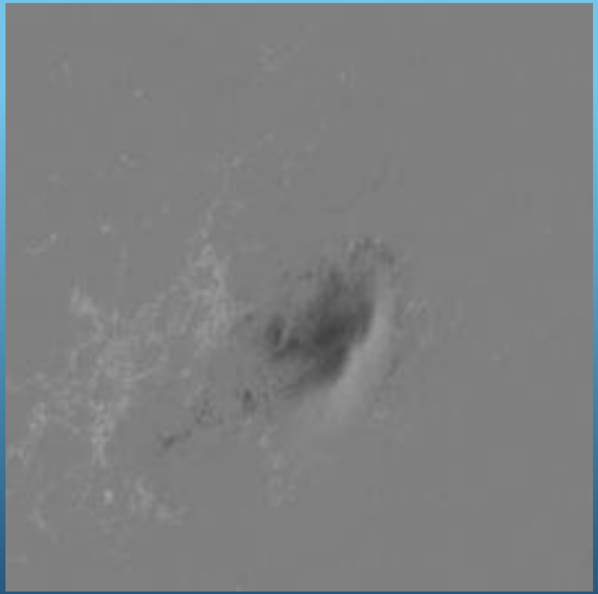
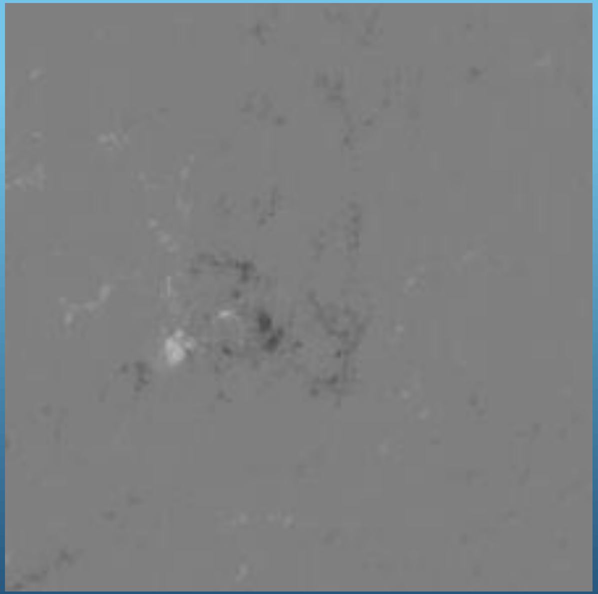
Solar Flare Prediction and Active Region Monitoring Using Multimodal Machine Learning



By: Zaid Abuhassan and Jonathan Davis



PROBLEM WITH EXISTING SOLAR FLARE PREDICTION

Problem:

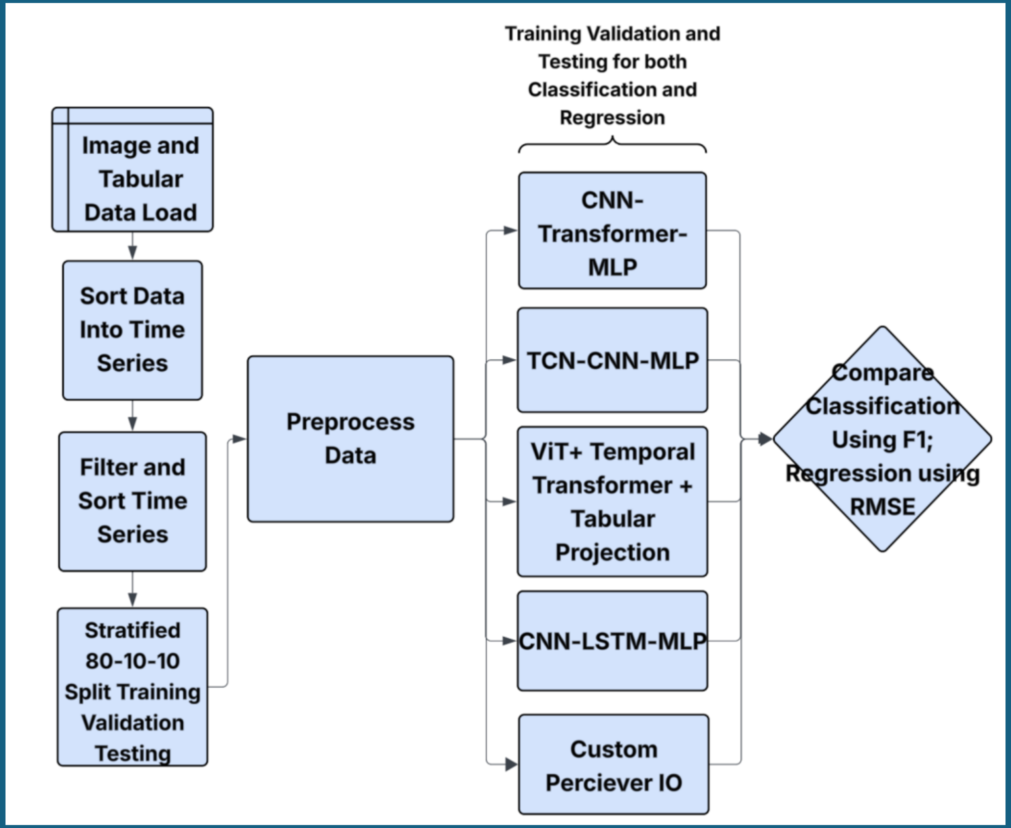
- In 2024 a large geomagnetic storm caused **severe damage** to power grids, GPS and radio interruptions, and satellite disruptions
- Solar flares are direct precursors to similar damaging geomagnetic storms
- Current solar flare detection technology identifies if a solar flare will occur, **not how the active region develops**
- Understanding and characterizing the development of active regions into solar flares would provide astronomers with **greater understanding of solar weather** and help Earth prepare for the effects

Prior solutions do not address the **intermediate development of active regions**

Engineering Goal: Develop models that forecasts **solar flares** and characterizes the **development of active regions** through the prediction of active region magnetic field features with **strong performance.**

ENGINEERING METHODOLOGY

PROJECT OUTLINE



Pre-Processing:

1. Upload the magnetogram data to the Amazon Web Services
2. Sort Magnetograms into time series, filter by time series length and cut into 24-hour period and perform a stratified 80-10-10 training-validation-testing split, ensuring that class counts remain roughly equal in each.
3. Normalize and preprocess all magnetogram frames
4. Split the 24 hour windows into 10.5 and 13.5 hour-long sections

Training

5. Use data magnetogram time series and corresponding tabular data within the 10.5 hour time section to train classification models to forecast the class of solar flares within the next 13.5 hours.
6. Repeat for the regression models to train them for predicting the numerical values at 1 hour intervals after the end of the 10.5 hour section.

Post-Training

1. Test all models **over 160 trials** and use the macro average performance

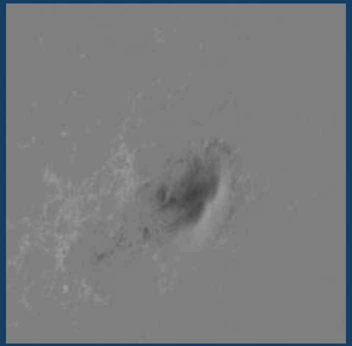
a. Compare the solar flare forecasting and active region monitoring models

b. Identify the best overall model for both solar flare forecasting and active region monitoring

PREPROCESSING OF DATA

- Sort magnetograms and corresponding tabular data by active region
- Cut the data for each active region to a 24-hour period
- Series with not enough images filtered out and series with too much images were cut down
- Split the data for each 24-hour window into a 10.5-hour background window and a 13.5-hour prediction window
- Split the data into a training, validation, and testing sets using a stratified split, thereby ensuring that relative class counts remain consistent through the splits.
- Normalize the magnetograms using Imagnet normalization weights

