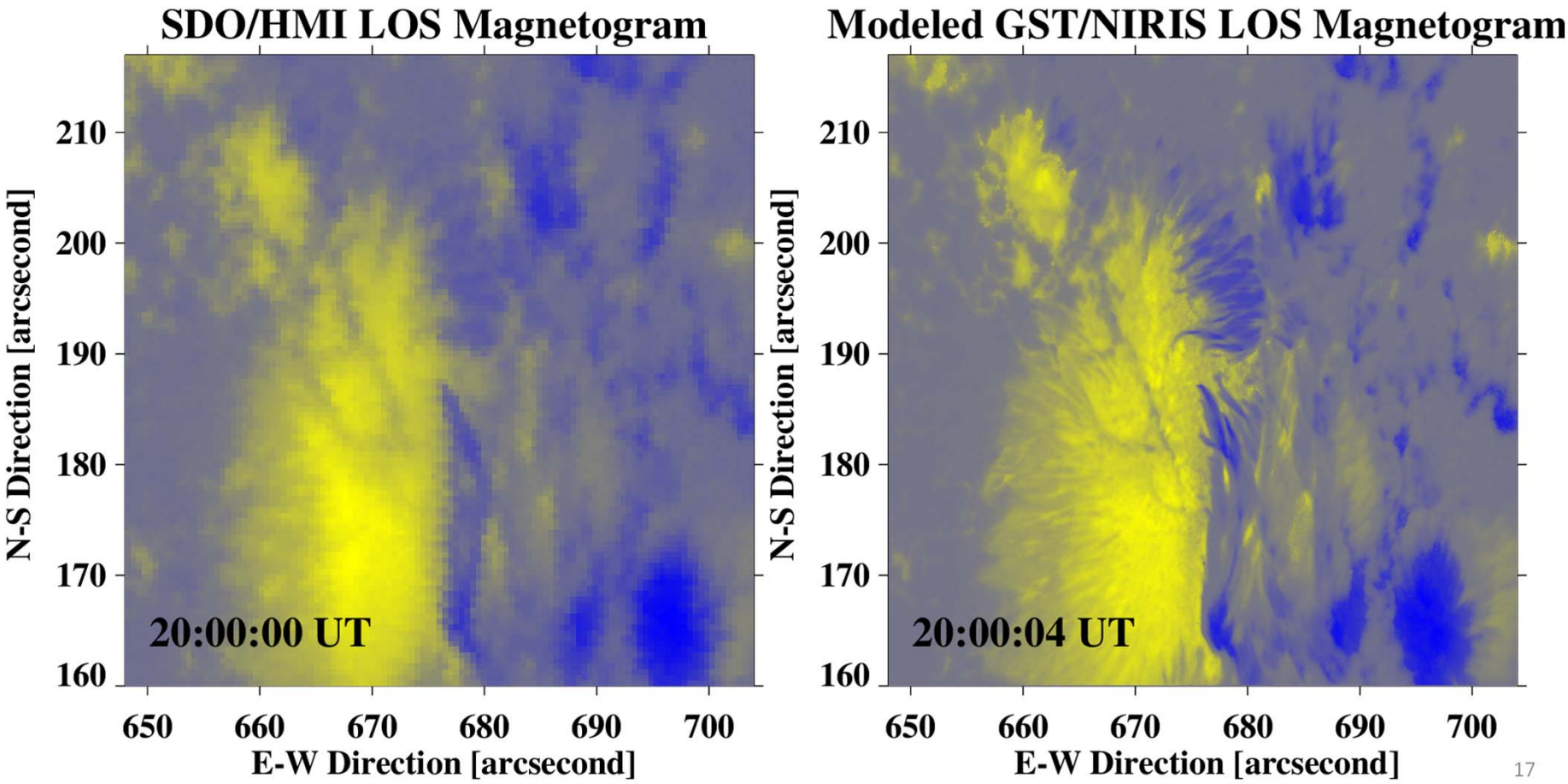


AI/ML for Stokes Inversion



Q, U, V (calibrated by I) are used as input.

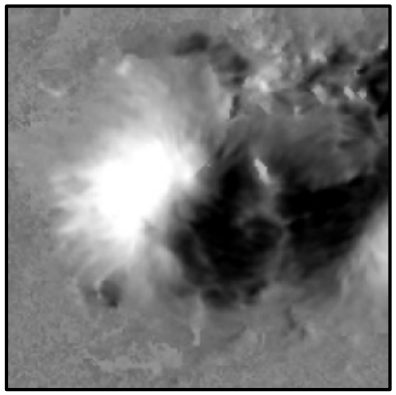
AI/ML for Stokes Inversion

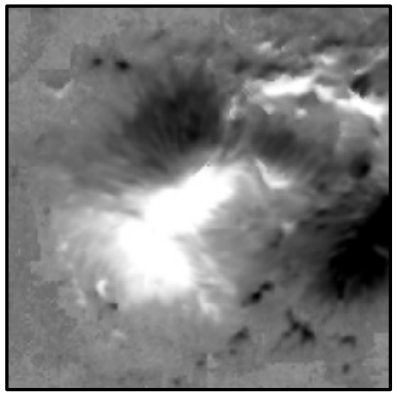


AI/ML for Solar Vector Magnetogram Generation

[Jiang et al. Solar Physics 2023]

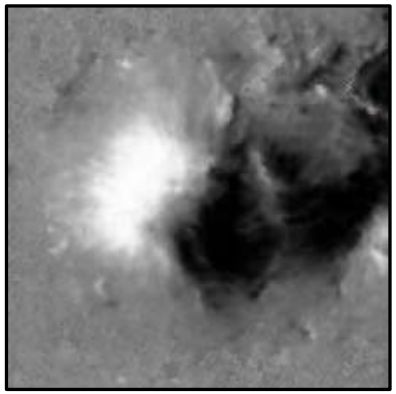
Original Bx

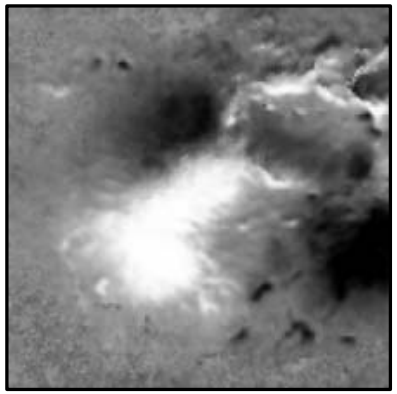




Original By

AI-Generated Bx



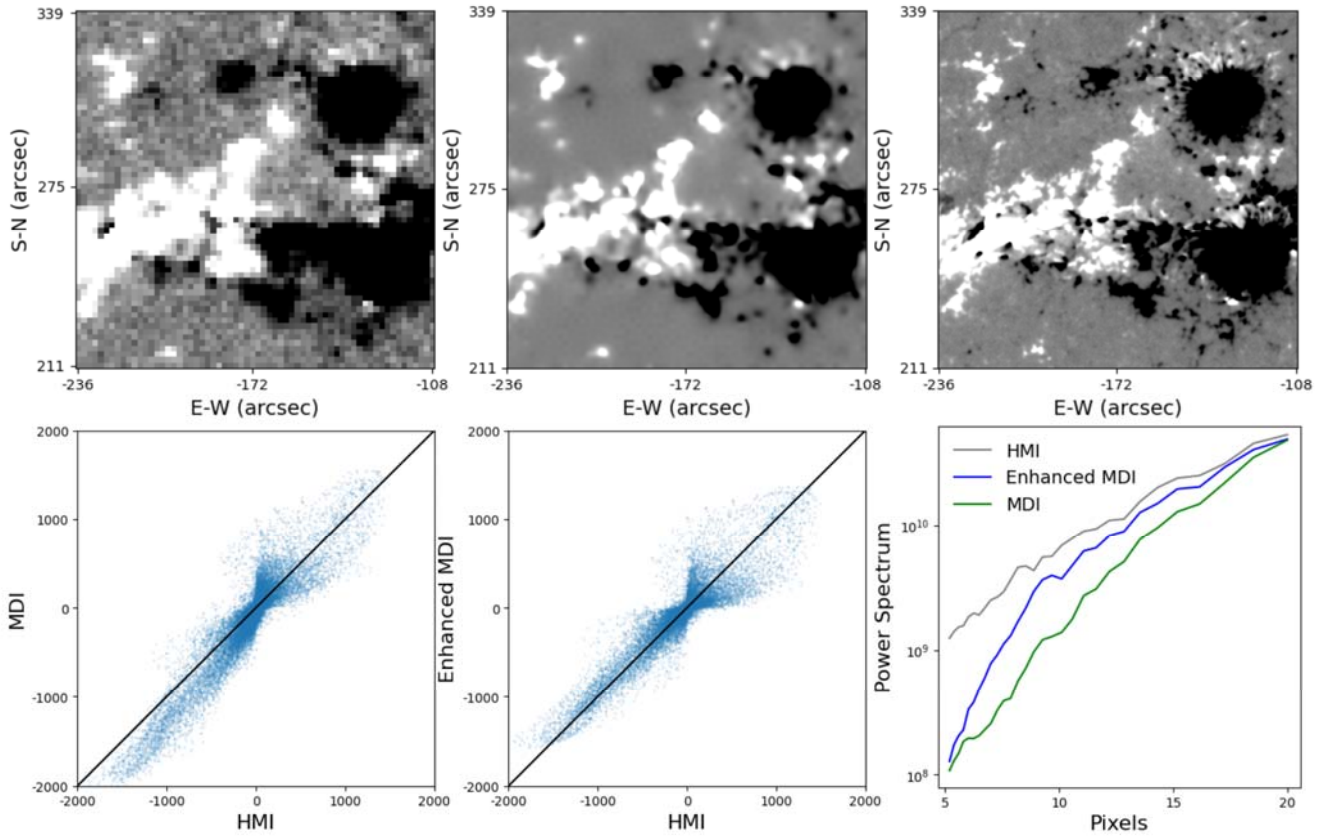


AI-Generated By

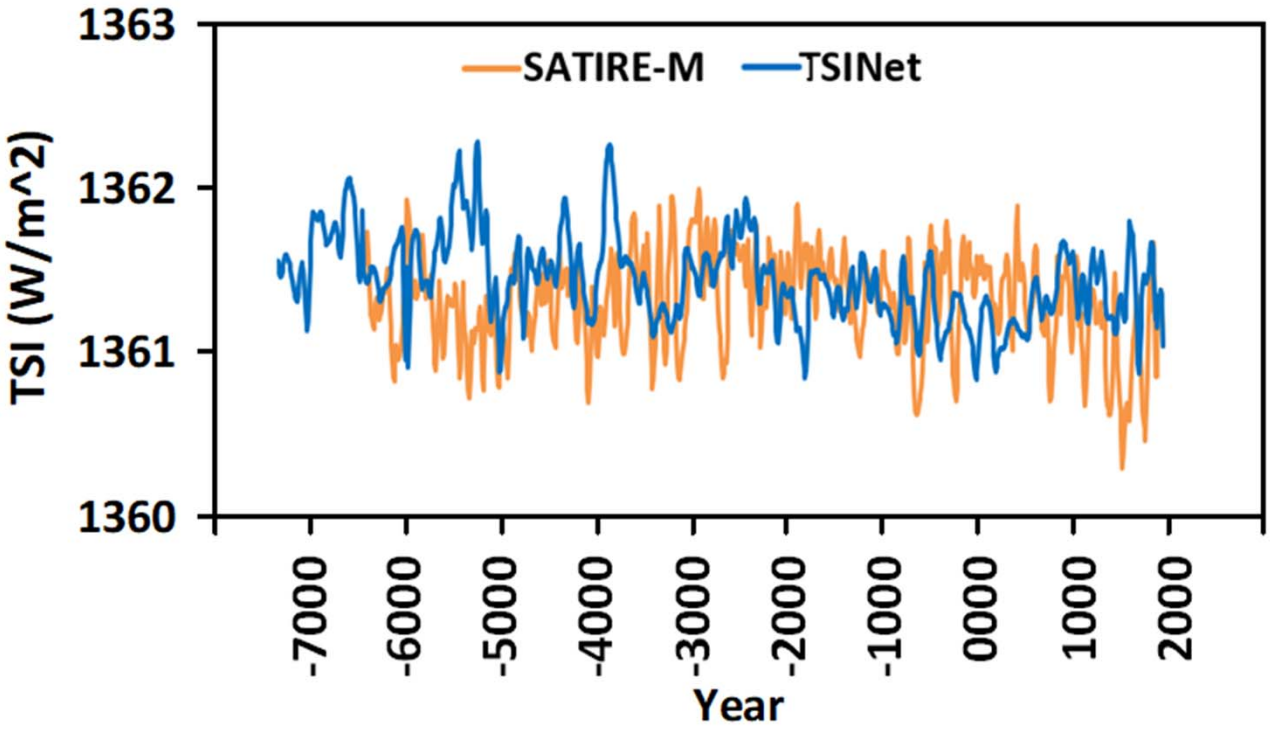
Use AI-Generated Transverse Components (Bx, By) of Magnetic Field to Study Solar Flare Activity in Multiple Solar Cycles

