

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:

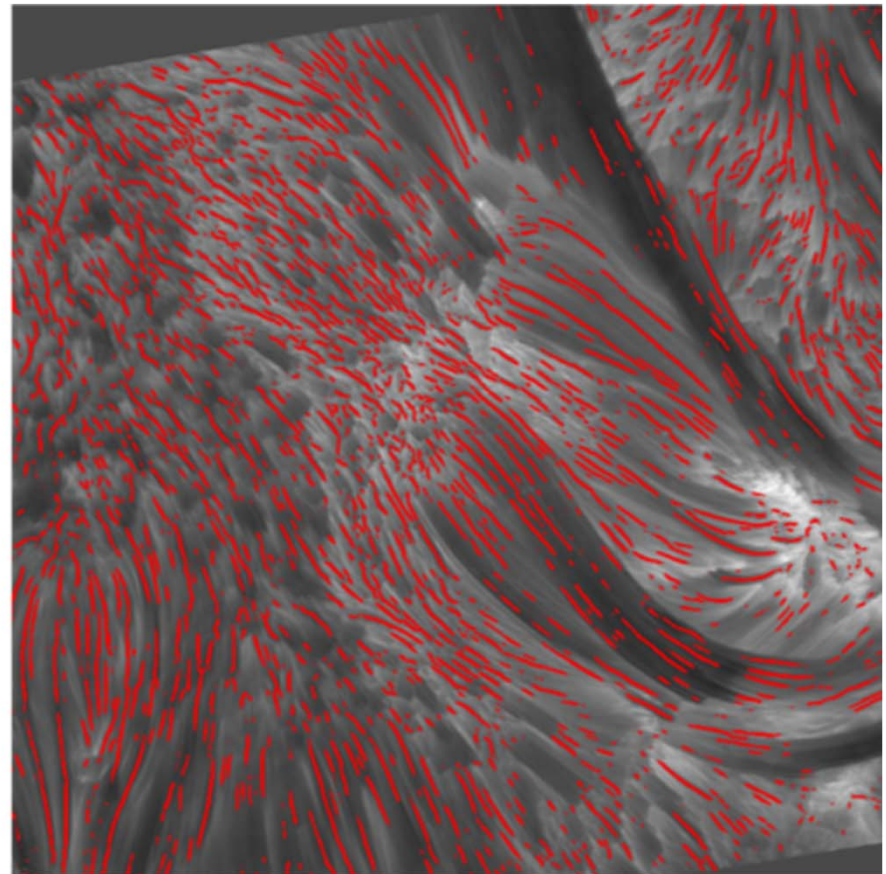
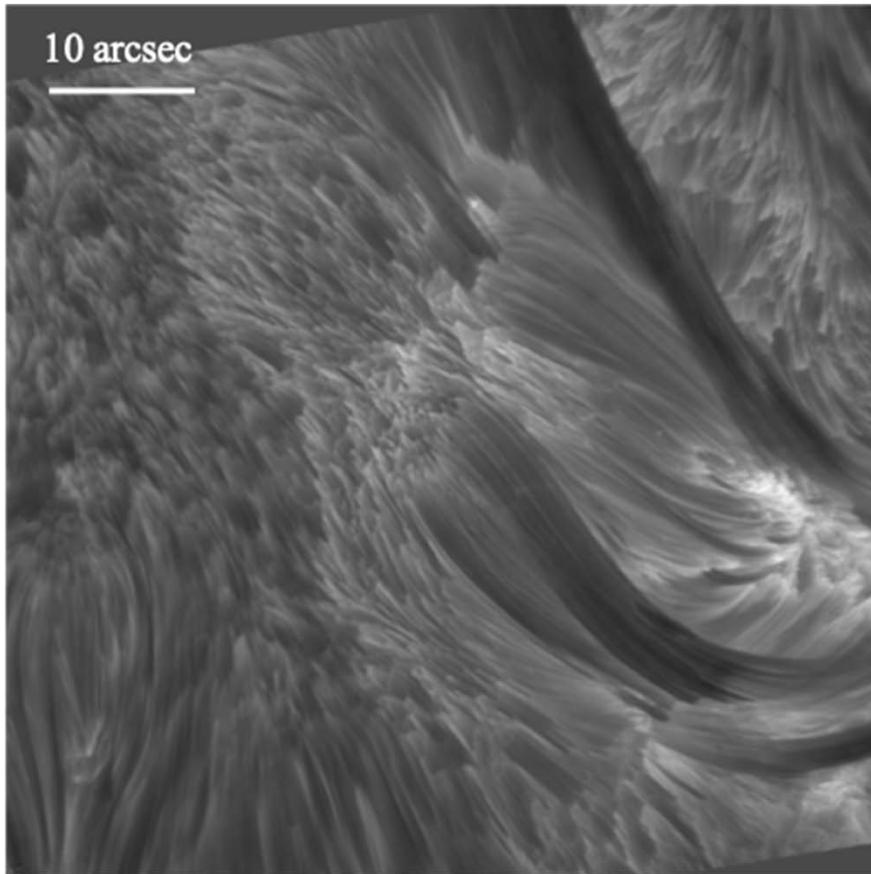
1. AI/ML detects fibrils through making predictions, and hence is often faster than computational tools which involve time-consuming computation. We aim at building near real-time operational and forecasting systems, so AI/ML better meets our need.
2. AI/ML models have inference and generalization capabilities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

Outline

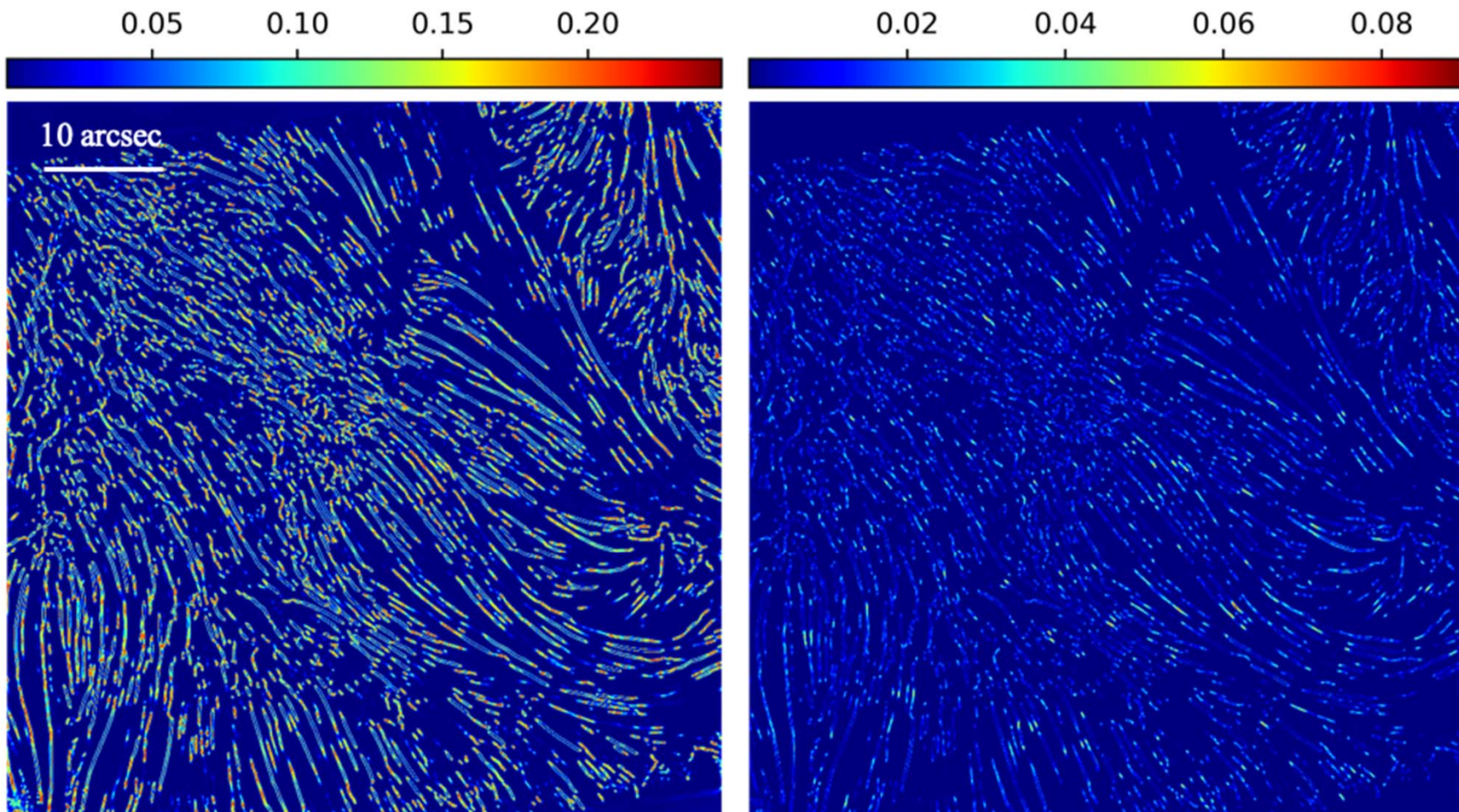
- Solar Data Science: An Emerging Interdisciplinary Field
- AI/ML for Solar Data Analysis
- AI/ML for Space Weather Forecasting
- Roles of Generative AI (GenAI) and Explainable AI (XAI)
- Big Data-Enabled AI-Powered Community-Coordinated Cyberinfrastructure
- Conclusion

AI/ML for Fibril Detection

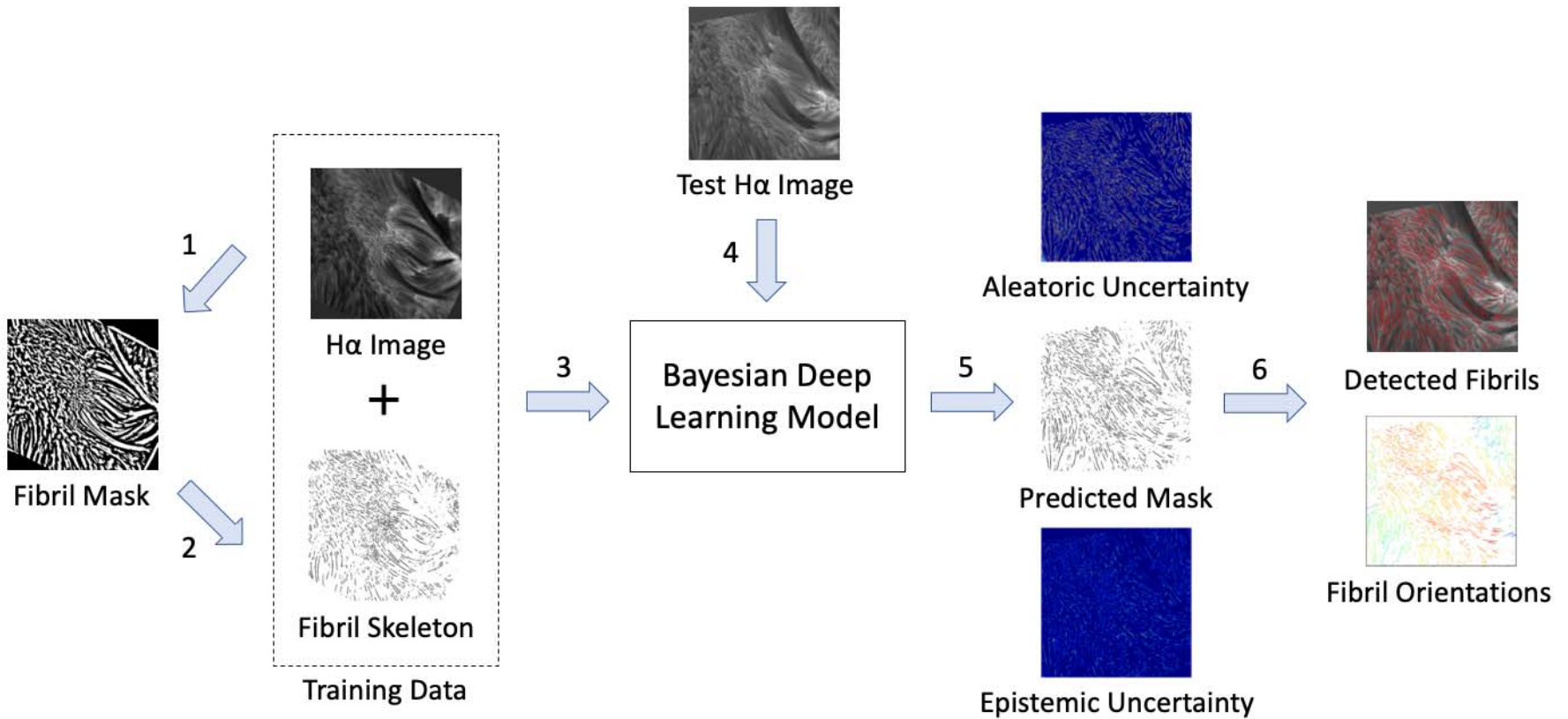
[Jiang et al. ApJS 2021]