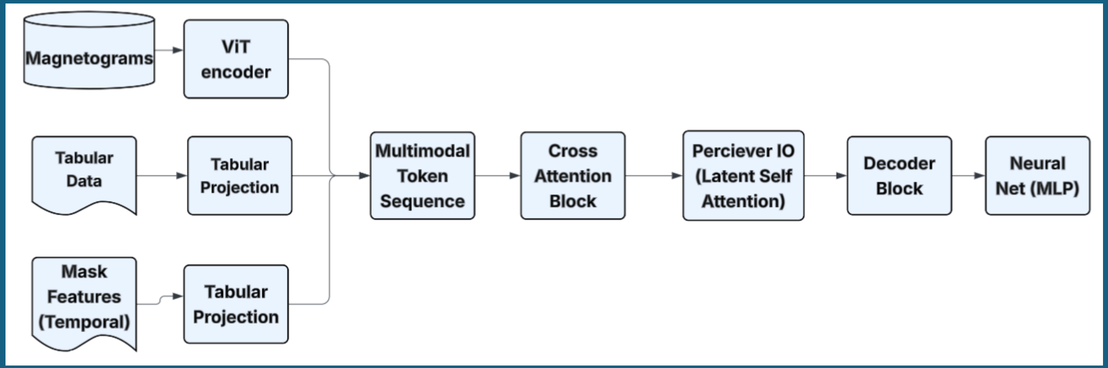
What is a magnetogram?



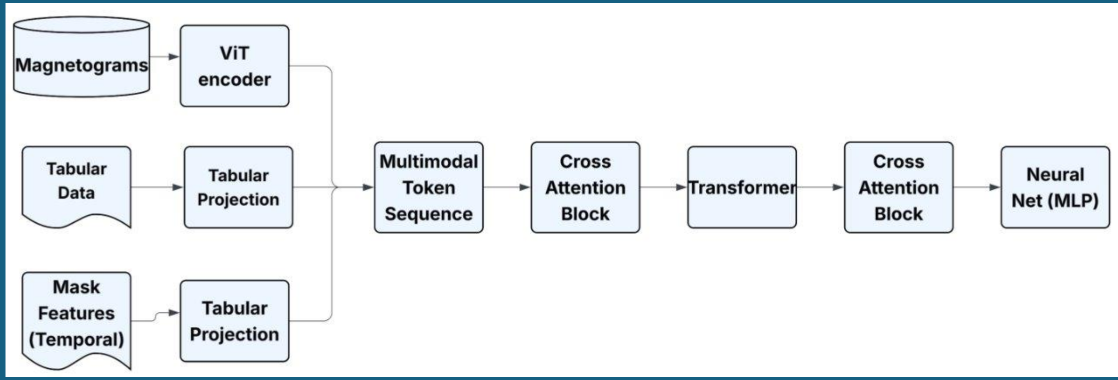
Magnetogram: Image that quantifies the strength of the magnetic field of the sun's surface at each point in the image

Multi-Modal Machine Learning Model Architecture Development

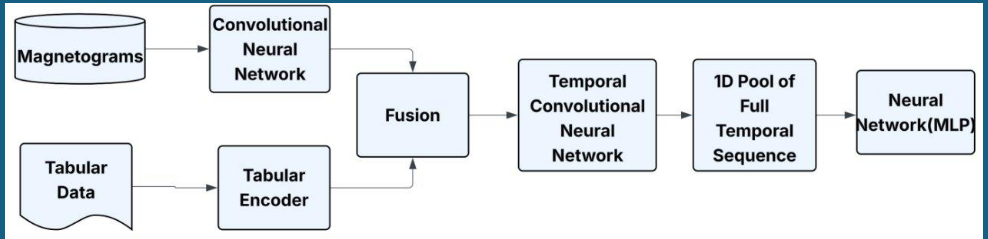
Custom Perceiver IO:



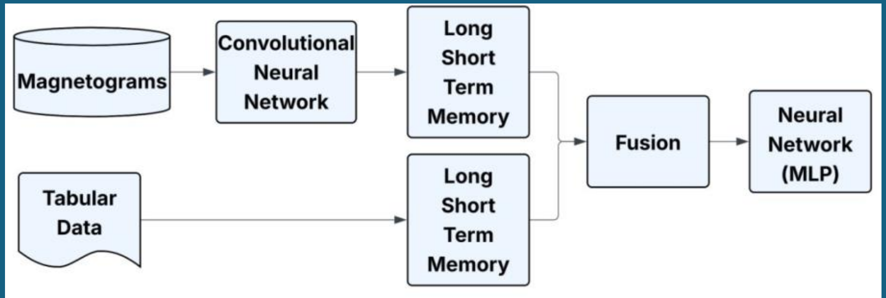
Vision Transformer



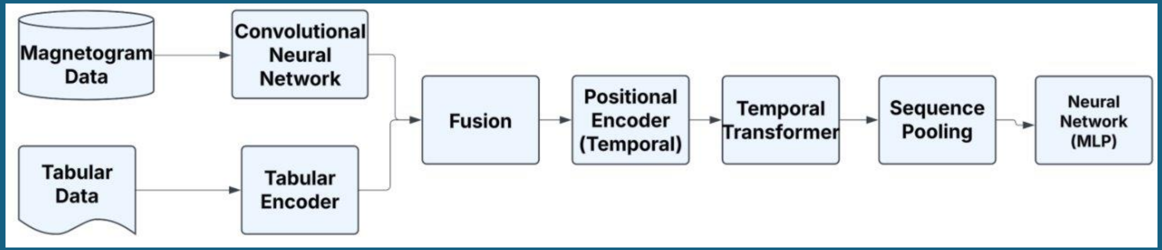
Convolutional Neural Network + Temporal Convolutional Neural Network



Convolutional Neural Network + Long Short Term Memory



Convolutional Neural Network + Transformer



COMPARISON OF MODEL PERFORMANCE

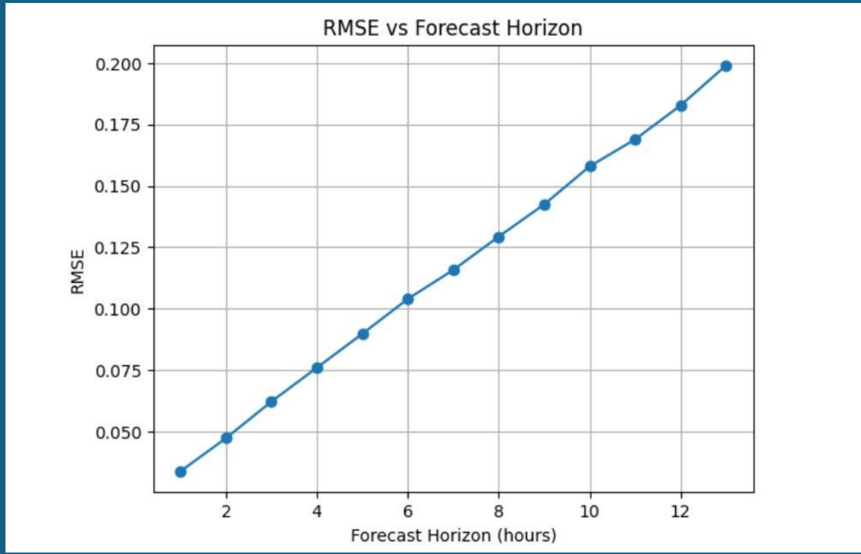
F1: Metric that holistically evaluates model performance by combining both specificity and sensitivity metrics for solar flare forecasting

RMSE (Root Mean Squared Error): Quantifies the error of the models in predicting the numerical active region characteristics