AI/ML for Improving Spatial Resolution of SOHO/MDI Magnetograms

[Xu et al. Solar Physics 2024]

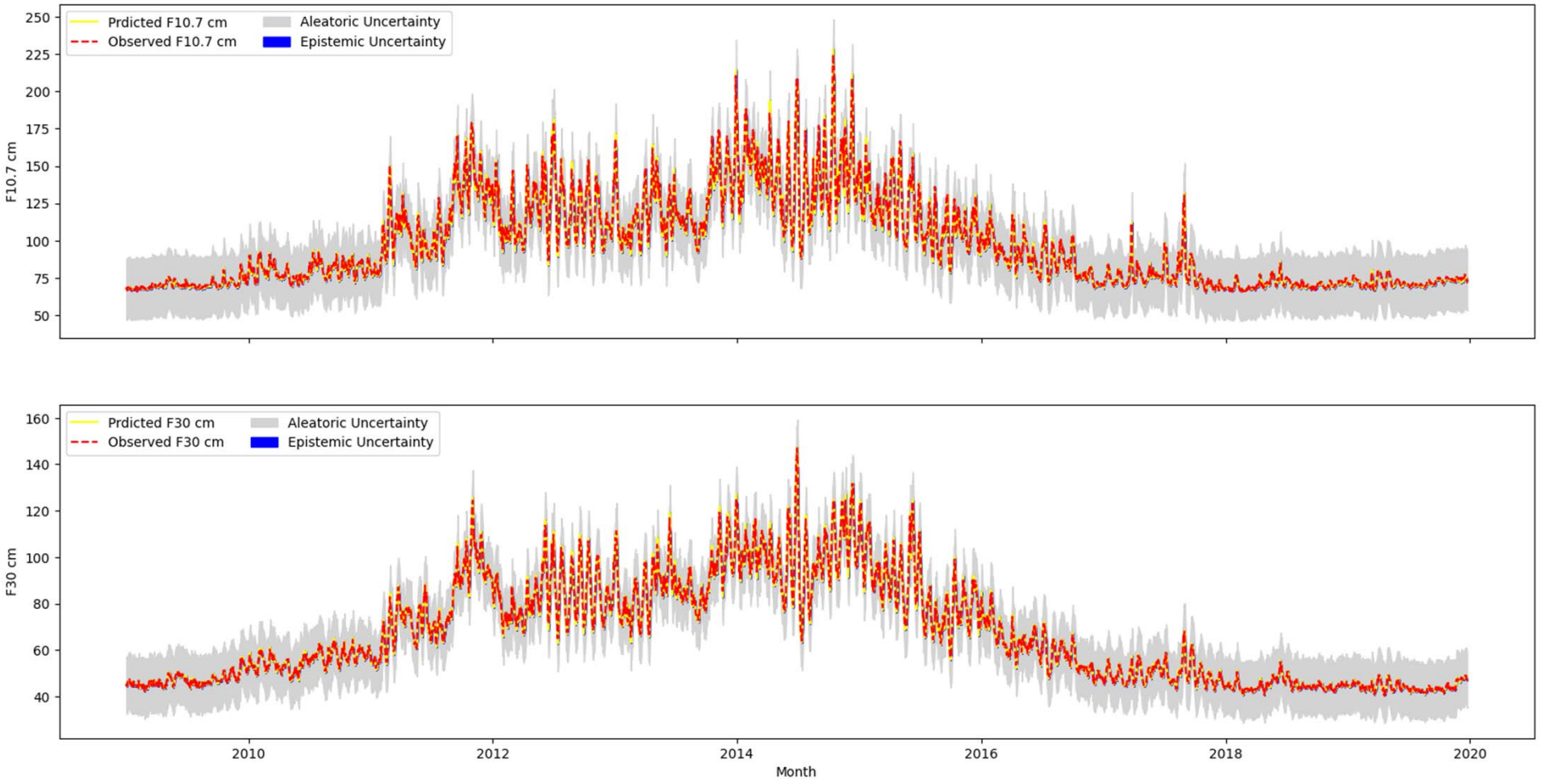


AI/ML for Solar Irradiance Reconstruction



SATIRE-M (Spectral And Total Irradiance REconstruction model - Millennia), based on sunspot numbers, provides decennial averages and reconstructs the total solar irradiance (TSI) over the last 9,000 years. Our TSINet reconstructs TSI for additional 1,000 years. Currently we work on solar EUV reconstruction (SOHO/SEM, TIMED/SEE, SDO/EVE).

AI/ML for Predicting Solar (Activity) Indices F10.7 and F30 (Proxies of Solar EUV Irradiance)



Outline

- Solar Data Science: An Emerging Interdisciplinary Field
- AI/ML for Solar Data Analysis
- AI/ML for Space Weather Forecasting
- Roles of Generative AI (GenAI) and Explainable AI (XAI)
- Big Data-Enabled AI-Powered Community-Coordinated Cyberinfrastructure
- Conclusion

Space Weather Events: Solar Flares

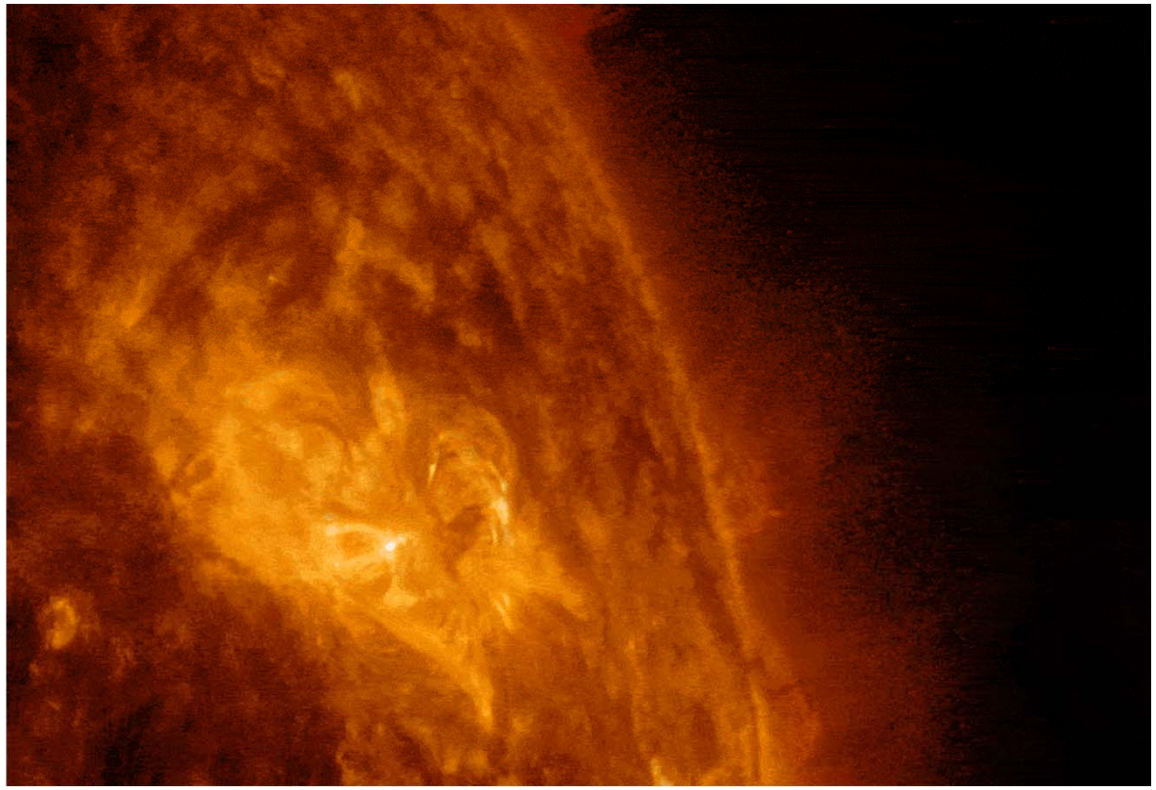
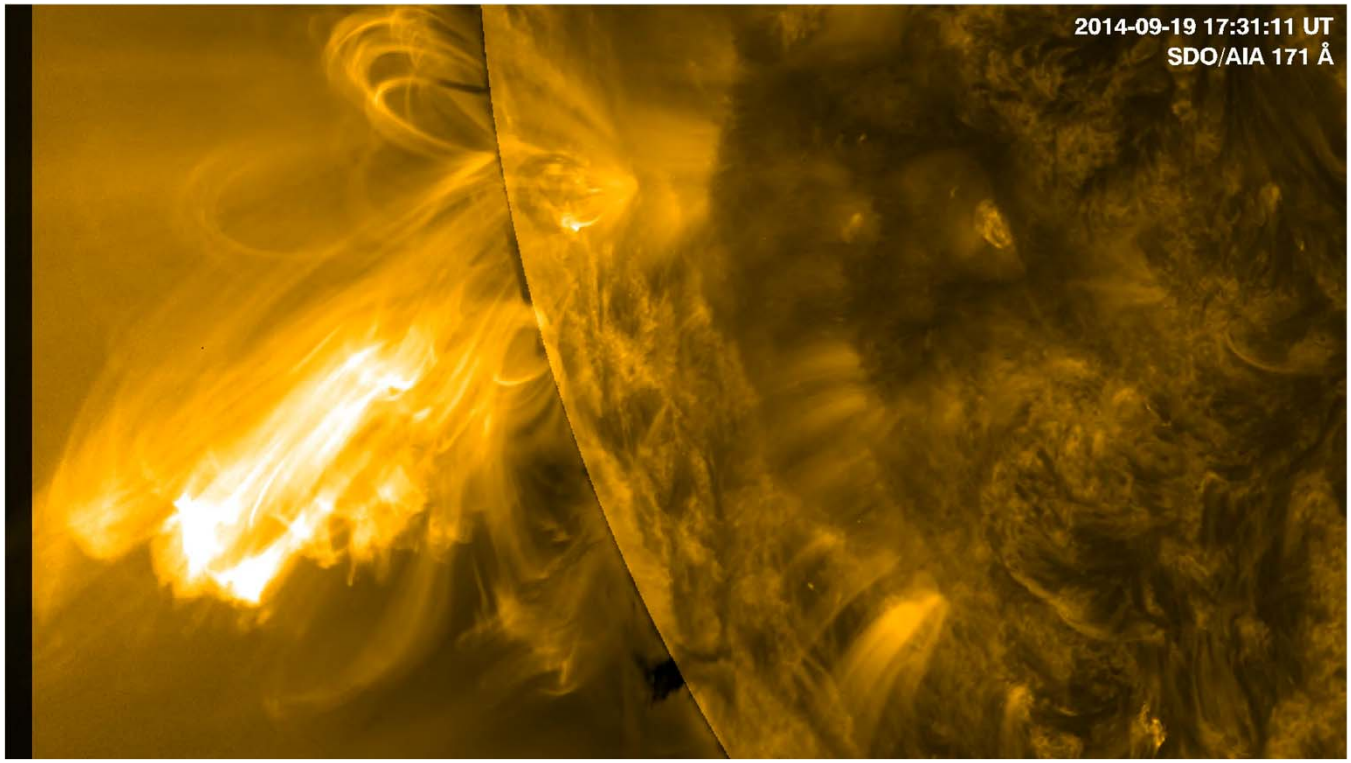