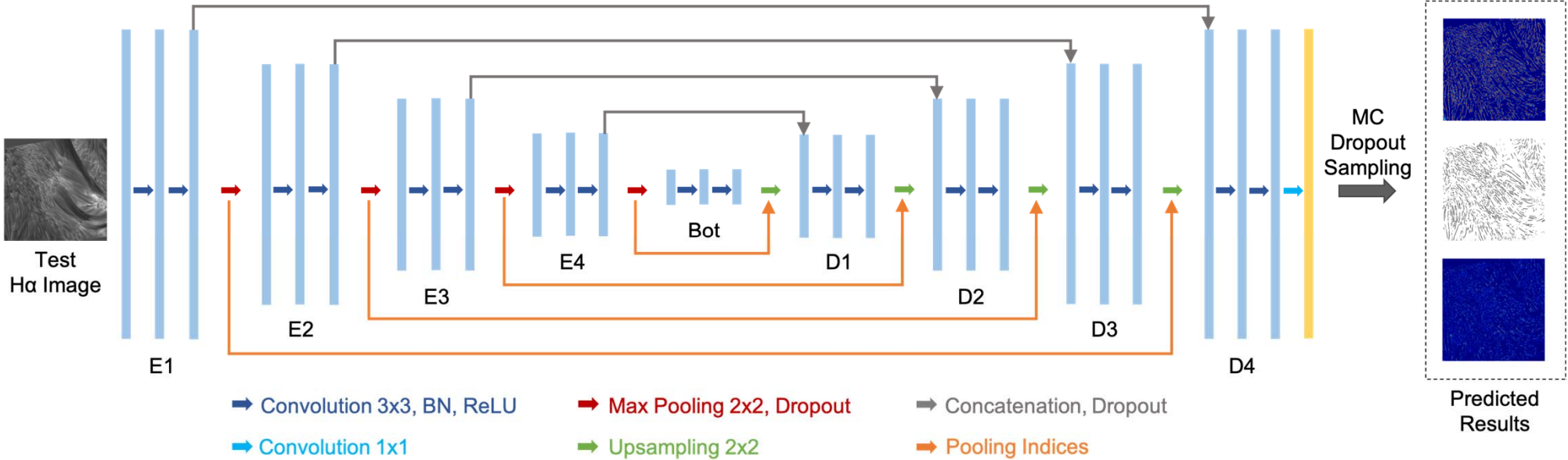
AI/ML for Fibril Detection



AI/ML for Fibril Detection

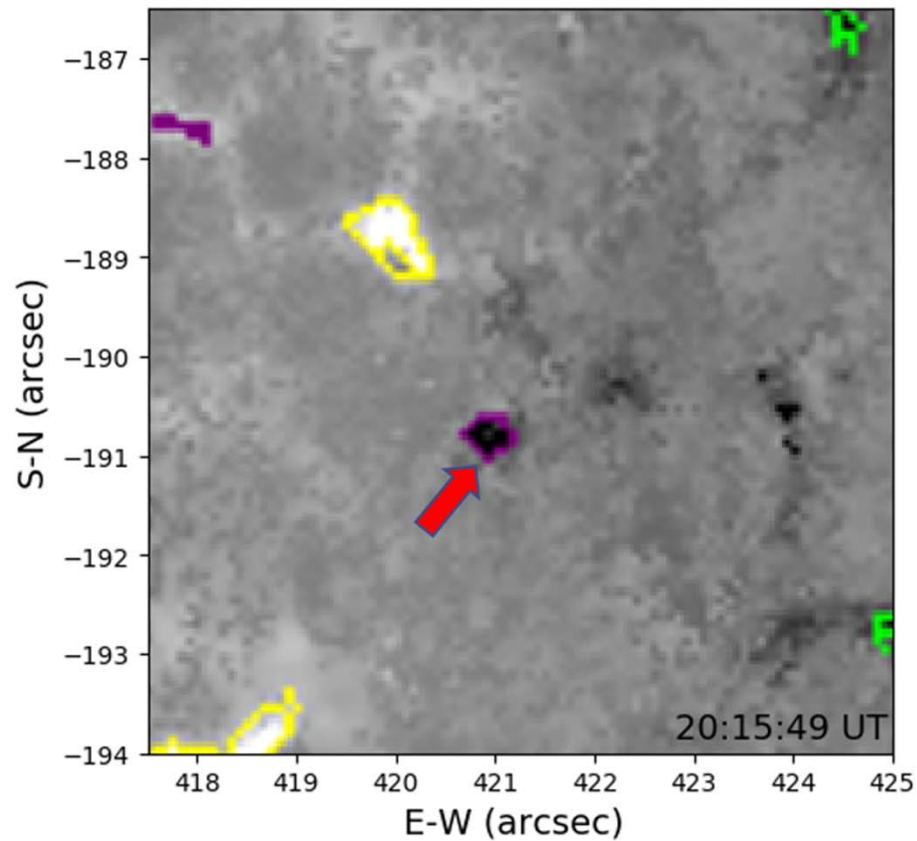


Bayesian Deep Learning Model for Fibril Detection

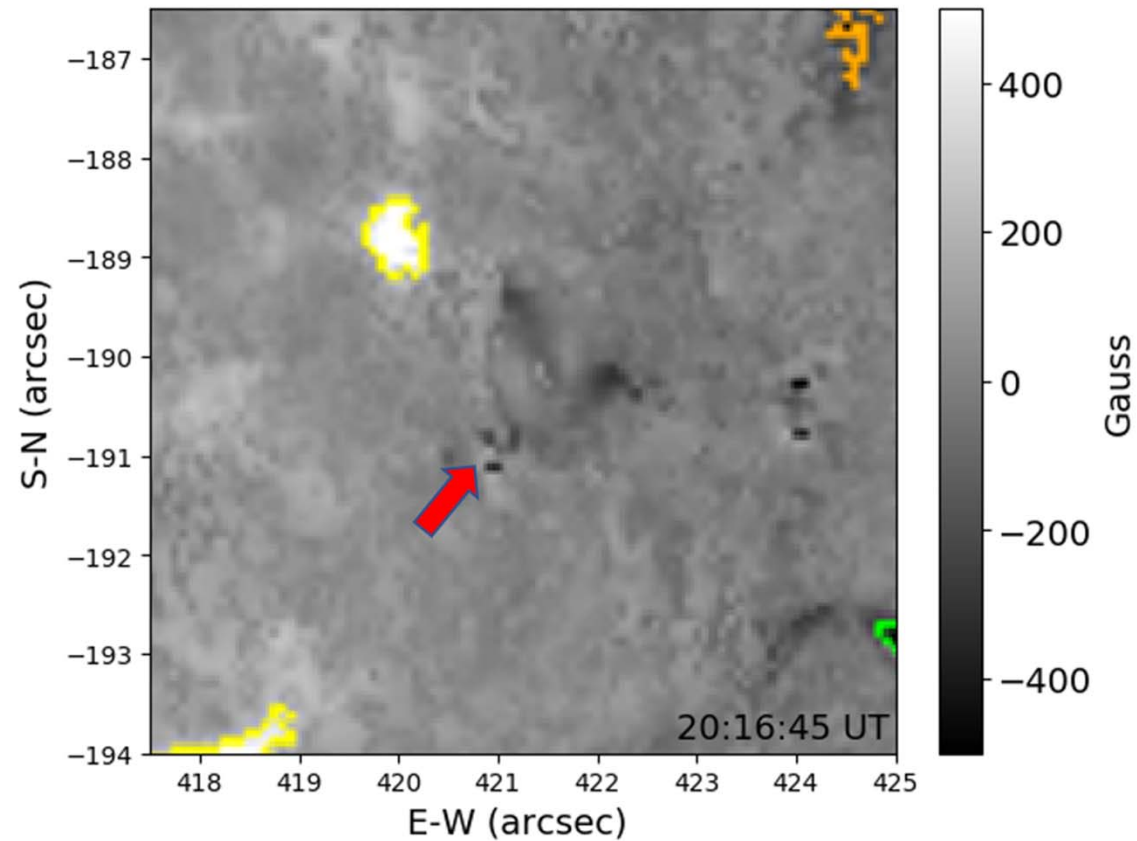


AI/ML for Magnetic Flux Tracking

flux element disappearance [Jiang et al. ApJS 2020]



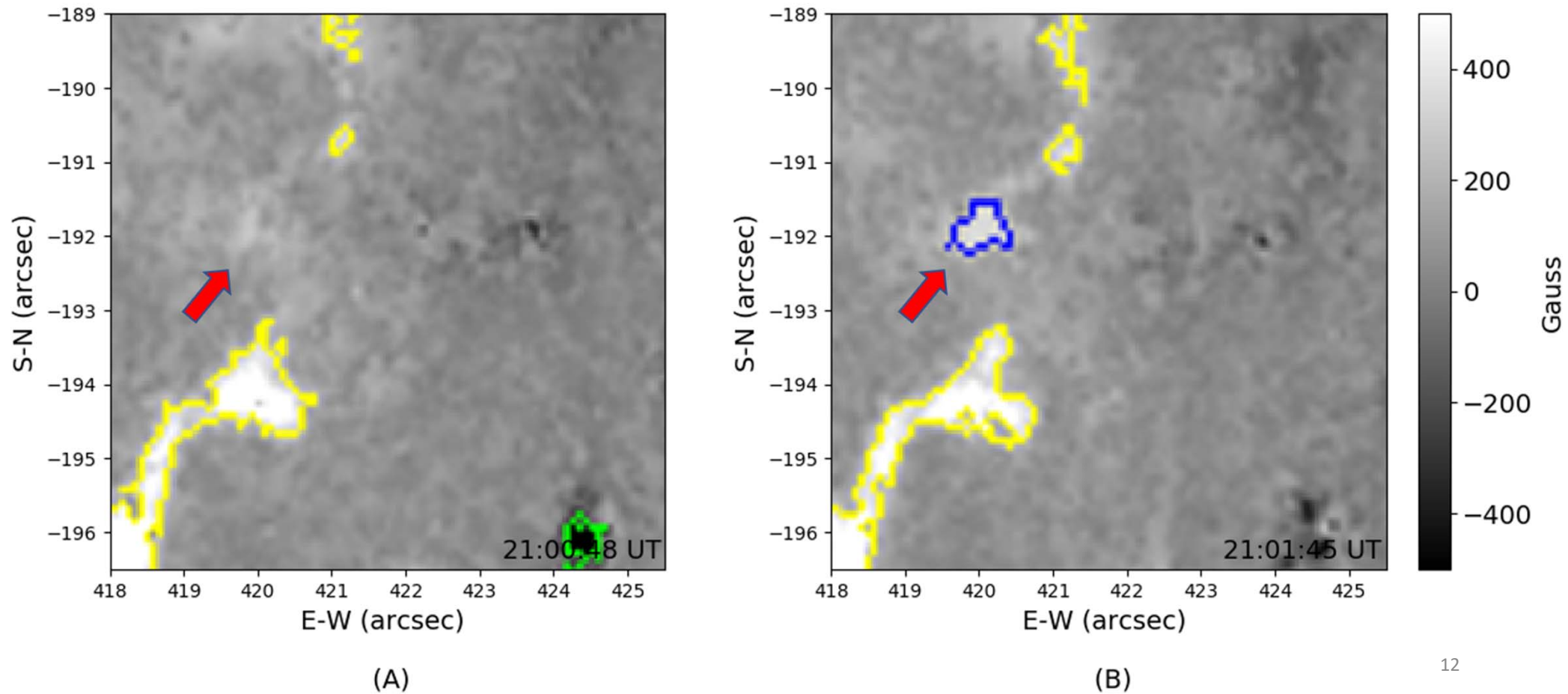
(A)



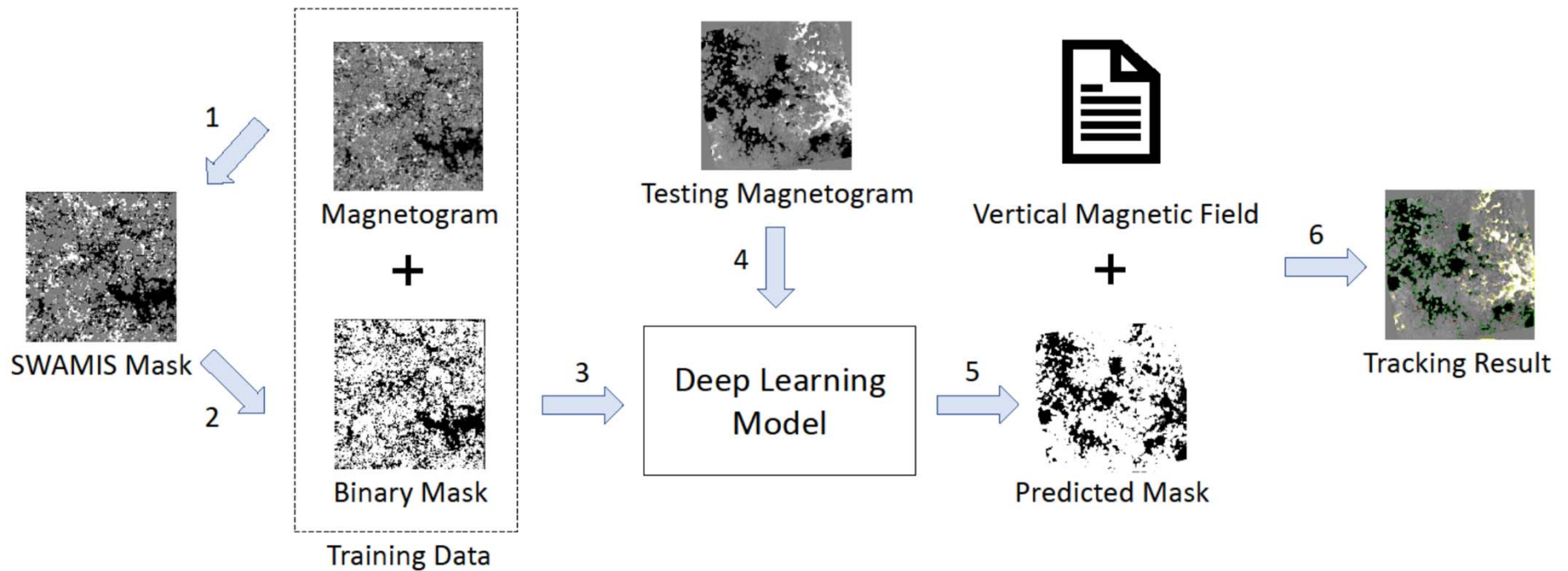
(B)