Image credit: NASA/SDO/Goddard
https://spaceplace.nasa.gov/solar-activity/en/

Space Weather Events: Coronal Mass Ejections (CMEs)

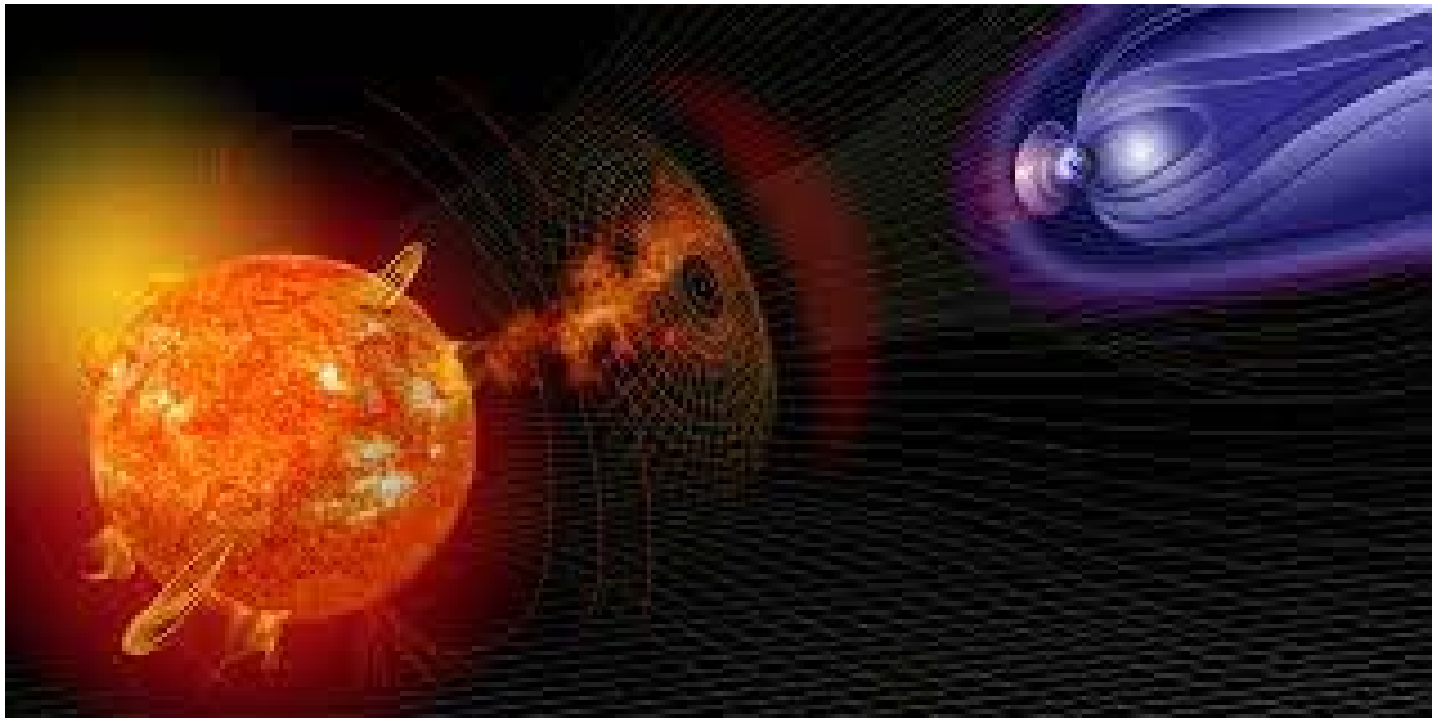
Filament Eruption
Leading to a CME



Observed
CME arrival time:
18 hr – 138 hr

Source: AIA Filament Eruption Catalog

Space Weather Events: Solar Energetic Particles (SEPs)



SEP arrival time:
<= 1 hr
Events last for
days-weeks

Source: NASA

AI/ML for Predicting Solar Eruptions, Space Weather Events, and Geomagnetic Activity

- Predicting Solar Flares Using SDO/HMI Vector Magnetic Data Products & Random Forest Algorithm [Liu et al. ApJ 2017]
- Predicting Solar Flares Using a Long Short-term Memory Network [Liu et al. ApJ 2019]
- Predicting Coronal Mass Ejections Using SDO/HMI Vector Magnetic Data Products and Recurrent Neural Networks [Liu et al. ApJ 2020]
- Predicting Solar Energetic Particles Using SDO/HMI Vector Magnetic Data Products and a Bidirectional LSTM Network [Abduallah et al. ApJS 2022]
- Solar Flare Index Prediction Using SDO/HMI Vector Magnetic Data Products with Statistical and Machine-learning Methods [Zhang et al. ApJS 2022]
- Predicting CME Arrival Time through Data Integration and Ensemble Learning [Alobaid et al. Frontiers in Astronomy and Space Sciences 2022]
- Operational Prediction of Solar Flares Using a Transformer-Based Framework [Abduallah et al. Scientific Reports 2023]
- Estimating CME Mass and Kinetic Energy by Fusion of Multiple Deep-Learning Models [Alobaid et al. ApJL 2023]
- A Machine Learning Approach to Understanding the Physical Properties of Magnetic Flux Ropes in the Solar Wind at 1 au [Farooki et al. ApJ 2024]

Geomagnetic storms due to CMEs,
March 8, 2012, Iceland
Source: nasa.gov

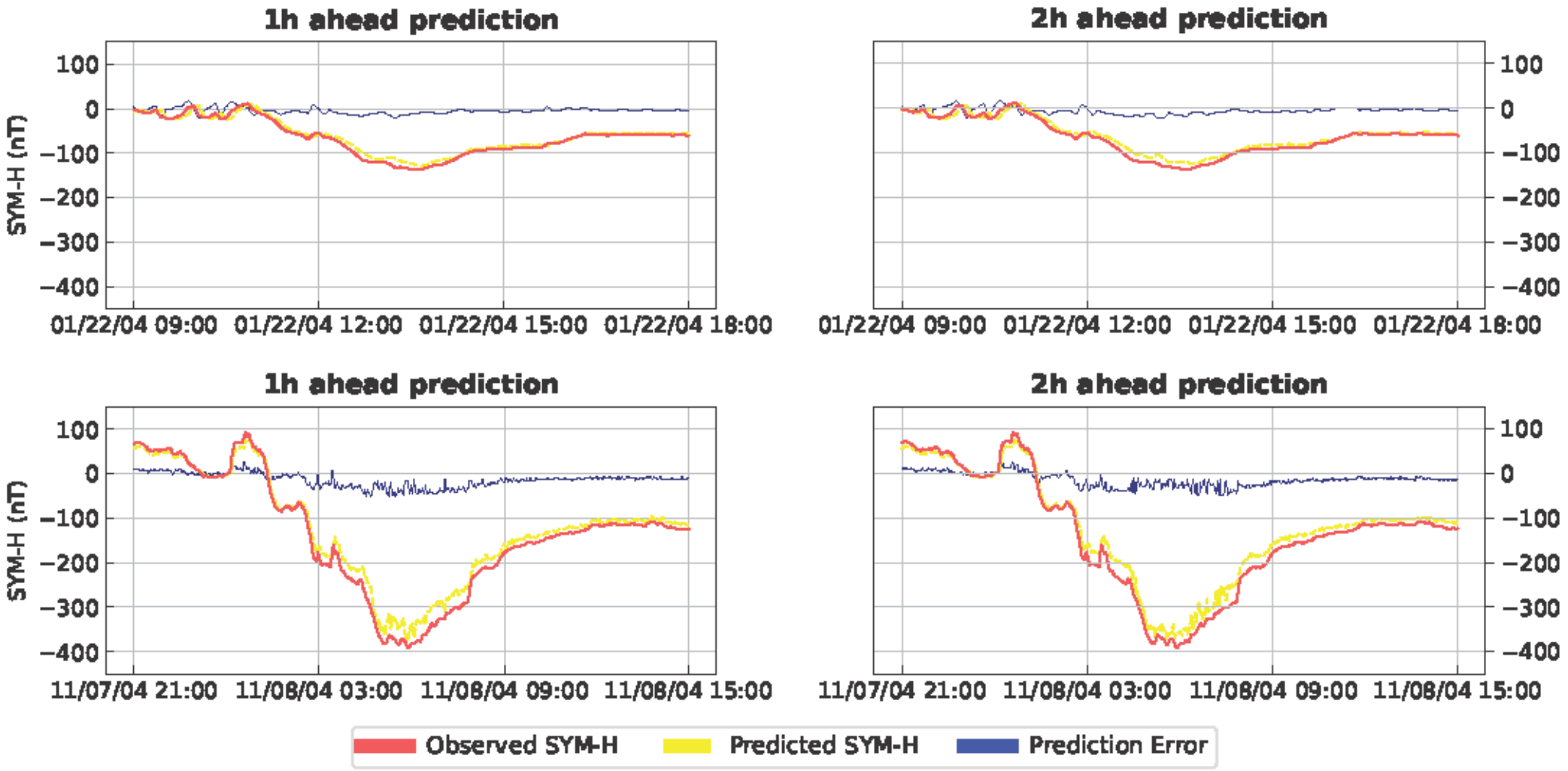
AI/ML for Geomagnetic Index Prediction

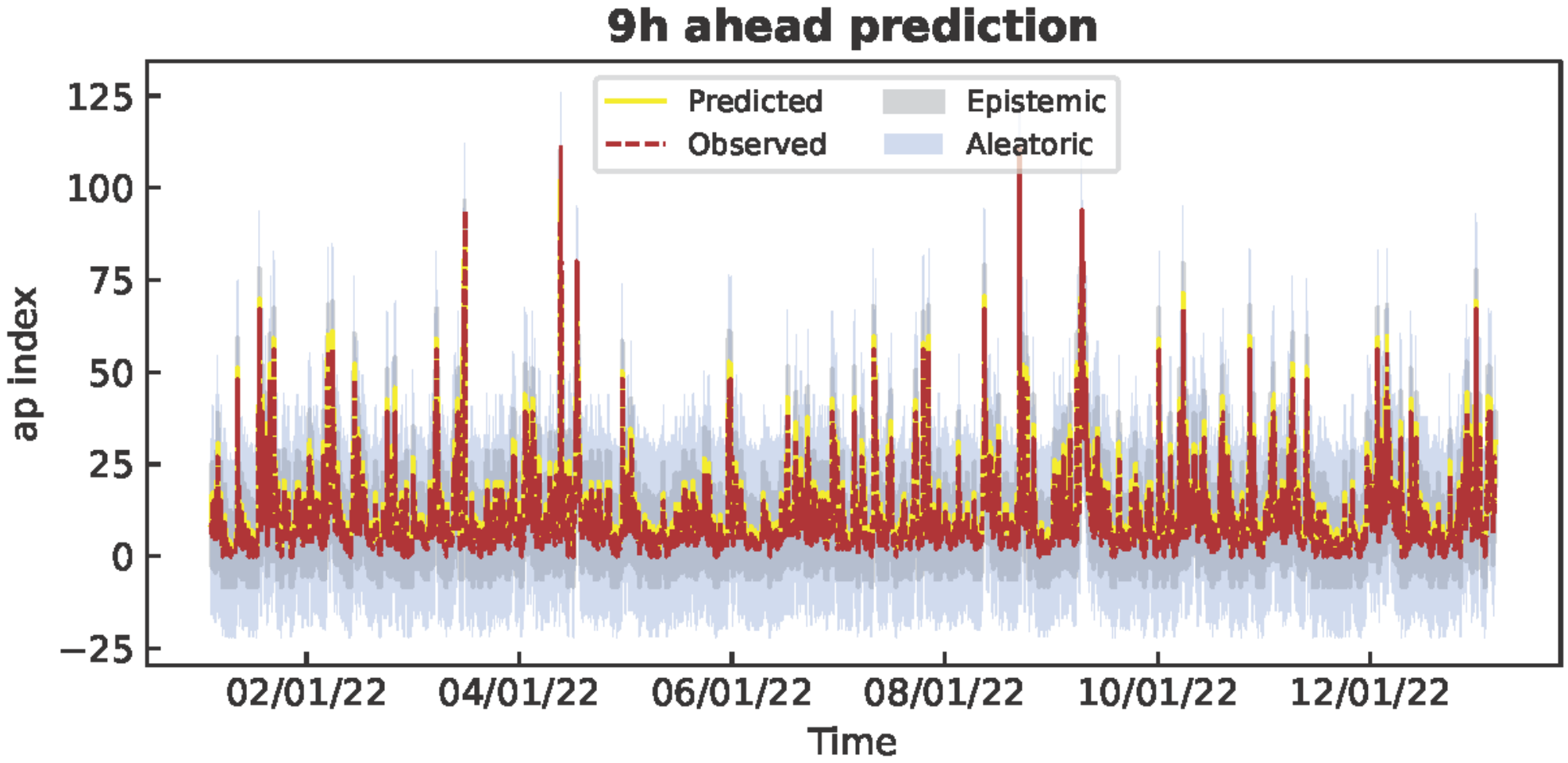
We use deep learning to predict Dst (disturbance storm time) index values focusing on intense storms (-250 nT < Dst < -50 nT) and super storms (Dst < -250 nT).

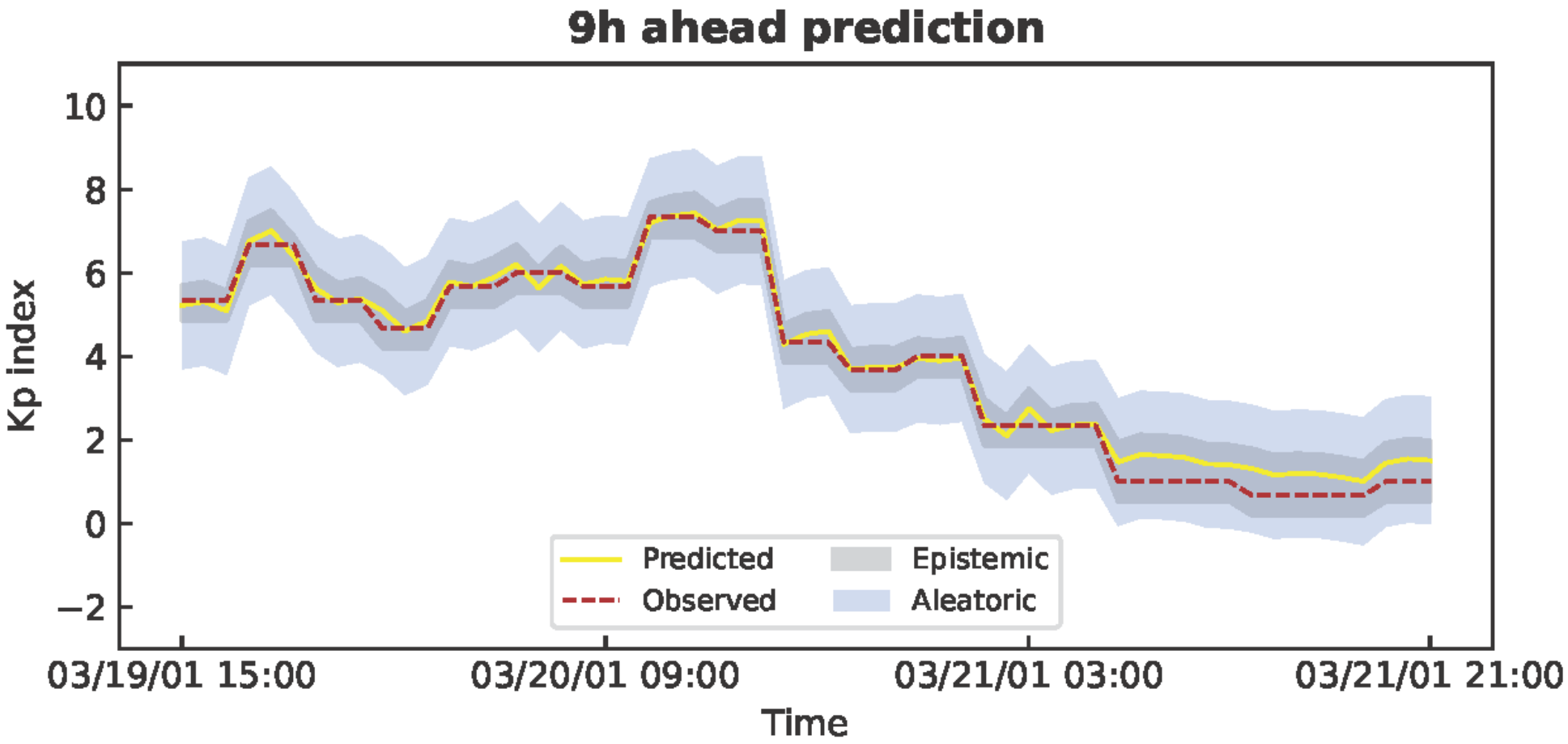
There are many geomagnetic indices: Dst, Kp, ap, Ap, SYM-H, etc.

Use deep learning to predict geomagnetic indices in storm time and quiet time.

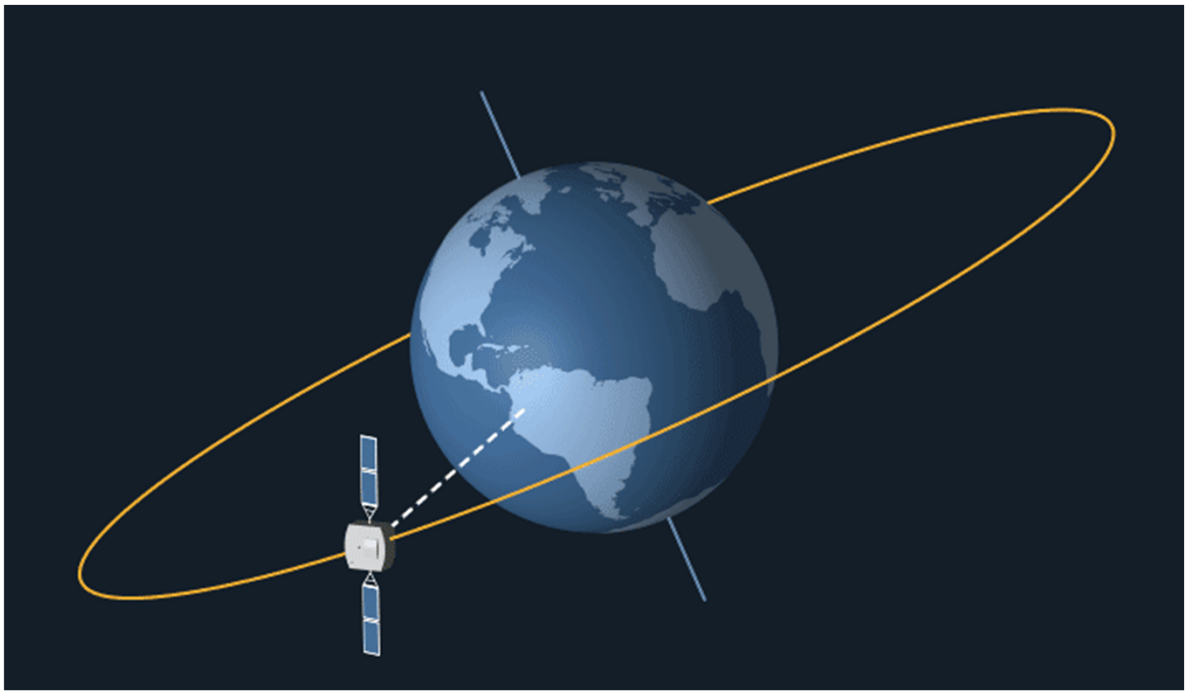
[Abduallah et al. Space Weather 2024]




9h ahead prediction
Predicted
Observed
Epistemic
Aleatoric
125
100
75
50
25
0
−25
ap index
02/01/22
04/01/22
06/01/22
08/01/22
10/01/22
12/01/22
Time


9h ahead prediction
10
8
6
4
2
0
−2
Kp index
Predicted
Observed
Epistemic
Aleatoric
03/19/01 15:00
03/20/01 09:00
03/21/01 03:00
03/21/01 21:00
Time