AI/ML for Magnetic Flux Tracking

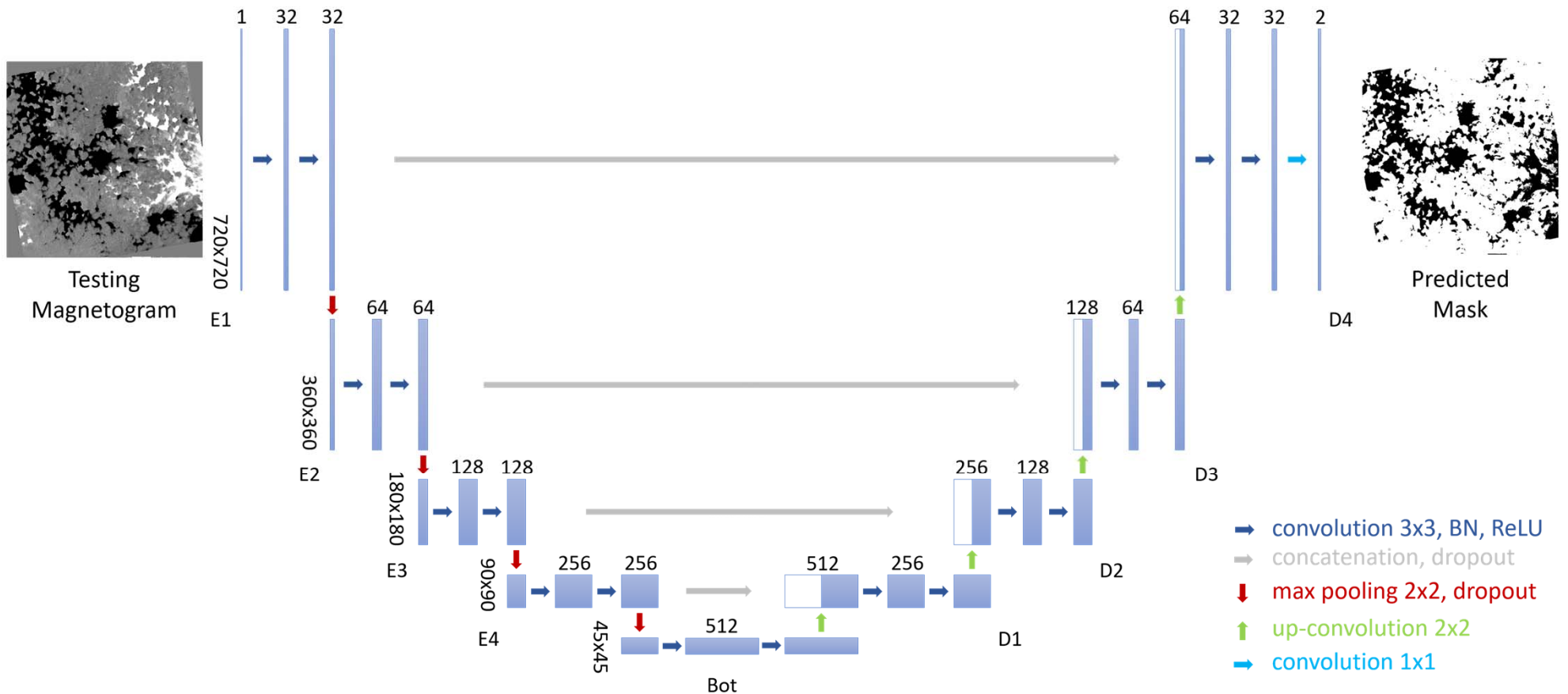
flux element appearance



AI/ML for Magnetic Flux Tracking

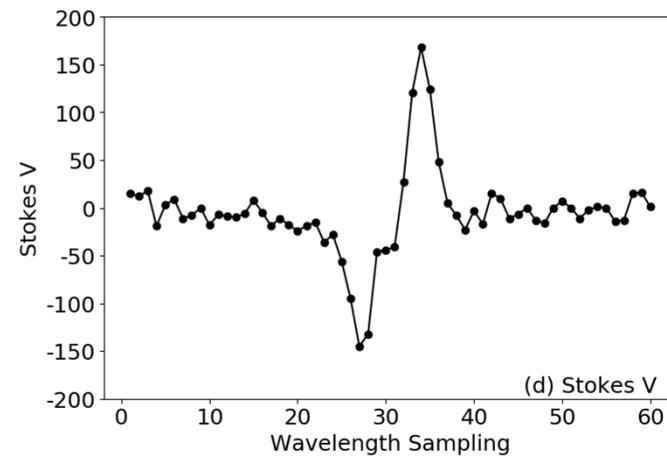
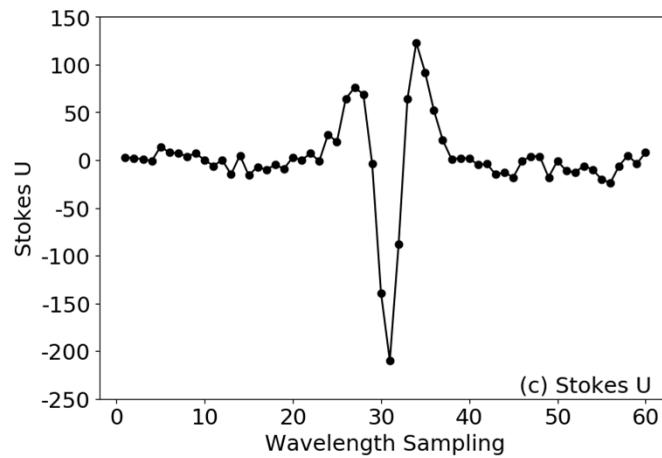
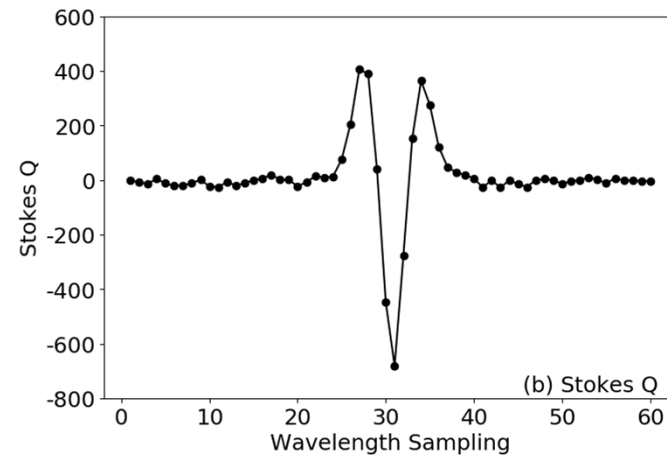
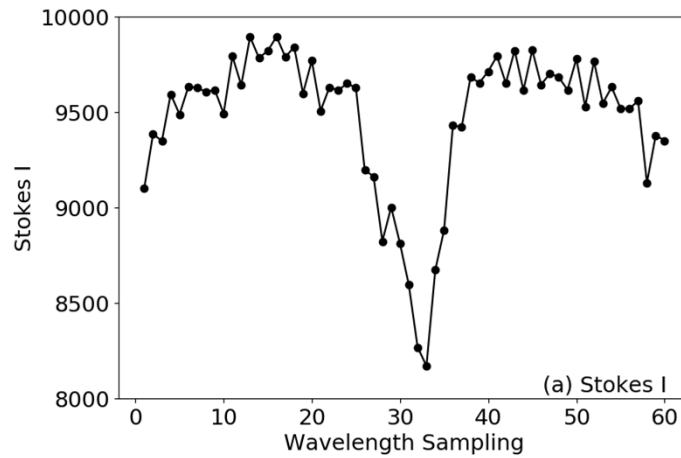


Deep Learning Model for Magnetic Flux Tracking

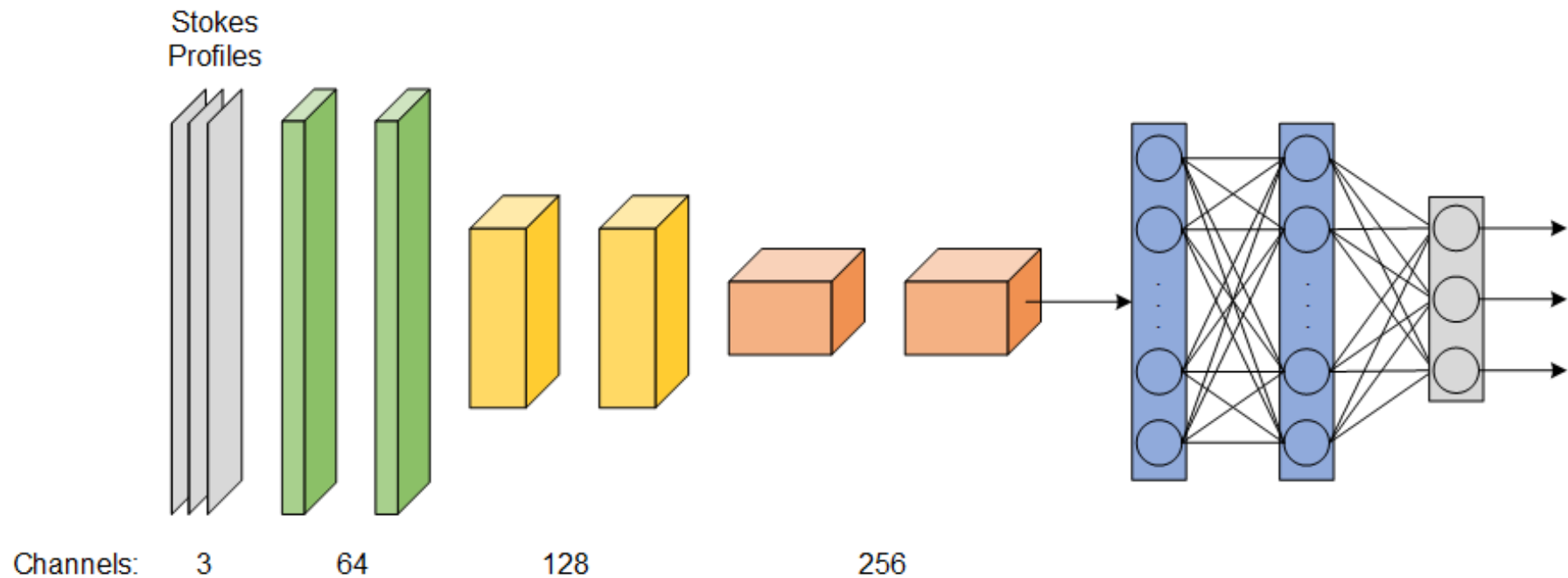


AI/ML for Stokes Inversion

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

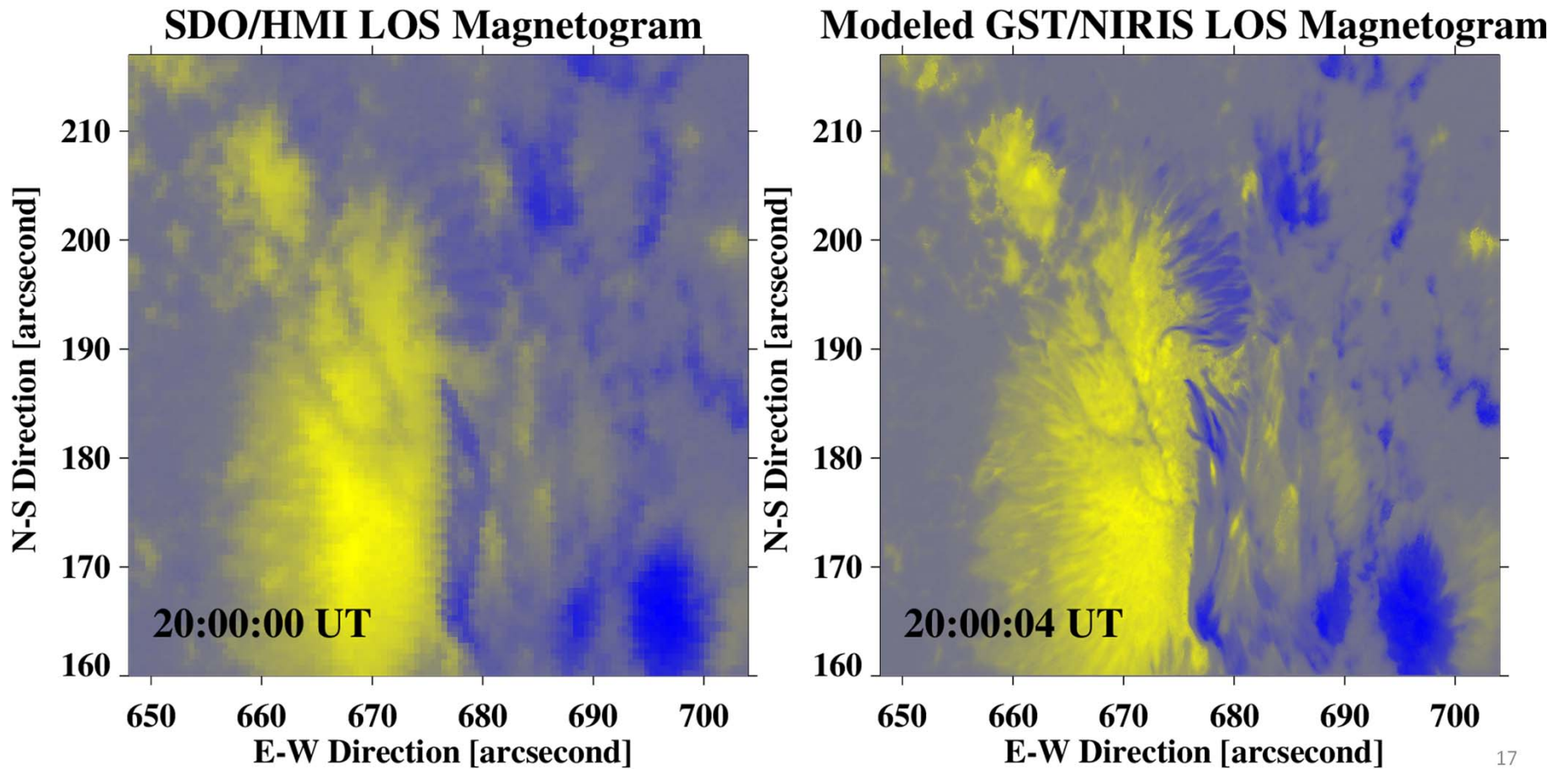


AI/ML for Stokes Inversion



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

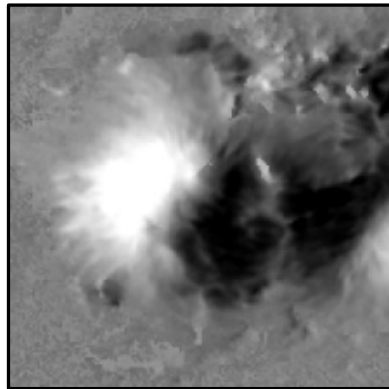
AI/ML for Stokes Inversion



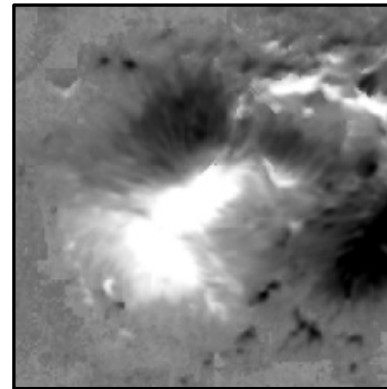
AI/ML for Solar Vector Magnetogram Generation