Investigating how space weather impacts satellite drag and orbits



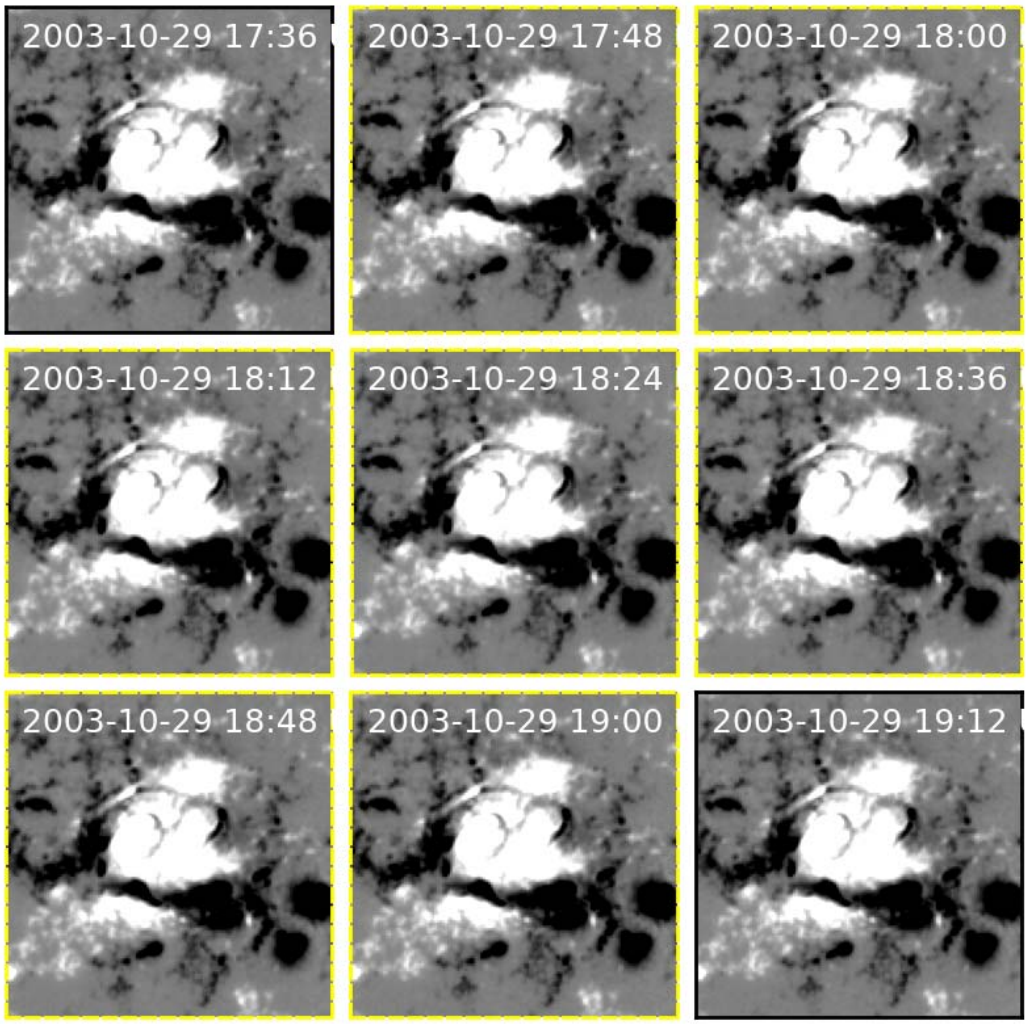
Source: NASA

Investigate how solar indices and geomagnetic indices affect satellite drag and orbits (thermospheric neutral density model)

Joint work with Rutgers University

Outline

- Solar Data Science: An Emerging Interdisciplinary Field
- AI/ML for Solar Data Analysis
- AI/ML for Space Weather Forecasting
- Roles of Generative AI (GenAI) and Explainable AI (XAI)
- Big Data-Enabled AI-Powered Community-Coordinated Cyberinfrastructure
- Conclusion



AI/ML for Improving Temporal Resolution

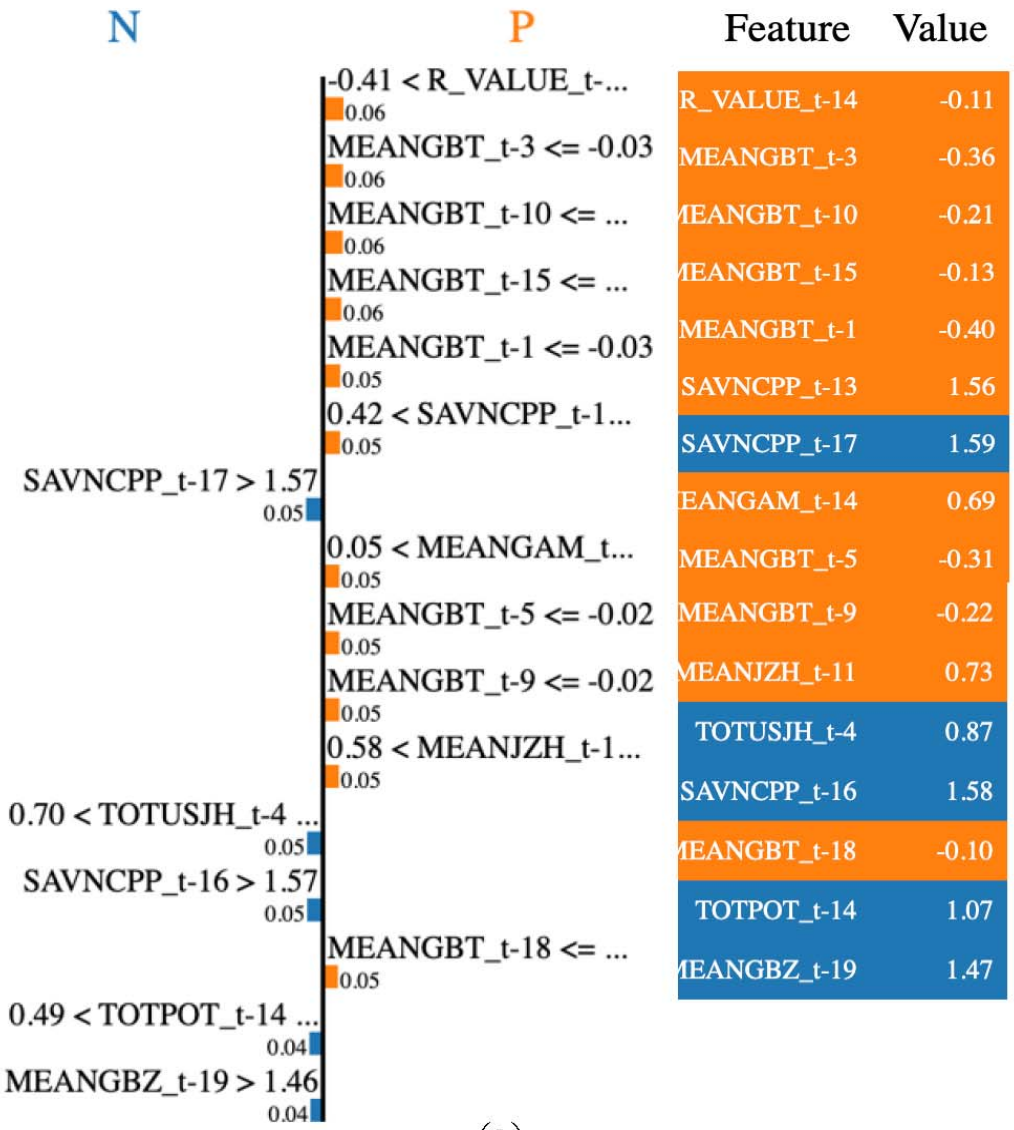
Use GenAI to improve the cadence of SOHO/MDI magnetograms from 96 minutes to 12 minutes. True (synthetic, respectively) magnetograms are enclosed by black (yellow, respectively) boundary lines.

XAI Methods

Global interpretation (provide explanation based on the whole test set)
For example, a home loan is denied (reasons: bad credit, short credit history, insufficient income)

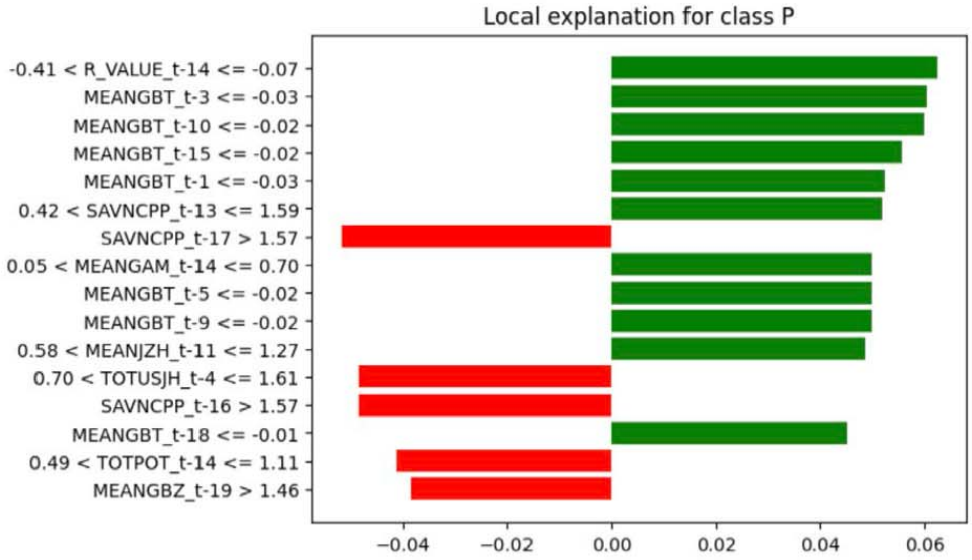
Local interpretation (provide explanation based on a specific test instance)
For example, John's home loan is denied (reason: John is bankrupt recently)

Both global and local interpretations are used to explain the decision-making process of an AI/ML model. XAI methods are particularly useful when the AI/ML model uses many physical parameters (magnetic field parameters, solar wind parameters, etc.) to make predictions.



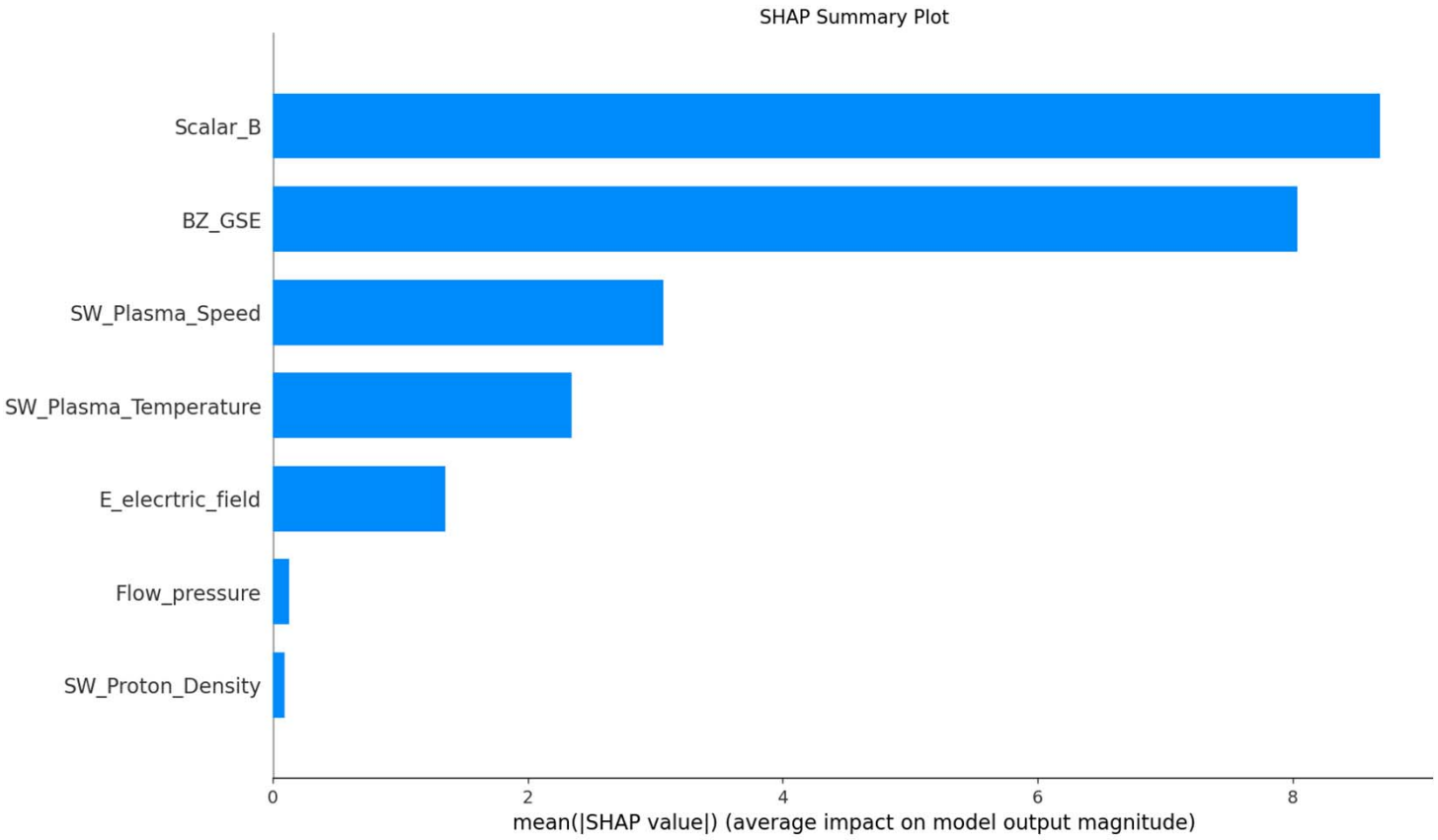
(a)

Feature	Value
R_VALUE_t-14	-0.11
MEANGBT_t-3	-0.36
MEANGBT_t-10	-0.21
MEANGBT_t-15	-0.13
MEANGBT_t-1	-0.40
SAVNCPP_t-13	1.56
SAVNCPP_t-17	1.59
EANGAM_t-14	0.69
MEANGBT_t-5	-0.31
MEANGBT_t-9	-0.22
MEANJZH_t-11	0.73
TOTUSJH_t-4	0.87
SAVNCPP_t-16	1.58
MEANGBT_t-18	-0.10
TOTPOT_t-14	1.07
MEANGBZ_t-19	1.47



(Local Interpretation) Use an XAI tool, LIME, to explain a positive prediction for solar flare forecasting where the input is a time series with 20 time steps each having 16 SHARP magnetic field parameters

(b)



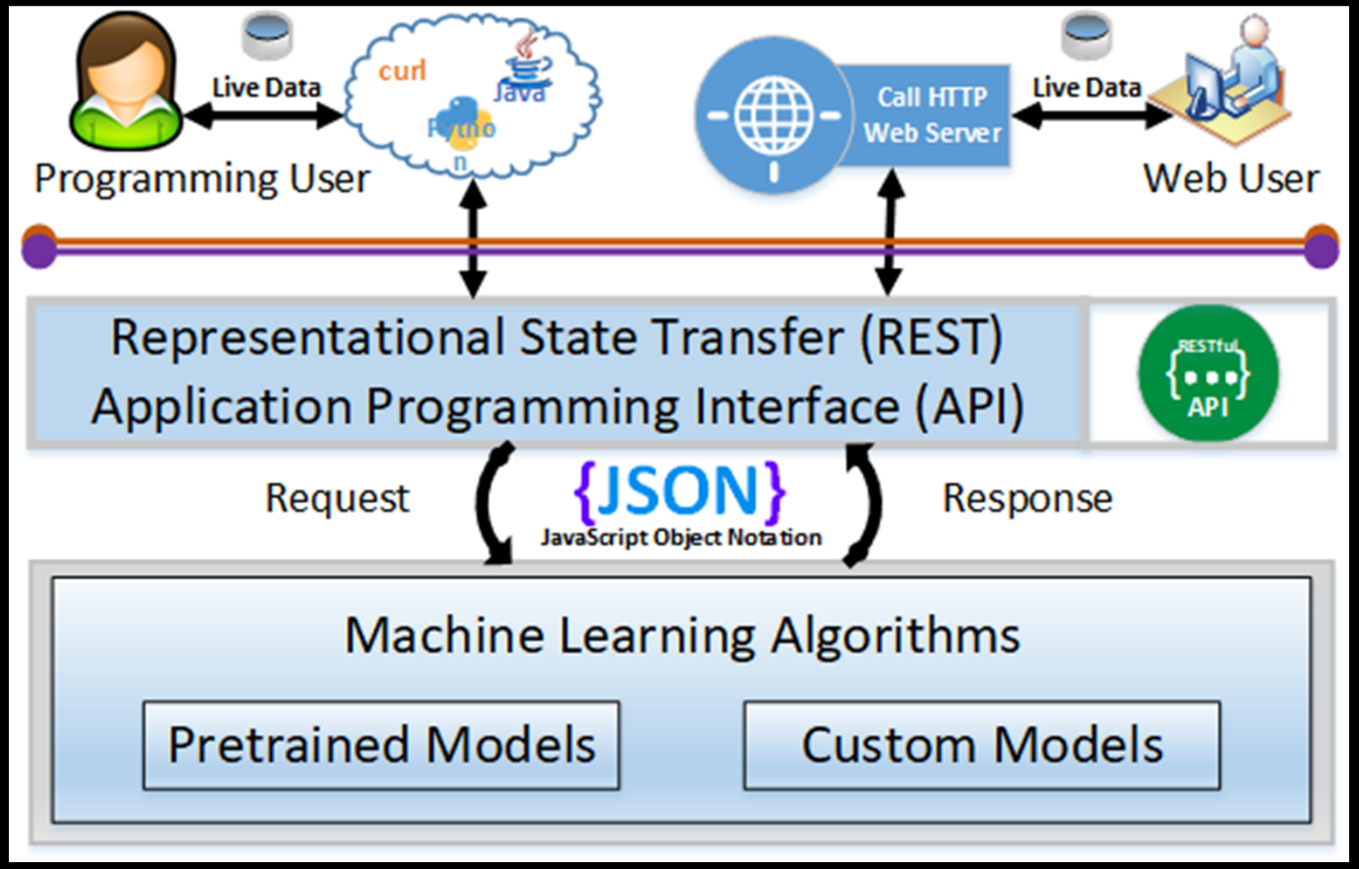
(Global Interpretation) Use an XAI tool, SHAP, to explain/rank the importance of 7 solar wind parameters in the input for Disturbance Storm Time (Dst) index prediction

Outline

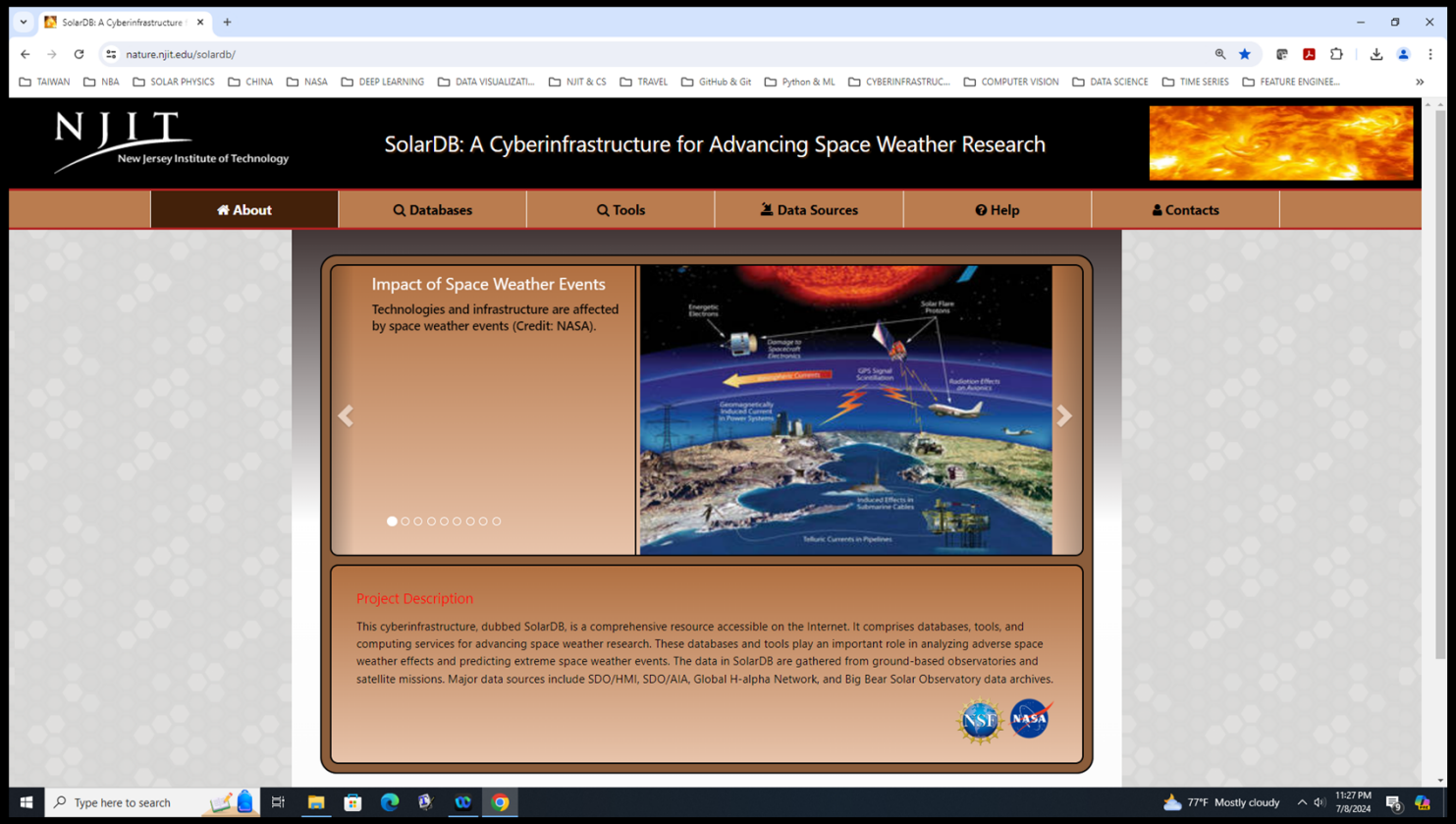
- Solar Data Science: An Emerging Interdisciplinary Field
- AI/ML for Solar Data Analysis
- AI/ML for Space Weather Forecasting
- Roles of Generative AI (GenAI) and Explainable AI (XAI)
- Big Data-Enabled AI-Powered Community-Coordinated Cyberinfrastructure
- Conclusion