[Jiang et al. Solar Physics 2023]

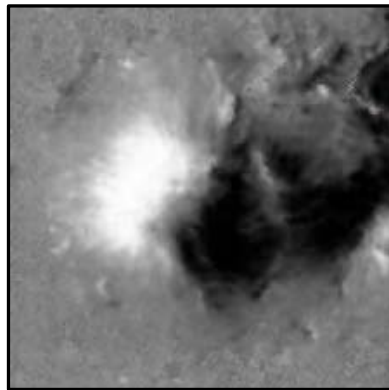
Original Bx



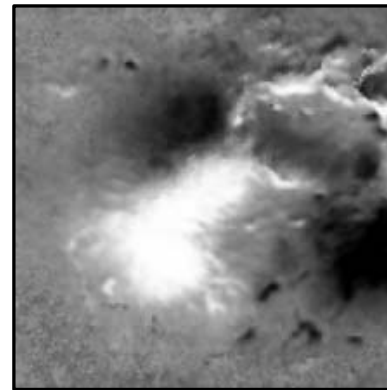
Original By



AI-Generated Bx



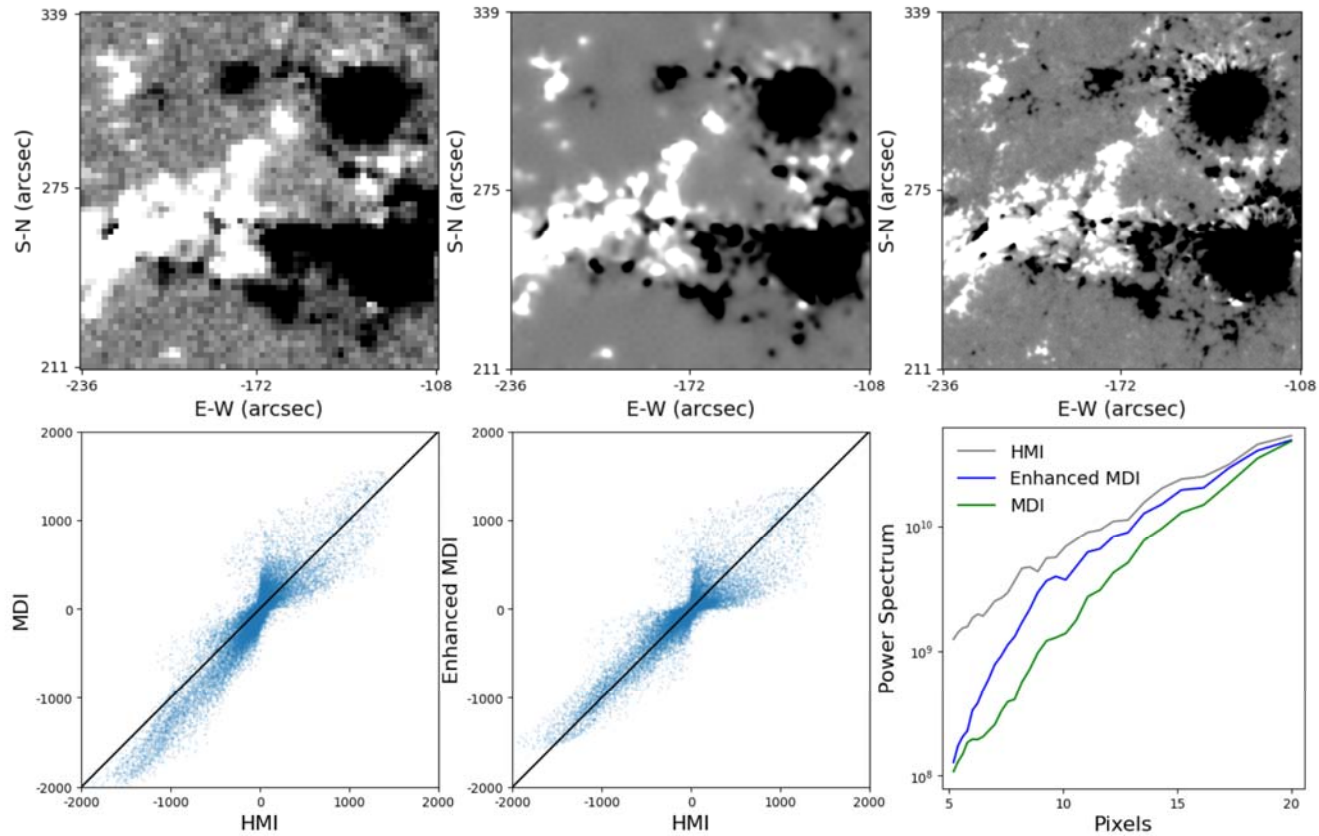
AI-Generated By



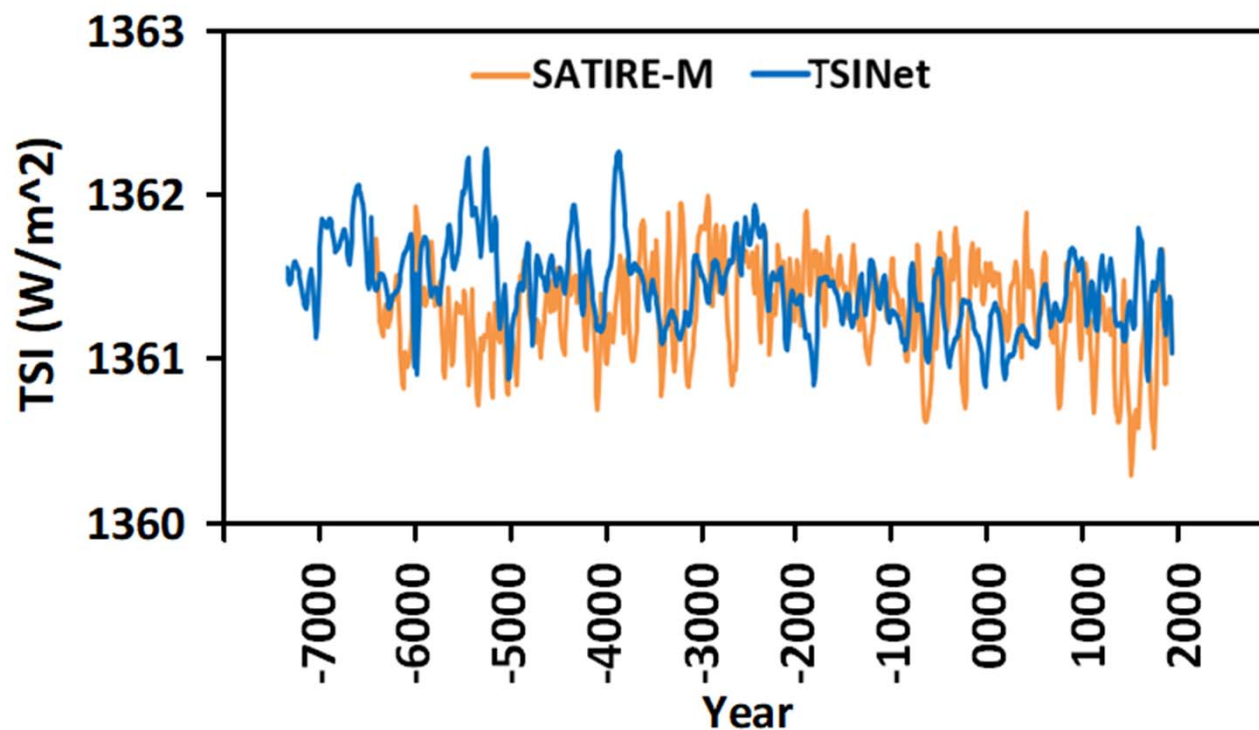
Use AI-Generated Transverse Components (Bx, By) of Magnetic Field to Study Solar Flare Activity in Multiple Solar Cycles

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

[Xu et al. Solar Physics 2024]

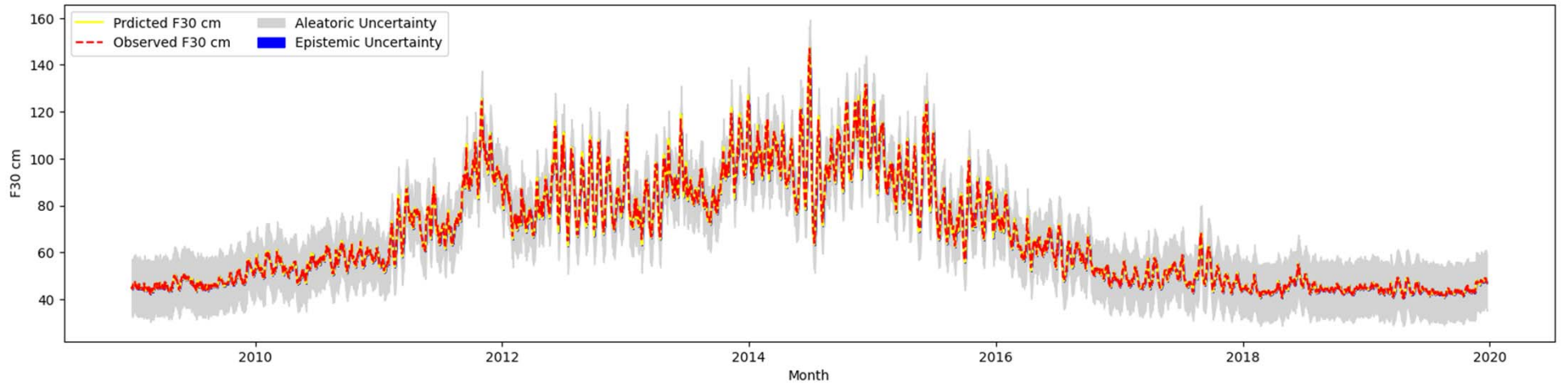
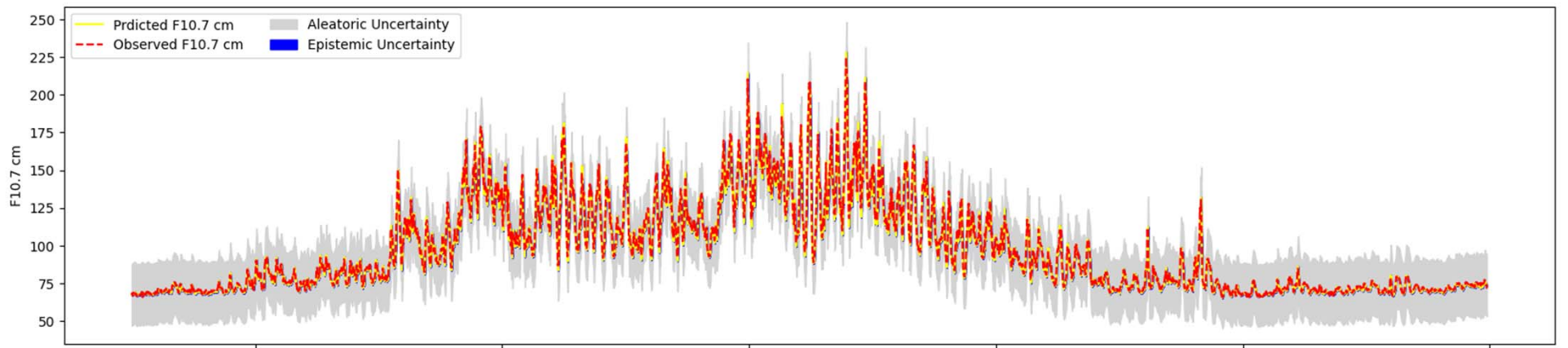