DeepSun: ML-as-a-Service for Solar Flare Prediction

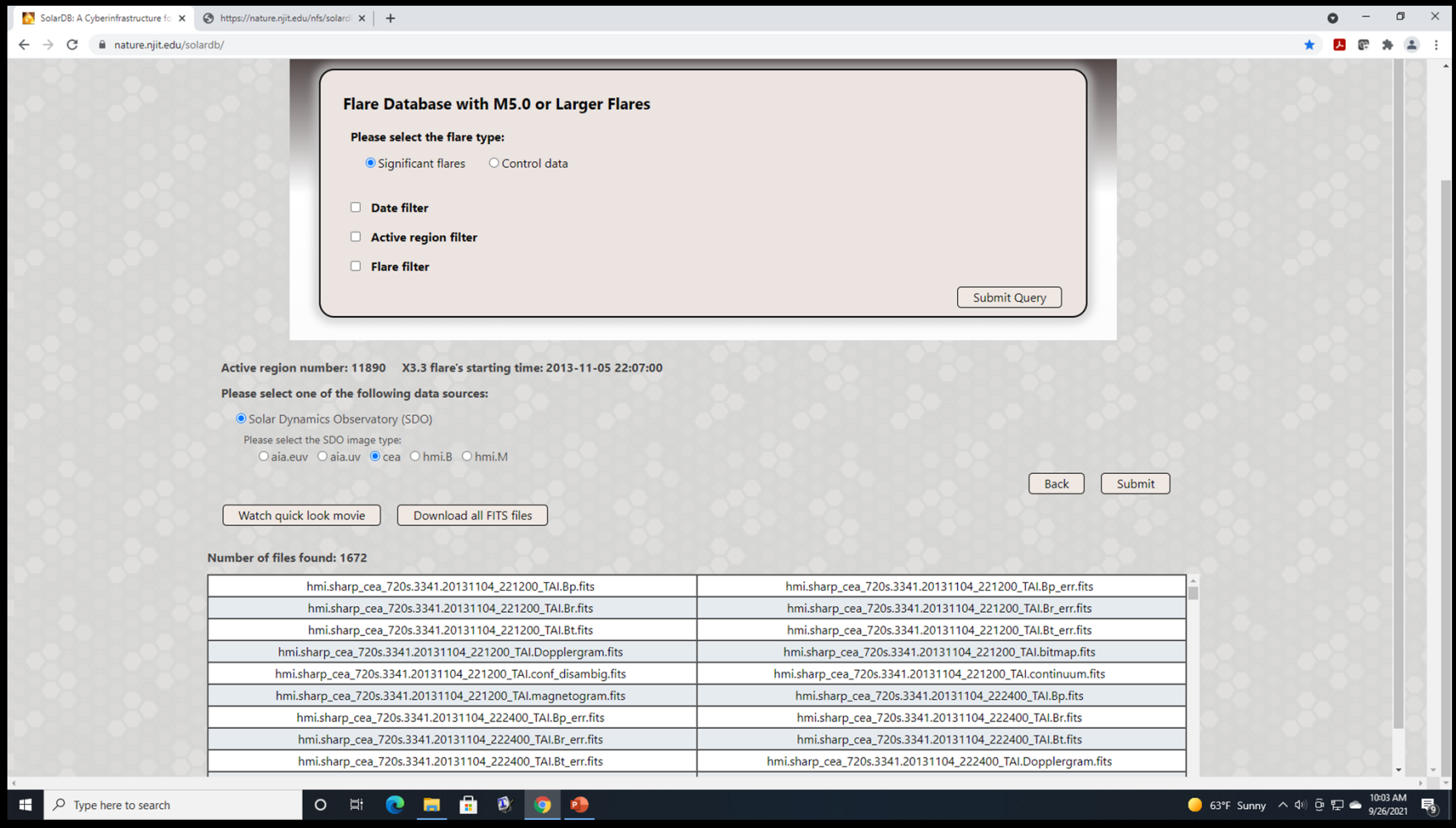
[Abduallah et al. RAA 2021]

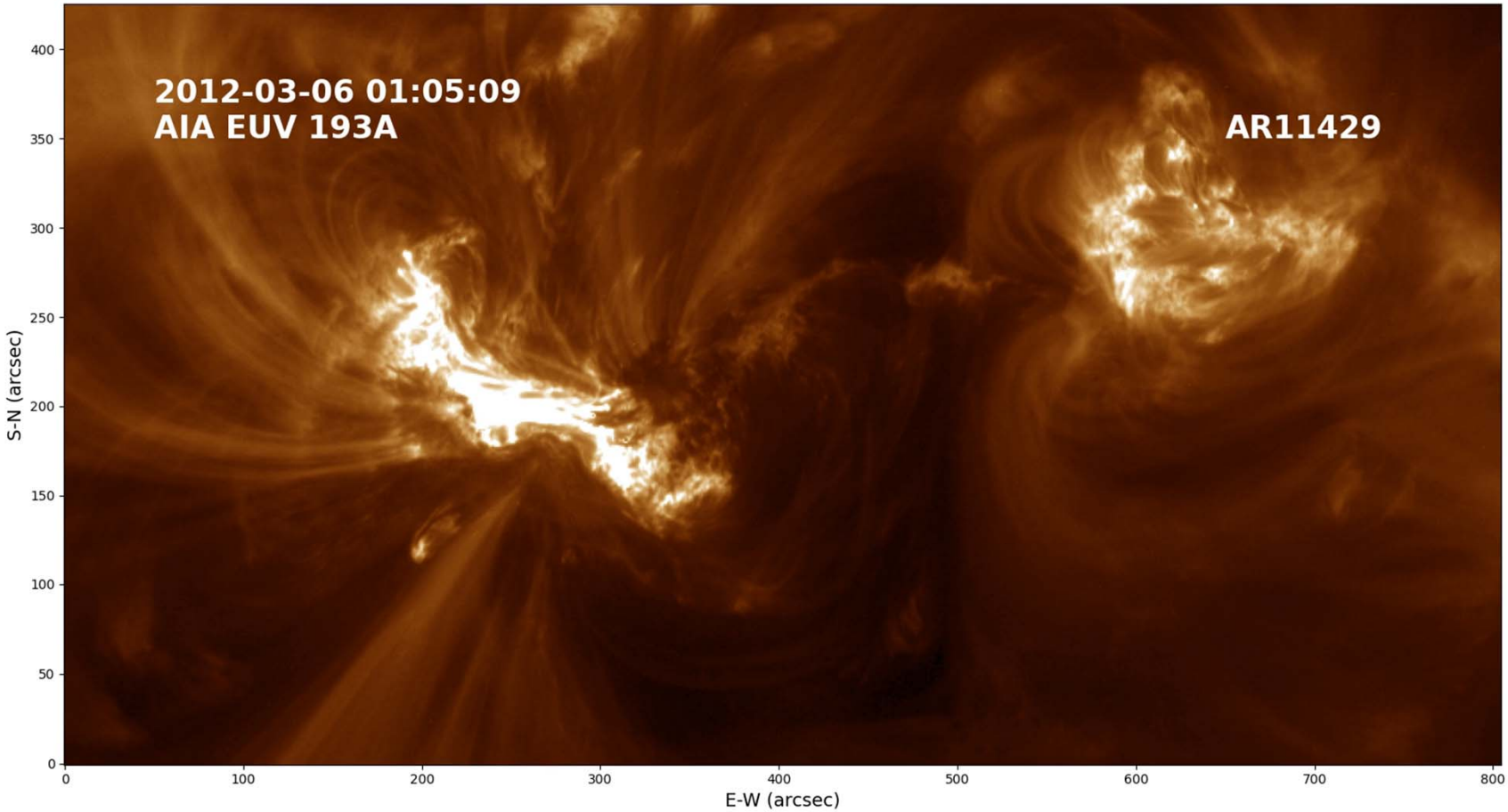


Cyberinfrastructure for Advancing Space Weather Research



Cyberinfrastructure for Advancing Space Weather Research




2012-03-06 01:05:09
AIA EUV 193A
AR11429
S-N (arcsec)
E-W (arcsec)
0
50
100
150
200
250
300
350
400
0
100
200
300
400
500
600
700
800

NJIT New Jersey Institute of Technology

SolarDB: A Cyberinfrastructure for Advancing Space Weather Research

Big Bear Solar Observatory (Historic Observations)

- ☐ Date filter
- ☐ Active region filter
- ☐ Instrument filter

Submit Query

Download displayed data in JSON format | Clear Result

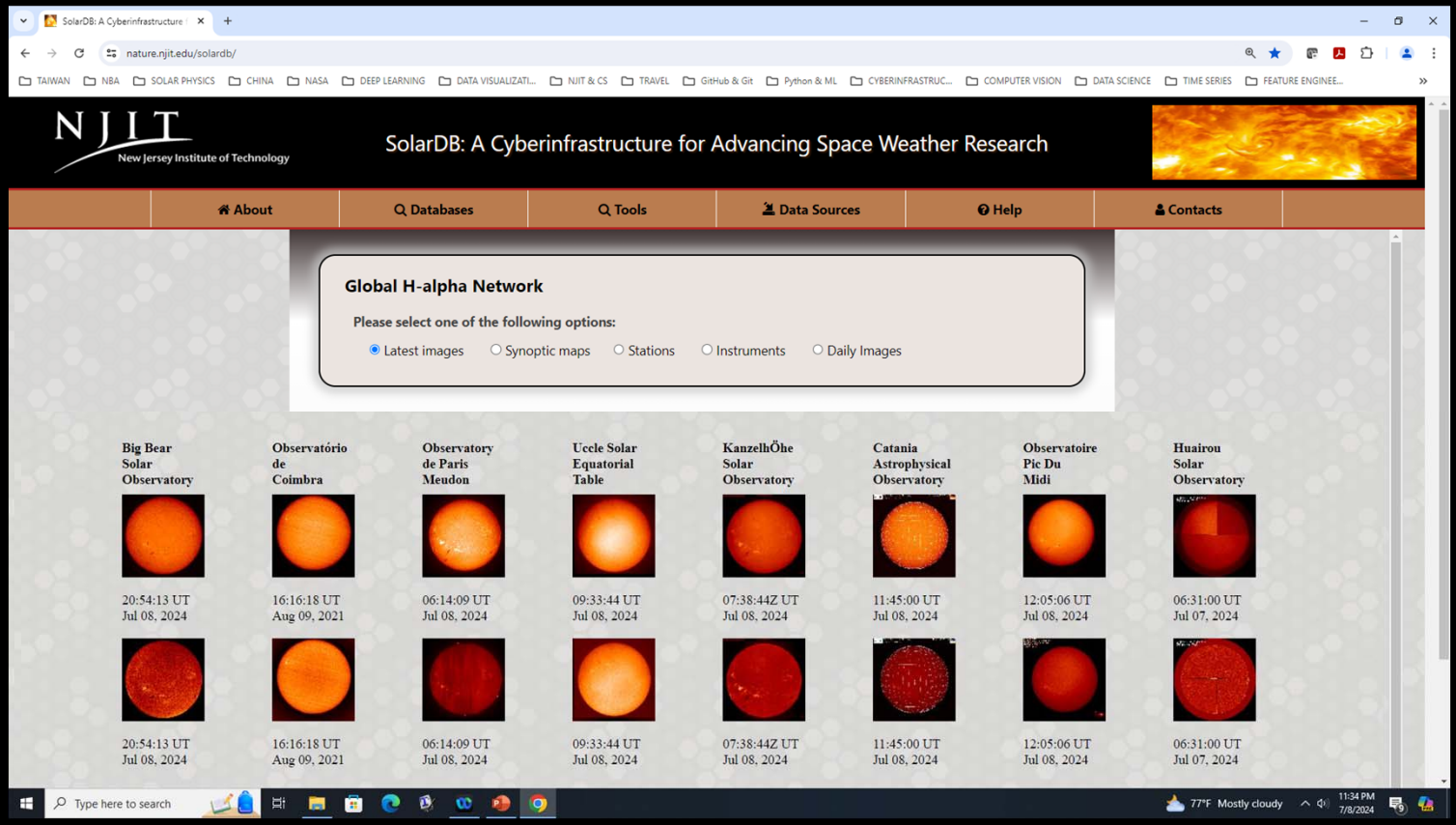
Table of reocrds

1147 records are retrieved based on the selected search criteria. 1 - 100 records are displayed.