AI/ML for Solar Irradiance Reconstruction



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

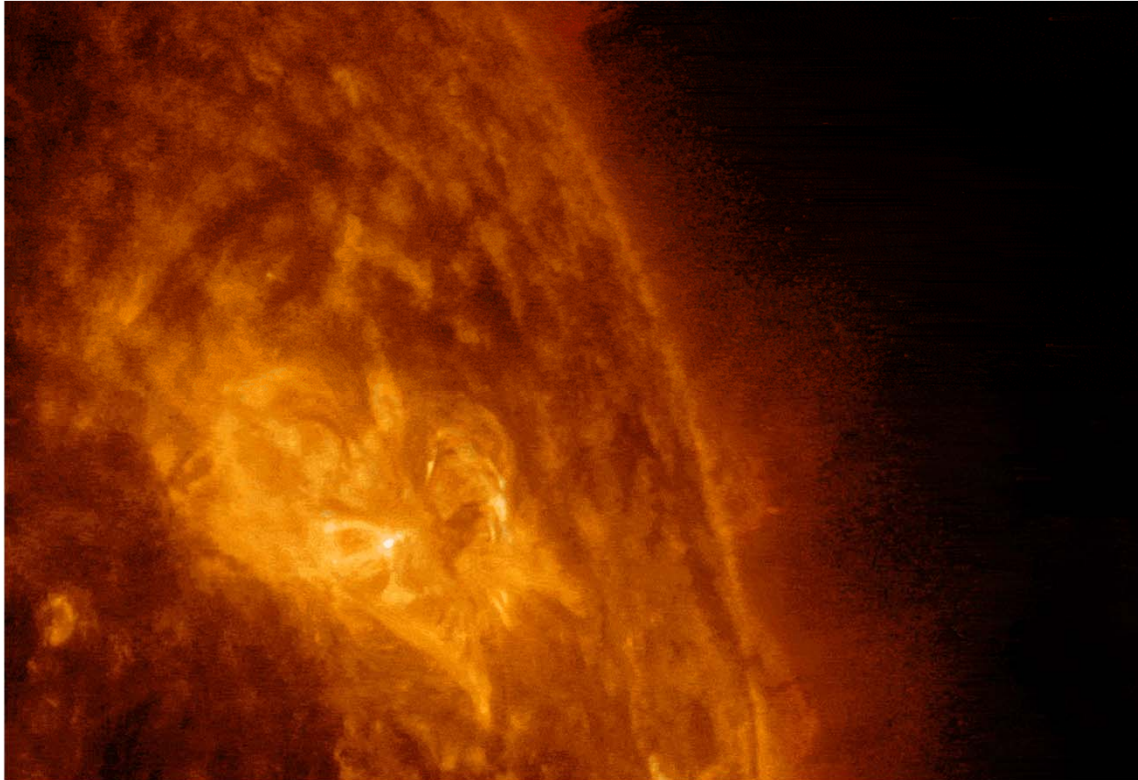
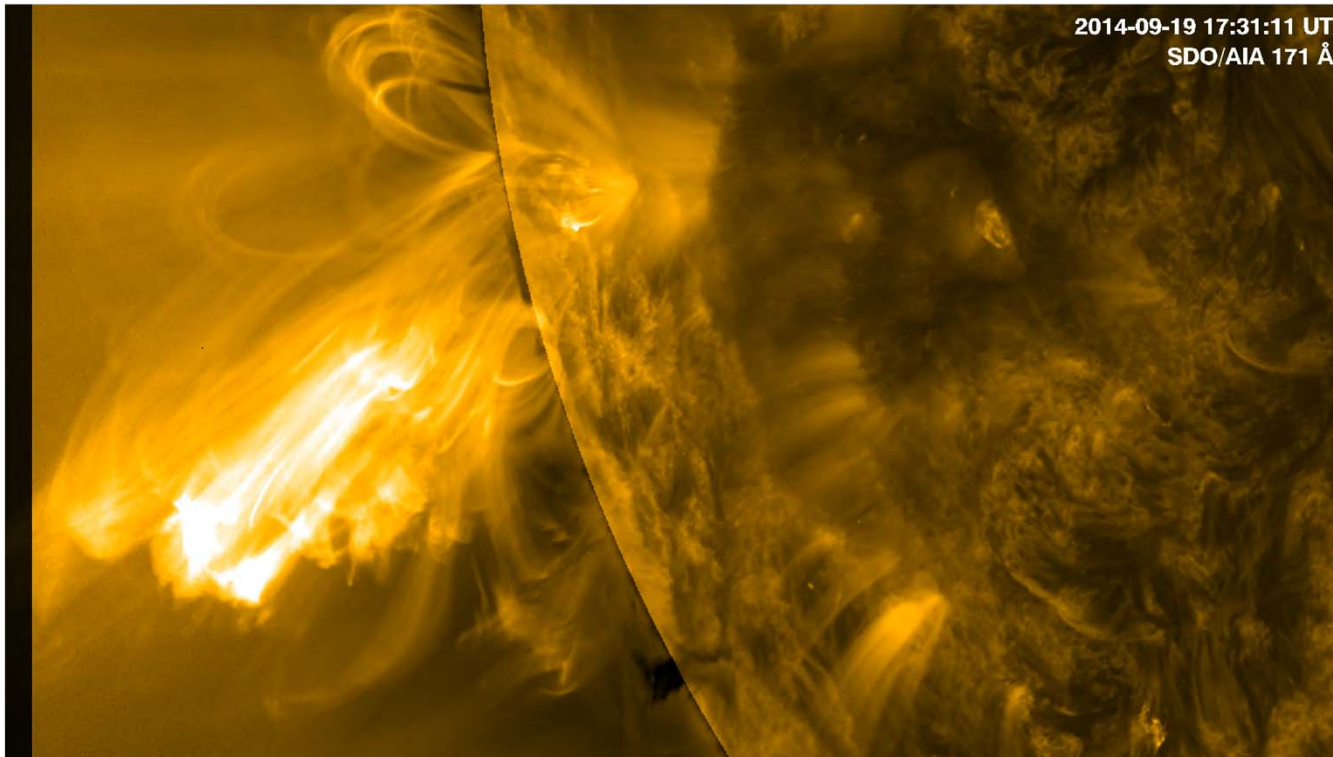


Image credit: NASA/SDO/Goddard
<https://spaceplace.nasa.gov/solar-activity/en/>

Space Weather Events: Coronal Mass Ejections (CMEs)

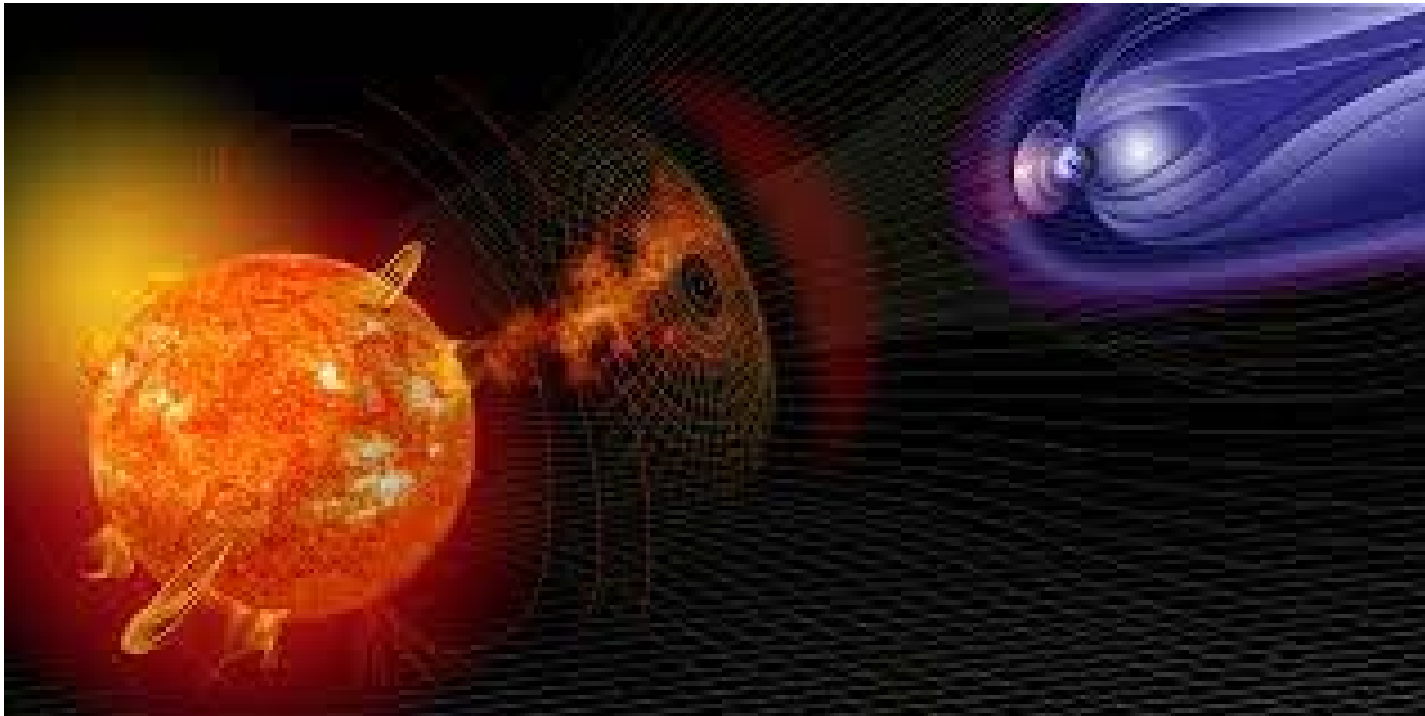
Filament Eruption
Leading to a CME



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

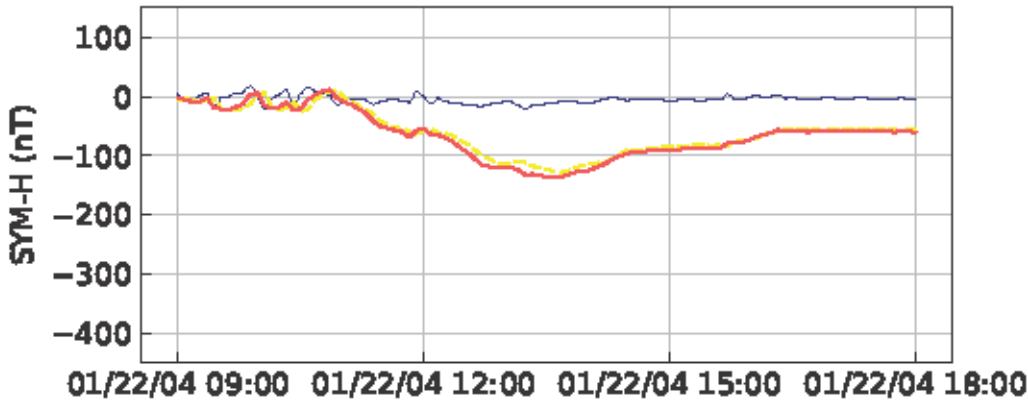
We use deep learning to predict Dst (disturbance storm time) index values focusing on intense storms ($-250 \text{ nT} < \text{Dst} < -50 \text{ nT}$) and super storms ($\text{Dst} < -250 \text{ nT}$).

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

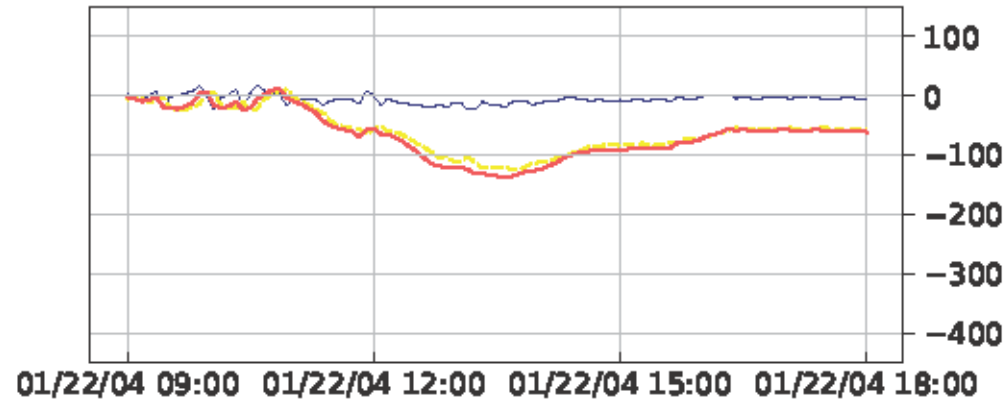
Use deep learning to predict geomagnetic indices in storm time and quiet time.

[Abduallah et al. Space Weather 2024]

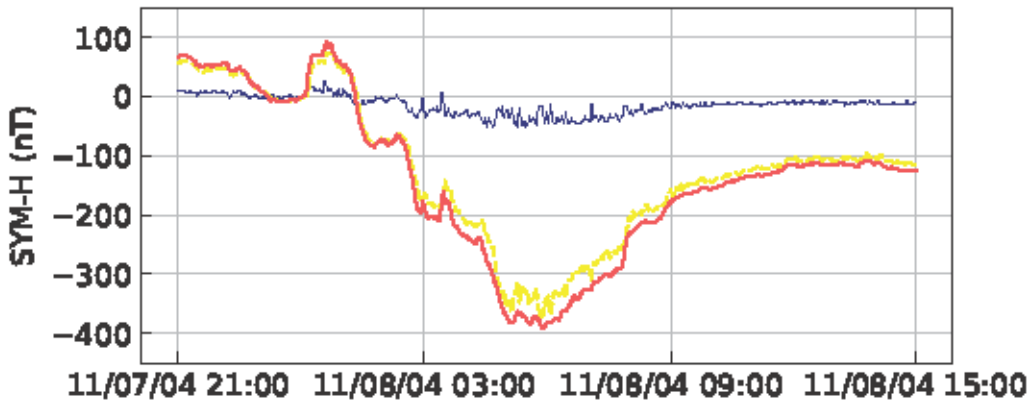
1h ahead prediction



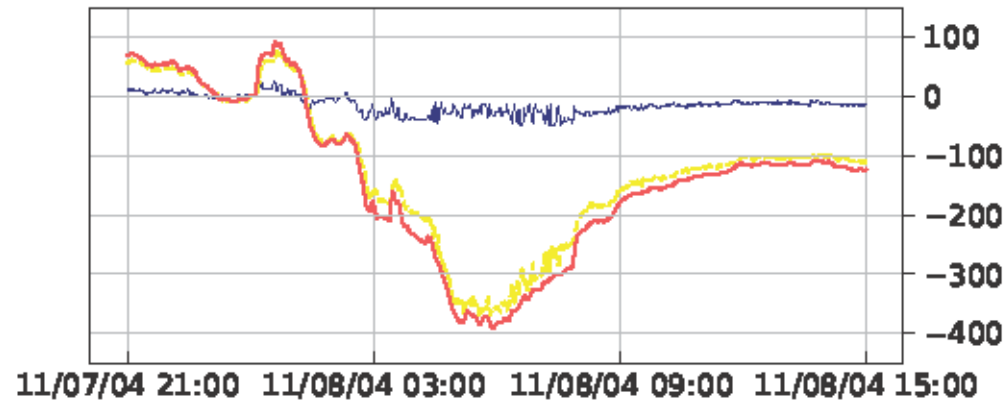
2h ahead prediction



1h ahead prediction

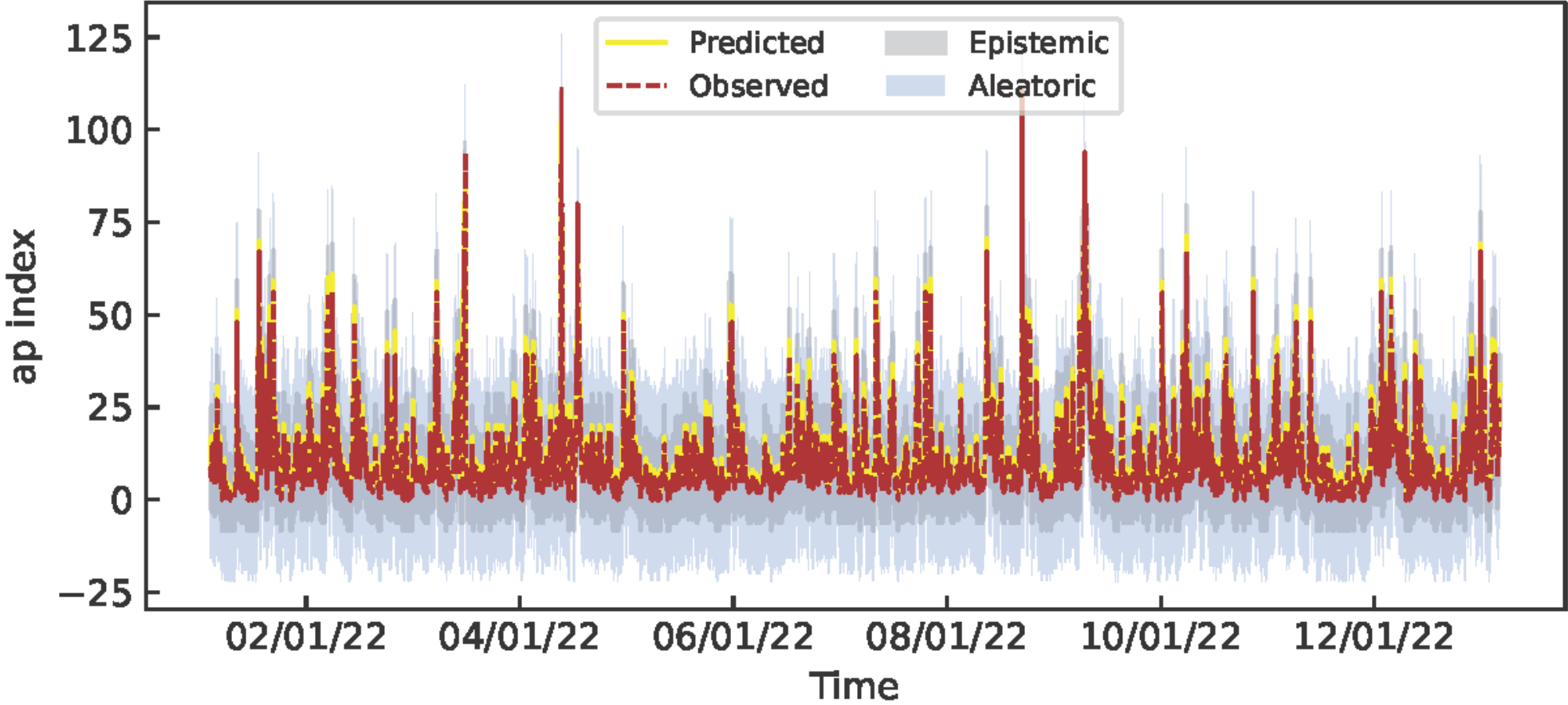


2h ahead prediction

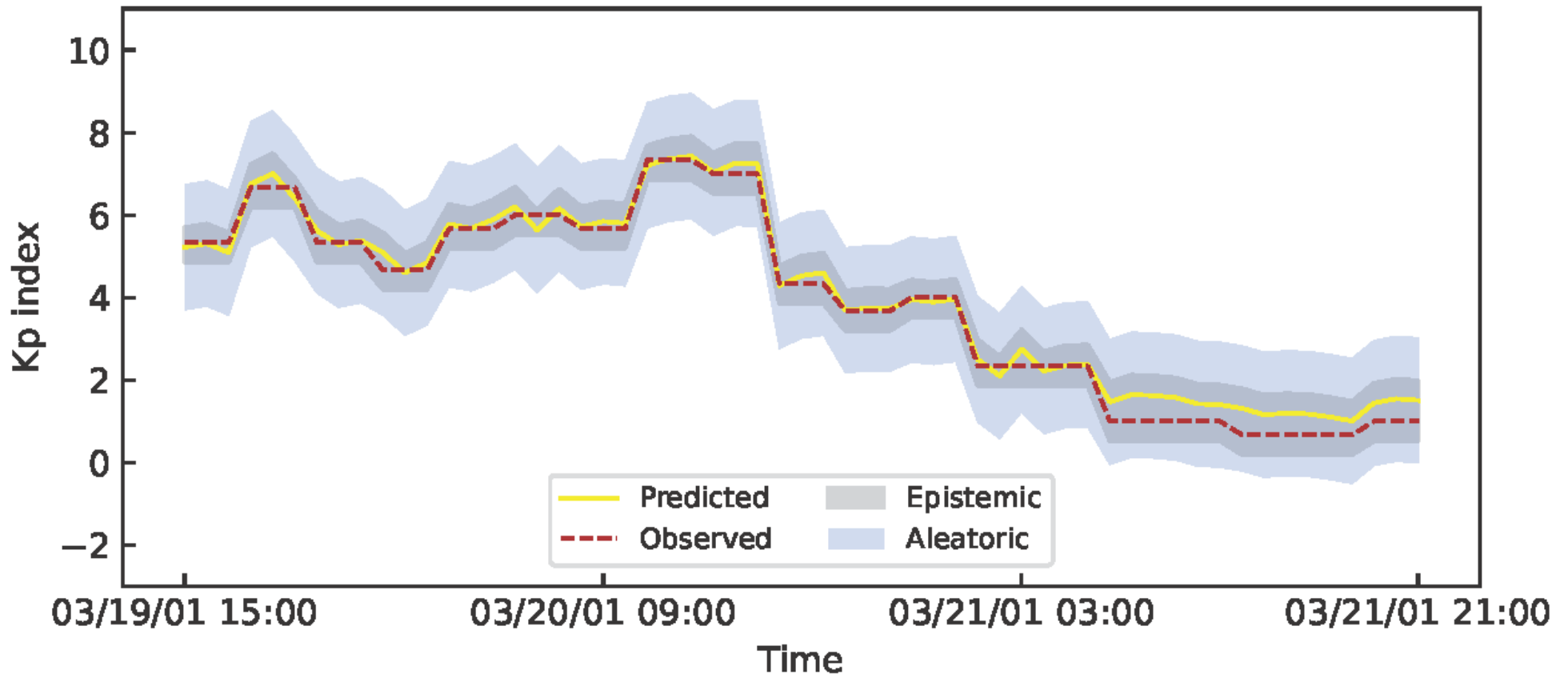


Observed SYM-H Predicted SYM-H Prediction Error

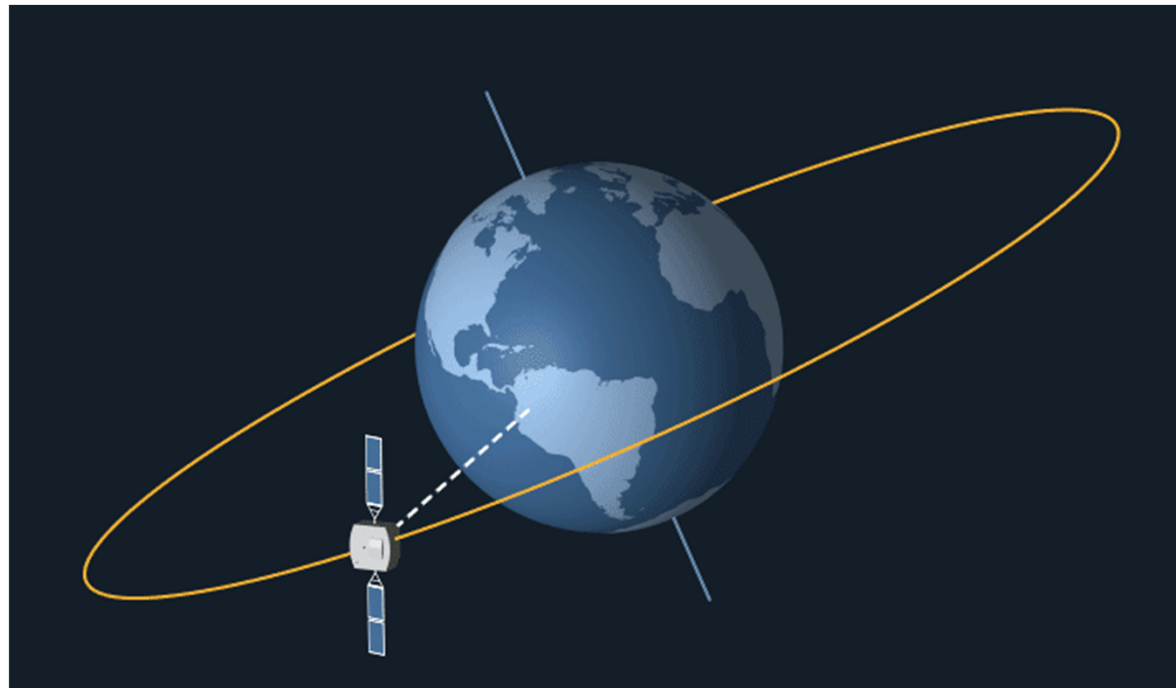
9h ahead prediction



9h ahead prediction



Investigating how space weather impacts satellite drag and orbits



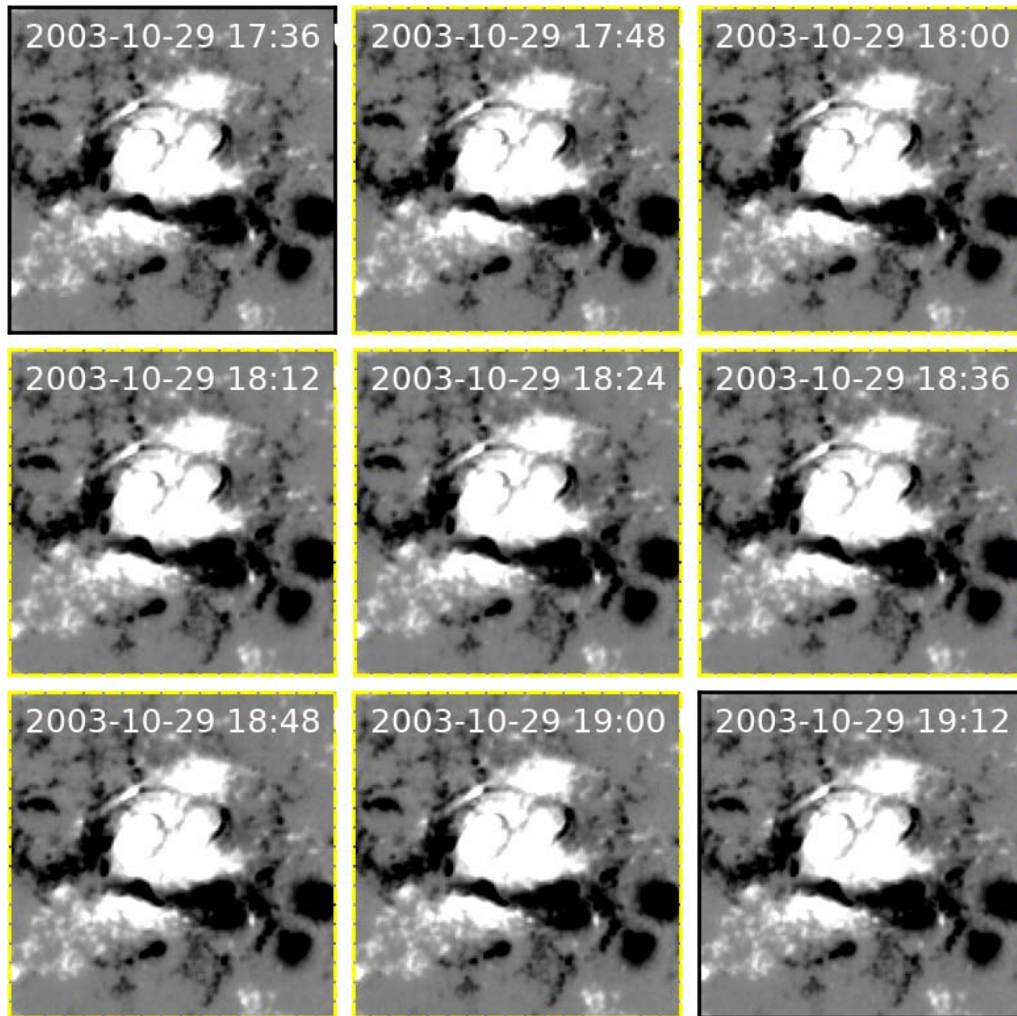
Source: NASA

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

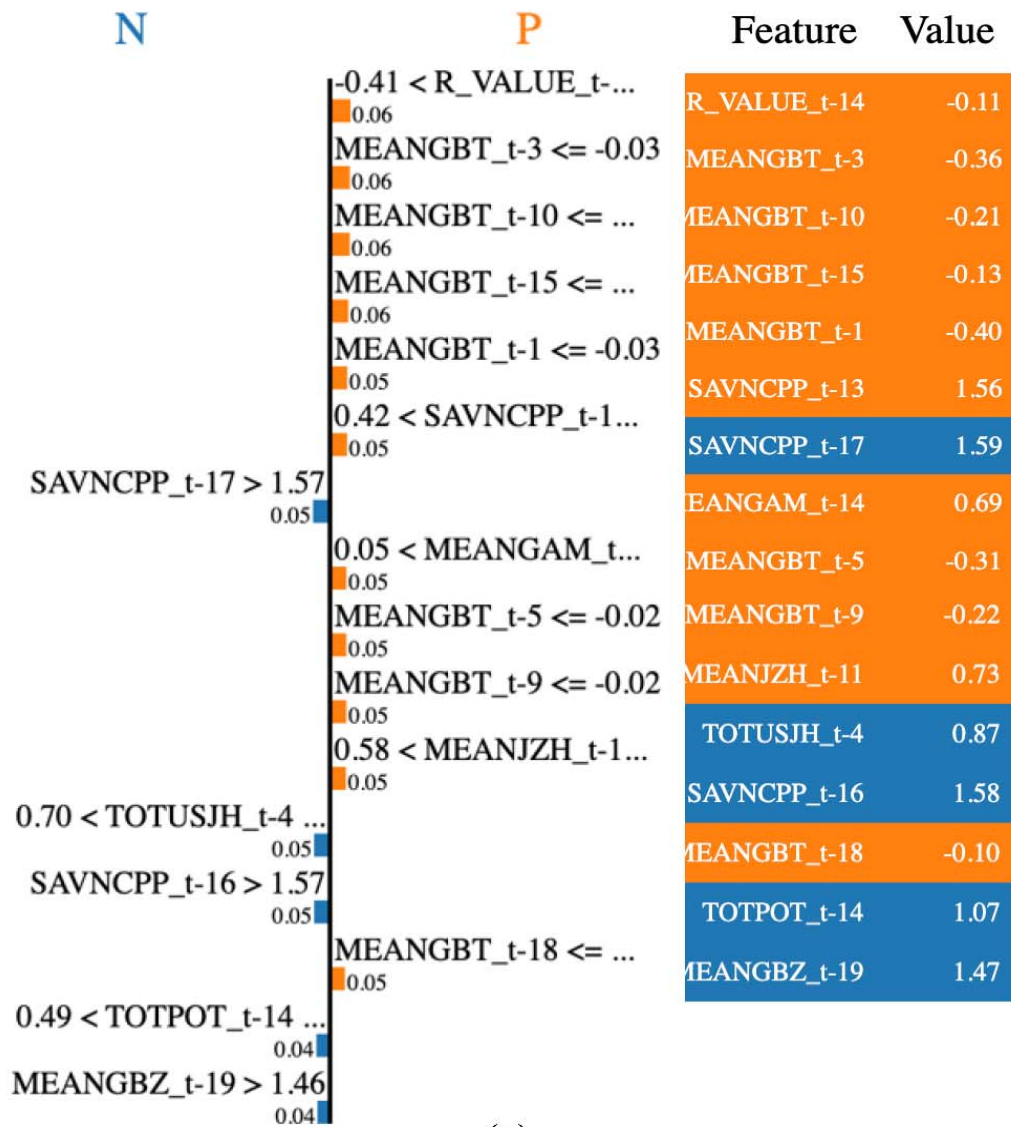
Global interpretation (provide explanation based on the whole test set)

For example, a home loan is denied (reasons: bad credit, short credit history, insufficient income)

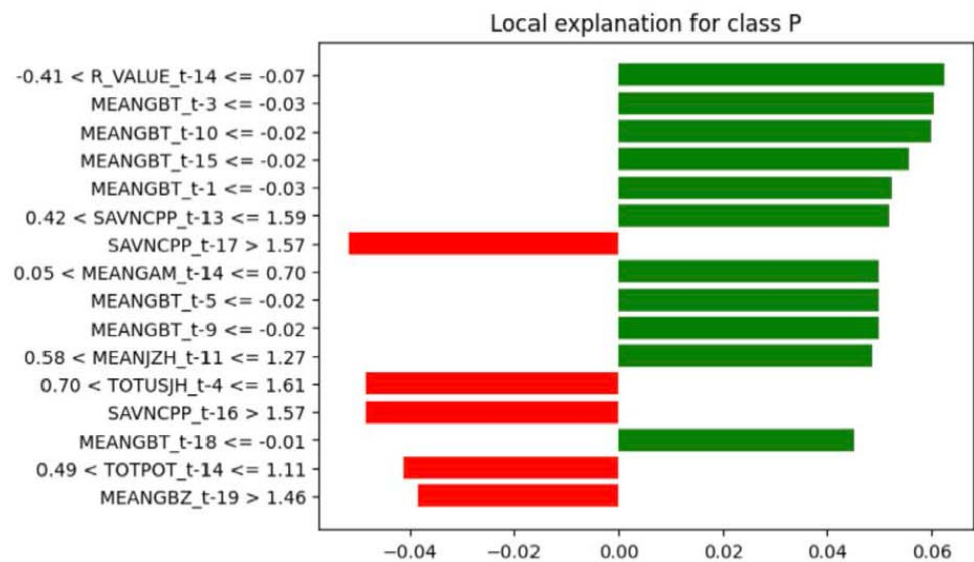
Local interpretation (provide explanation based on a specific test instance)

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.



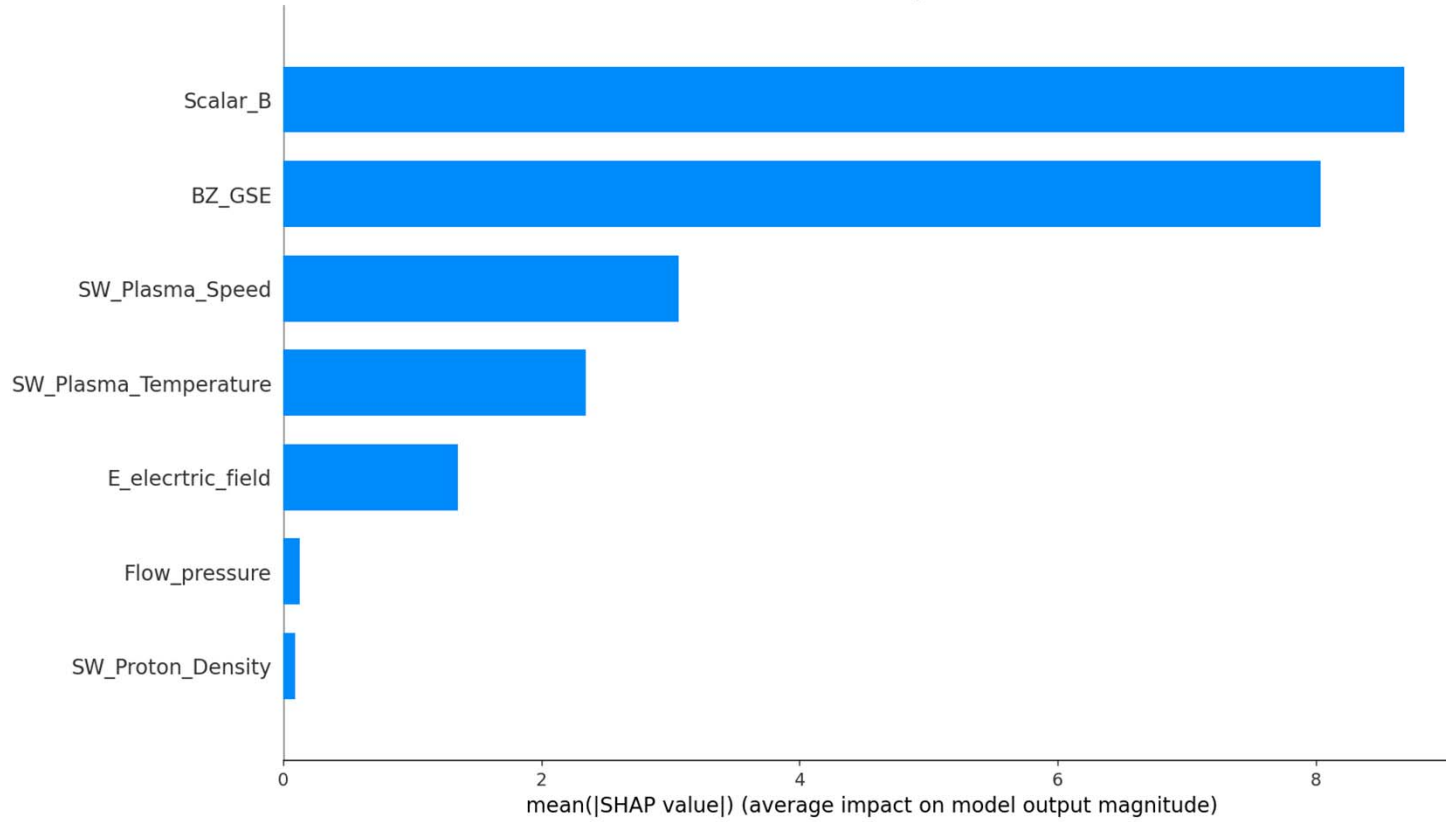
(a)



(b)

(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

SHAP Summary Plot



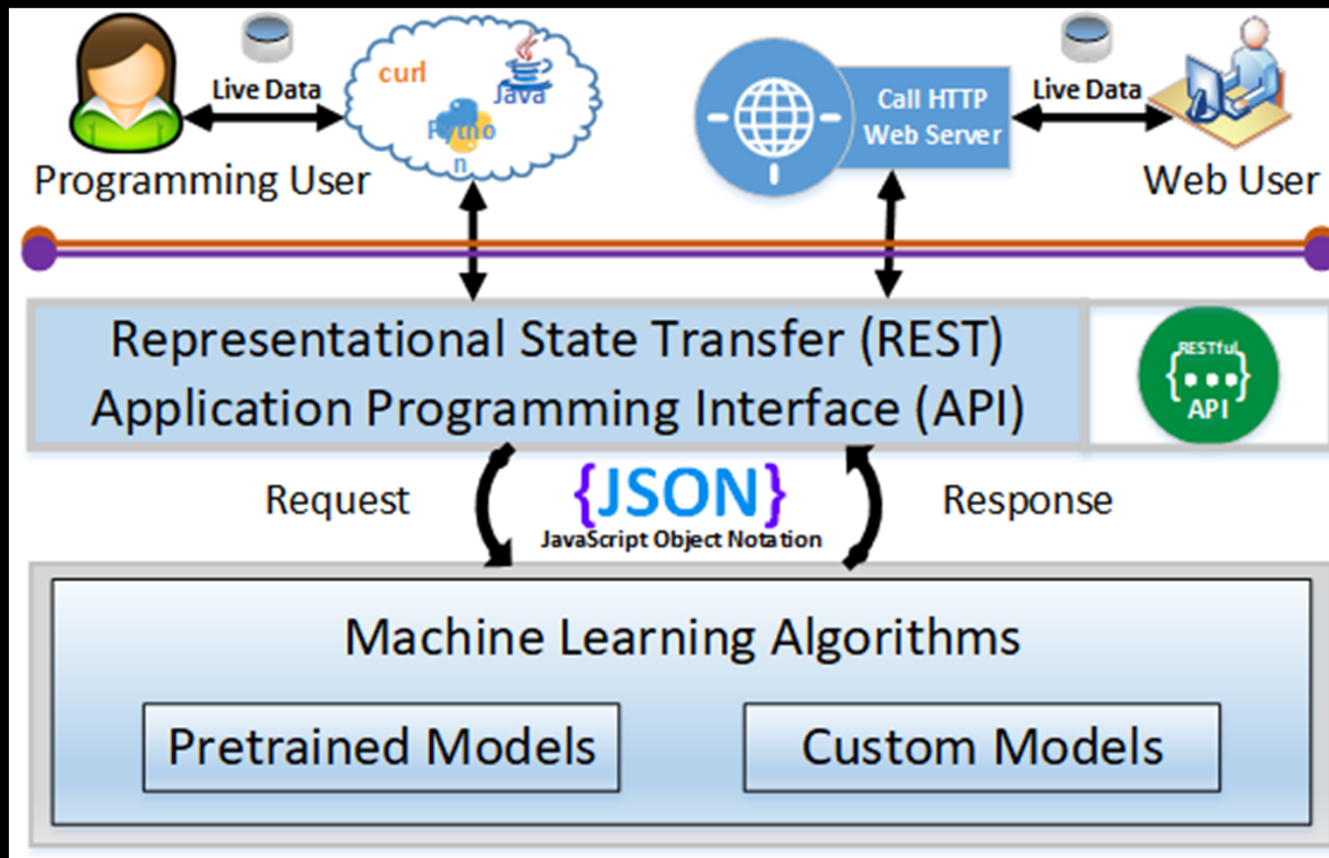
(Global Interpretation) Use an XAI tool, SHAP, to explain/rank the importance of 7 solar wind parameters in the input for Disturbance Storm Time (Dst) index prediction

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

The screenshot shows a web browser window displaying the SolarDB website. The browser's address bar shows the URL `nature.njit.edu/solardb/`. The website header includes the NJIT logo and the title "SolarDB: A Cyberinfrastructure for Advancing Space Weather Research". A navigation menu contains tabs for "About", "Databases", "Tools", "Data Sources", "Help", and "Contacts".

The main content area features a slide titled "Impact of Space Weather Events" with the text: "Technologies and infrastructure are affected by space weather events (Credit: NASA)." To the right of this text is a diagram illustrating the effects of space weather. The diagram shows a cross-section of Earth's atmosphere and space. Labels include: "Energetic Electrons" (pointing to the upper atmosphere), "Coronal Mass Ejections (CMEs)" (pointing to the solar wind), "Solar Flare Emissions" (pointing to the sun), "GPS Signal Scintillation" (pointing to a satellite), "Radiation Effects on Aircraft" (pointing to an airplane), "Geomagnetically Induced Currents in Power Systems" (pointing to power lines), "Radiation Effects on Satellites" (pointing to a satellite), "Subsolar Effects on Submarine Cables" (pointing to a cable), and "Turbulent Currents in Pipelines" (pointing to a pipeline). A "Space Weather" label is also present.