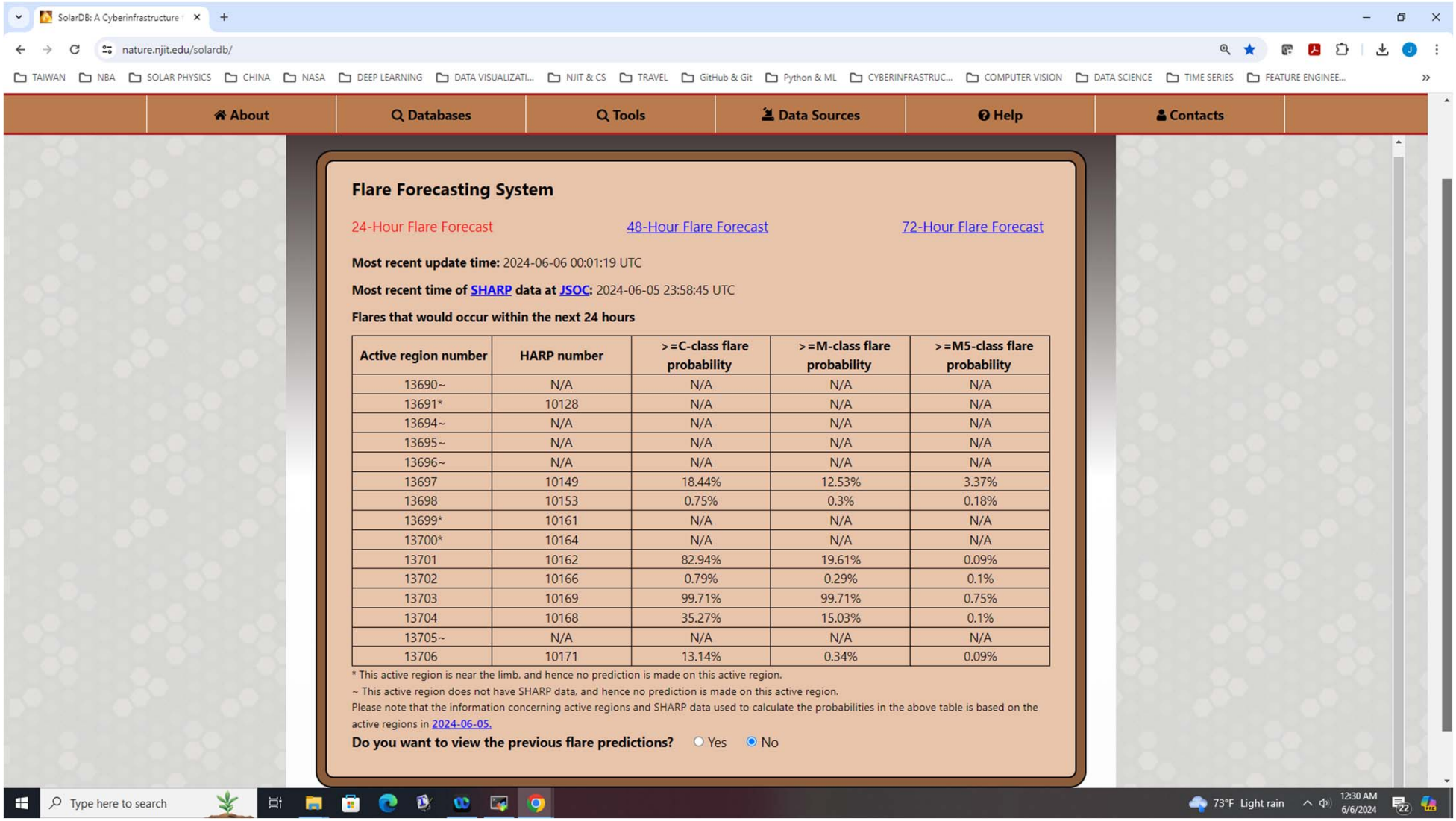
(Note: You can re-sort the results by clicking on the column arrows, the query will re-run.)

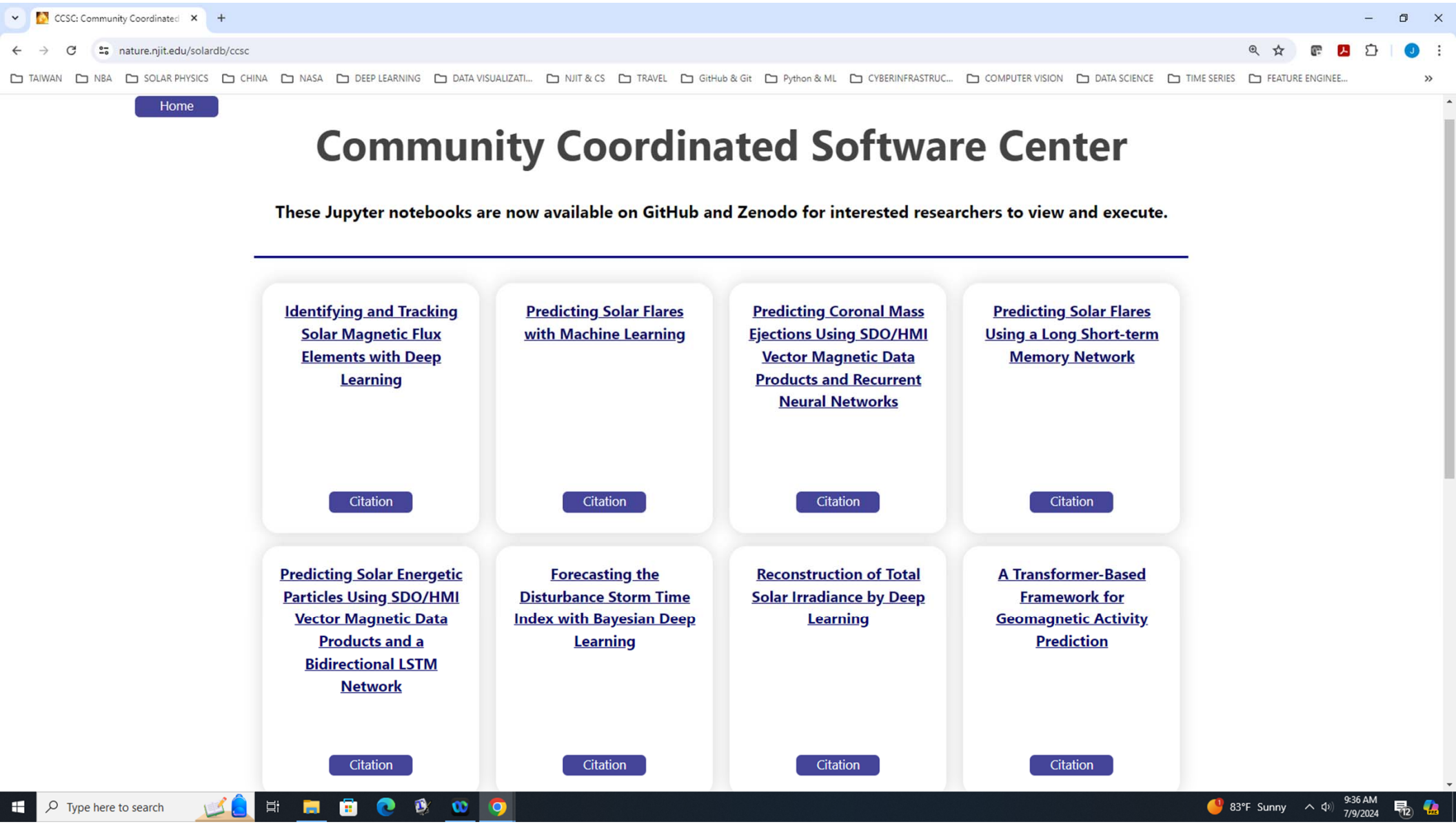
Active Region Number ↓↑	Date ↓↑	Instrument ↓↑
1574	1978-12-13	26E
1574	1978-12-14	26E
1574	1978-12-16	26E
1574	1978-12-26	26E
1574	1978-12-27	26E
1574	1979-01-13	26E
1574	1979-01-22	26E
1574	1979-01-23	26E
1574	1979-01-29	26E
1574	1979-02-07	26E
1574	1979-02-08	26E
1574	1979-02-08	26W

Cyberinfrastructure for Advancing Space Weather Research



[Abduallah et al. Scientific Reports 2023]




CCSC: Community Coordinated
nature.njit.edu/solardb/ccsc
TAIWAN
NBA
SOLAR PHYSICS
CHINA
NASA
DEEP LEARNING
DATA VISUALIZATI...
NJIT & CS
TRAVEL
GitHub & Git
Python & ML
CYBERINFRASTRUC...
COMPUTER VISION
DATA SCIENCE
TIME SERIES
FEATURE ENGINEE...
Home
Community Coordinated Software Center
These Jupyter notebooks are now available on GitHub and Zenodo for interested researchers to view and execute.
Identifying and Tracking Solar Magnetic Flux Elements with Deep Learning
Citation
Predicting Solar Flares with Machine Learning
Citation
Predicting Coronal Mass Ejections Using SDO/HMI Vector Magnetic Data Products and Recurrent Neural Networks
Citation
Predicting Solar Flares Using a Long Short-term Memory Network
Citation
Predicting Solar Energetic Particles Using SDO/HMI Vector Magnetic Data Products and a Bidirectional LSTM Network
Citation
Forecasting the Disturbance Storm Time Index with Bayesian Deep Learning
Citation
Reconstruction of Total Solar Irradiance by Deep Learning
Citation
A Transformer-Based Framework for Geomagnetic Activity Prediction
Citation
Type here to search
83°F Sunny
9:36 AM
7/9/2024

Published June 29, 2022 | Version v1.1

Software | Open

Forecasting the Disturbance Storm Time Index with Bayesian Deep Learning

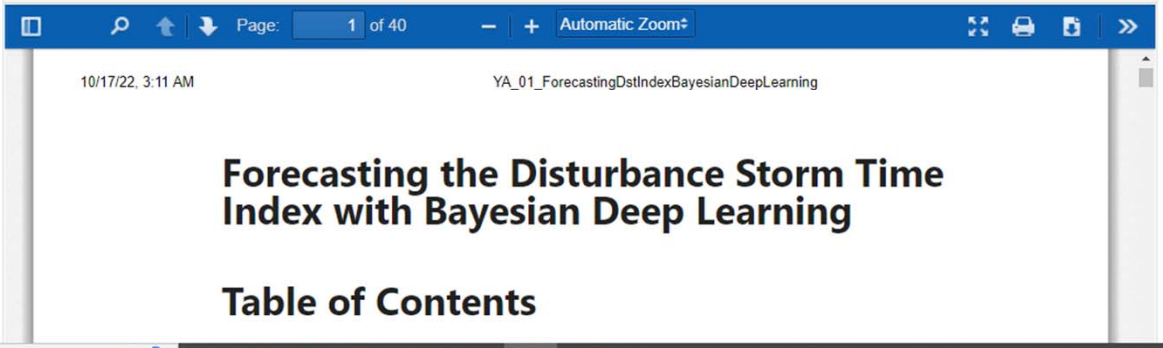
Yasser Abduallah[1]; Jason T. L. Wang[1]; Prianka Bose[1]; Genwei Zhang[1]; Firas Gerges[1]; Haimin Wang[2]

Show affiliations

Forecasting the Disturbance Storm Time Index with Bayesian Deep Learning

Files

Dst-prediction-v1.1.pdf



Conclusion

- AI/ML is part of data-driven science. When more data becomes available, AI/ML comes to play a role. Solar data science is filled with big complex data, so AI/ML is well suited for solar data science.
- AI/ML can not guarantee to yield the best results, but AI/ML can work well when no (better) solution exists!
- However, there are many challenges for AI/ML in solar data science:
  - How to identify a solar data science problem suitable for AI/ML?
  - Which AI/ML model to use (RF, LSTM, CNN, Transformer, GNN, GAN, Diffusion Model, PINN, RL)?
  - Which architecture to use?
  - Which parameters/hyperparameters to use?
  - Which loss function and optimizer to use?
  - How to prepare training data?
  - How to interpret predictions and validate prediction results? Are the results usable and useful?
- On the other hand, there are also many opportunities; our case studies demonstrate some of these opportunities.
- AI/ML in solar data science is still in its infancy. We look forward to more challenges and opportunities in this interdisciplinary field.

Thank you