Below the diagram is a "Project Description" section. The text reads: "This cyberinfrastructure, dubbed SolarDB, is a comprehensive resource accessible on the Internet. It comprises databases, tools, and computing services for advancing space weather research. These databases and tools play an important role in analyzing adverse space weather effects and predicting extreme space weather events. The data in SolarDB are gathered from ground-based observatories and satellite missions. Major data sources include SDO/HMI, SDO/AIA, Global H-alpha Network, and Big Bear Solar Observatory data archives." The section concludes with logos for NSTL and NASA.

The Windows taskbar at the bottom shows the search bar, several application icons, and system information: "77°F Mostly cloudy", "11:27 PM", and "7/8/2024".

Cyberinfrastructure for Advancing Space Weather Research

SolarDB: A Cyberinfrastructure for... x | https://nature.njit.edu/nfs/solardb/ x | +

nature.njit.edu/solardb/

Flare Database with M5.0 or Larger Flares

Please select the flare type:

Significant flares Control data

Date filter

Active region filter

Flare filter

Submit Query

Active region number: 11890 X3.3 flare's starting time: 2013-11-05 22:07:00

Please select one of the following data sources:

Solar Dynamics Observatory (SDO)

Please select the SDO image type:

aia.euv aia.uv cea hmi.B hmi.M

Back Submit

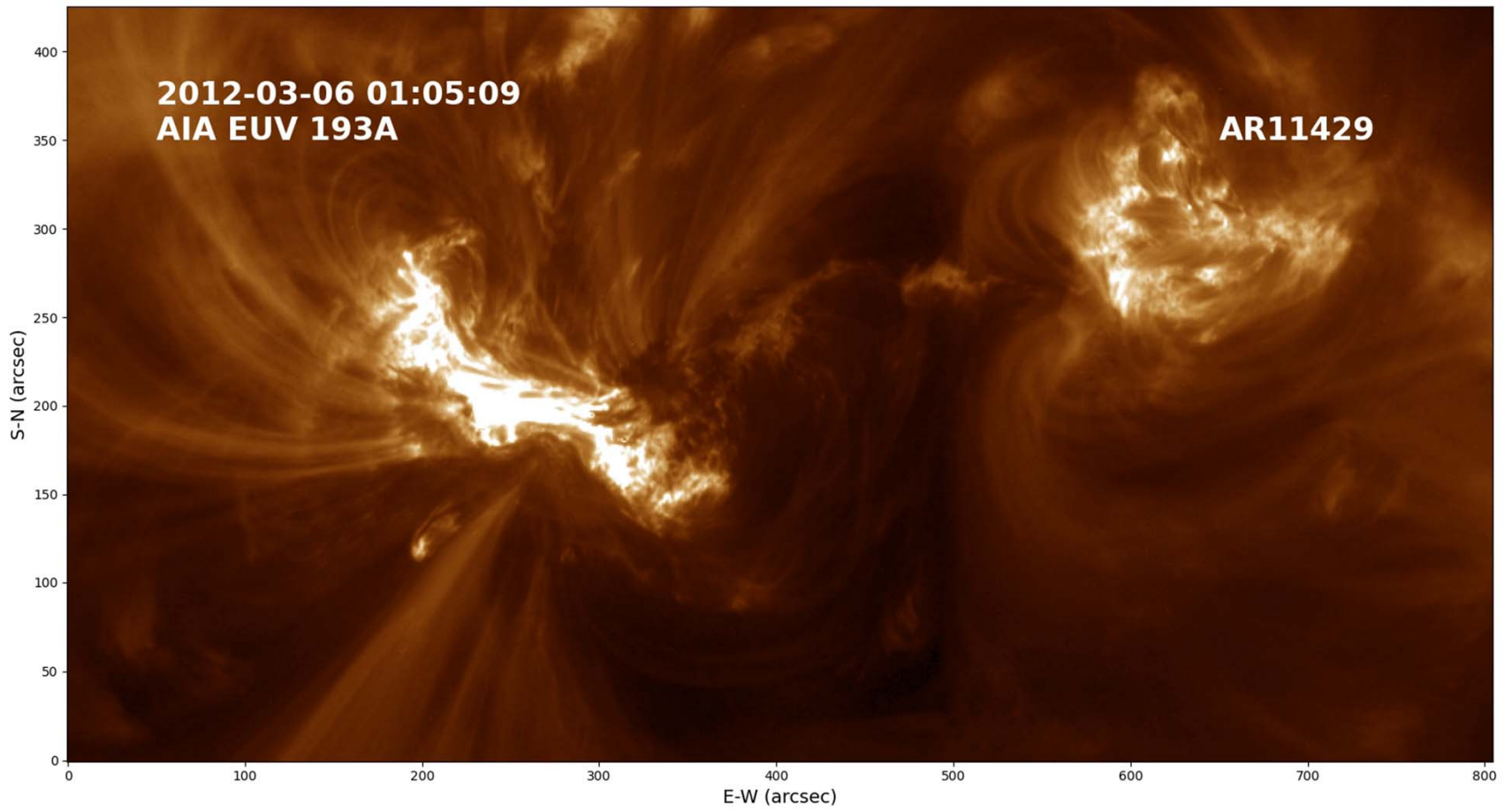
Watch quick look movie Download all FITS files

Number of files found: 1672

hmi.sharp_cea_720s.3341.20131104_221200_TAI.Bp.fits	hmi.sharp_cea_720s.3341.20131104_221200_TAI.Bp_err.fits
hmi.sharp_cea_720s.3341.20131104_221200_TAI.Br.fits	hmi.sharp_cea_720s.3341.20131104_221200_TAI.Br_err.fits
hmi.sharp_cea_720s.3341.20131104_221200_TAI.Bt.fits	hmi.sharp_cea_720s.3341.20131104_221200_TAI.Bt_err.fits
hmi.sharp_cea_720s.3341.20131104_221200_TAI.Dopplergram.fits	hmi.sharp_cea_720s.3341.20131104_221200_TAI.bitmap.fits
hmi.sharp_cea_720s.3341.20131104_221200_TAI.conf_disambig.fits	hmi.sharp_cea_720s.3341.20131104_221200_TAI.continuum.fits
hmi.sharp_cea_720s.3341.20131104_221200_TAI.magnetogram.fits	hmi.sharp_cea_720s.3341.20131104_222400_TAI.Bp.fits
hmi.sharp_cea_720s.3341.20131104_222400_TAI.Bp_err.fits	hmi.sharp_cea_720s.3341.20131104_222400_TAI.Br.fits
hmi.sharp_cea_720s.3341.20131104_222400_TAI.Br_err.fits	hmi.sharp_cea_720s.3341.20131104_222400_TAI.Bt.fits
hmi.sharp_cea_720s.3341.20131104_222400_TAI.Bt_err.fits	hmi.sharp_cea_720s.3341.20131104_222400_TAI.Dopplergram.fits

Type here to search

63°F Sunny 10:03 AM 9/26/2021





SolarDB: A Cyberinfrastructure for Advancing Space Weather Research



- About
- Databases
- Tools
- Data Sources
- Help
- Contacts

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 records
 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

The screenshot shows the SolarDB website interface. At the top, the NJIT logo and the title "SolarDB: A Cyberinfrastructure for Advancing Space Weather Research" are visible. A navigation bar includes links for "About", "Databases", "Tools", "Data Sources", "Help", and "Contacts". A central panel titled "Global H-alpha Network" offers options: "Latest images" (selected), "Synoptic maps", "Stations", "Instruments", and "Daily Images". Below this, a grid of 16 solar H-alpha images is displayed, each with its name, a timestamp, and a date. The observatories listed are: Big Bear Solar Observatory, Observatório de Coimbra, Observatoire de Paris Meudon, Uccle Solar Equatorial Table, Kanzelhöhe Solar Observatory, Catania Astrophysical Observatory, Observatoire Pic Du Midi, and Huairou Solar Observatory. The Windows taskbar at the bottom shows the system time as 11:34 PM on 7/8/2024.

Observatory	Image 1 Time	Image 1 Date	Image 2 Time	Image 2 Date
Big Bear Solar Observatory	20:54:13 UT	Jul 08, 2024	20:54:13 UT	Jul 08, 2024
Observatório de Coimbra	16:16:18 UT	Aug 09, 2021	16:16:18 UT	Aug 09, 2021
Observatoire de Paris Meudon	06:14:09 UT	Jul 08, 2024	06:14:09 UT	Jul 08, 2024
Uccle Solar Equatorial Table	09:33:44 UT	Jul 08, 2024	09:33:44 UT	Jul 08, 2024
Kanzelhöhe Solar Observatory	07:38:44Z UT	Jul 08, 2024	07:38:44Z UT	Jul 08, 2024
Catania Astrophysical Observatory	11:45:00 UT	Jul 08, 2024	11:45:00 UT	Jul 08, 2024
Observatoire Pic Du Midi	12:05:06 UT	Jul 08, 2024	12:05:06 UT	Jul 08, 2024
Huairou Solar Observatory	06:31:00 UT	Jul 07, 2024	06:31:00 UT	Jul 07, 2024

[Abduallah et al. Scientific Reports 2023]

The screenshot displays the SolarDB: A Cyberinfrastructure website. The browser address bar shows the URL nature.njit.edu/solardb/. The website has a navigation menu with links for About, Databases, Tools, Data Sources, Help, and Contacts. The main content area is titled "Flare Forecasting System" and includes links for 24-Hour, 48-Hour, and 72-Hour flare forecasts. It provides the most recent update time (2024-06-06 00:01:19 UTC) and the most recent time of SHARP data at JSOC (2024-06-05 23:58:45 UTC). A table lists active regions with their HARP numbers and probabilities for flares of different classes (C, M, M5) within the next 24 hours. Below the table, there are footnotes explaining that regions near the limb or without SHARP data have no prediction. At the bottom, there is a radio button option to view previous flare predictions, which is currently set to "No".

Flare Forecasting System

[24-Hour Flare Forecast](#) [48-Hour Flare Forecast](#) [72-Hour Flare Forecast](#)

Most recent update time: 2024-06-06 00:01:19 UTC
Most recent time of SHARP data at JSOC: 2024-06-05 23:58:45 UTC

Flares that would occur within the next 24 hours

Active region number	HARP number	>=C-class flare probability	>=M-class flare probability	>=M5-class flare probability
13690~	N/A	N/A	N/A	N/A
13691*	10128	N/A	N/A	N/A
13694~	N/A	N/A	N/A	N/A
13695~	N/A	N/A	N/A	N/A
13696~	N/A	N/A	N/A	N/A
13697	10149	18.44%	12.53%	3.37%
13698	10153	0.75%	0.3%	0.18%
13699*	10161	N/A	N/A	N/A
13700*	10164	N/A	N/A	N/A
13701	10162	82.94%	19.61%	0.09%
13702	10166	0.79%	0.29%	0.1%
13703	10169	99.71%	99.71%	0.75%
13704	10168	35.27%	15.03%	0.1%
13705~	N/A	N/A	N/A	N/A
13706	10171	13.14%	0.34%	0.09%

* This active region is near the limb, and hence no prediction is made on this active region.
~ This active region does not have SHARP data, and hence no prediction is made on this active region.
Please note that the information concerning active regions and SHARP data used to calculate the probabilities in the above table is based on the active regions in [2024-06-05](#).

Do you want to view the previous flare predictions? Yes No

Home

Community Coordinated Software Center

These Jupyter notebooks are now available on GitHub and Zenodo for interested researchers to view and execute.

<p><u>Identifying and Tracking Solar Magnetic Flux Elements with Deep Learning</u></p> <p>Citation</p>	<p><u>Predicting Solar Flares with Machine Learning</u></p> <p>Citation</p>	<p><u>Predicting Coronal Mass Ejections Using SDO/HMI Vector Magnetic Data Products and Recurrent Neural Networks</u></p> <p>Citation</p>	<p><u>Predicting Solar Flares Using a Long Short-term Memory Network</u></p> <p>Citation</p>
<p><u>Predicting Solar Energetic Particles Using SDO/HMI Vector Magnetic Data Products and a Bidirectional LSTM Network</u></p> <p>Citation</p>	<p><u>Forecasting the Disturbance Storm Time Index with Bayesian Deep Learning</u></p> <p>Citation</p>	<p><u>Reconstruction of Total Solar Irradiance by Deep Learning</u></p> <p>Citation</p>	<p><u>A Transformer-Based Framework for Geomagnetic Activity Prediction</u></p> <p>Citation</p>



Communities My dashboard

Log in Sign up

Published June 29, 2022 | Version v1.1

Software Open

Forecasting the Disturbance Storm Time Index with Bayesian Deep Learning

Yasser Abdullallah¹; Jason T. L. Wang¹; Prianka Bose¹; Genwei Zhang¹; Firas Gerges¹; Haimin Wang²

Show affiliations

Forecasting the Disturbance Storm Time Index with Bayesian Deep Learning

Files

Dst-prediction-v1.1.pdf

Page: 1 of 40 Automatic Zoom

10/17/22, 3:11 AM YA_01_ForecastingDstIndexBayesianDeepLearning

Forecasting the Disturbance Storm Time Index with Bayesian Deep Learning

Table of Contents

516 VIEWS 387 DOWNLOADS

Show more details

Versions

Version v1.1	Jun 29, 2022
10.5281/zenodo.7529705	
Version v1.1	Jun 29, 2022
10.5281/zenodo.7516651	
Version v1.1	Jun 29, 2022
10.5281/zenodo.7514481	
Version v1.1	Jun 29, 2022
10.5281/zenodo.7514431	
Version v1.1	Jun 29, 2022
10.5281/zenodo.6774658	

View all 6 versions

Cite all versions? You can cite all versions by using the DOI 10.5281/zenodo.6757941. This DOI

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