Model Name	CNN-LSTM	TCN-CNN	Vision Transformer	CNN-Transformer	Custom Perciever IO
F1 score (Solar Flare Forecasting)	0.65333	0.6713	0.6494	0.6662	0.7128
RMSE (Active Region Monitoring)	0.1202	0.129	0.128	0.1267	0.1382
			: best performance		

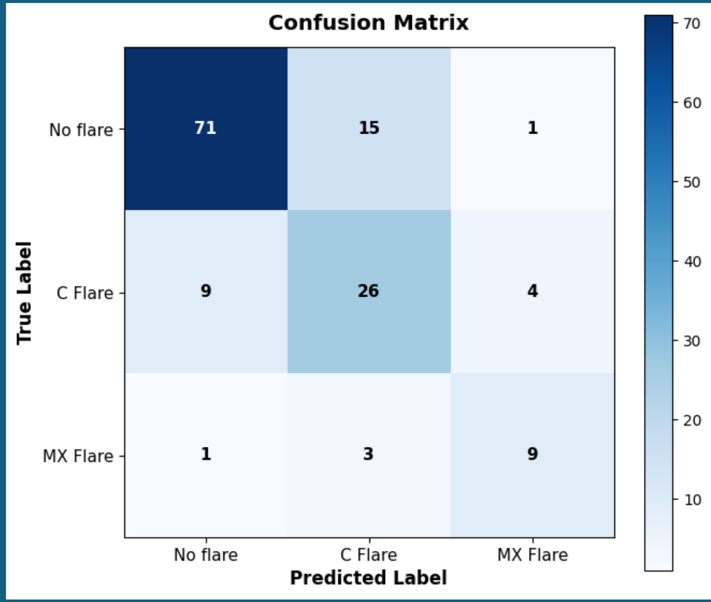
RESULTS FOR BEST PERFORMING MODELS

Active Region Monitoring: CNN-LSTM



	Performance
MAE	0.0681
RMSE	0.1202
R^2	0.9886

Solar Flare Forecasting: Custom Perceiver IO



	Performance
ROCAUC macro OVR	0.8731
TSS	0.6125
HSS	0.5789

CONCLUSION

The PerceiverIO was the best model for forecasting solar flares, showing the strongest performance in classification. In testing, it achieved strong performance with a ROCAUC macro OVR of 0.8761, a TSS of 0.6125, and an HSS of 0.5789. The CNN-LSTM was the best model for monitoring active regions via characteristic prediction, displaying the strongest regression performance with an MAE of 0.0681, an RMSE of 0.1202, and an R^2 of 0.9886.

The research also introduces the novel use of multimodal models for the hourly prediction of active region (AR) characteristics.

APPLICATIONS

- Machine learning models were able to forecast solar flares and prediction active region characteristics, **meeting all criteria** of the engineering goal
- Can be utilized by researchers to find trends and better understand solar flare development from active regions
- Can help astronomers **better anticipate solar flares** and coronal mass ejections, providing **more time to prepare sensitive infrastructure**
- Introduces machine learning for active region monitoring via numerical parameters, **encouraging further research**

FURTHER IMPROVEMENTS

- CME forecasting and risk assessment to predict geomagnetic storm strength with improved accuracy using active region characteristic forecasting
- Integration with active region identification and CME prediction for end-to-end earth threat analysis
- Utilize additional, novel machine learning model types to improve performance and efficiency of the machine learning application
- Build into app-based tool that provides easy, accessible use to astronomers for studying solar flare development
- Use forecasted active region parameters to train a magnetogram generation model to provide a forecast for the entire active region

REFERENCES

Boucheron, L. E., Vincent, T., Grajeda, J. A., & Wuest, E. (2023). Solar active region magnetogram image dataset for studies of space weather. *Scientific Data*, *10*(1). https://doi.org/10.1038/s41597-023-02628-8

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*(2), 121. https://doi.org/10.3847/1538-4357/ab1b3c

NASA. (2024, September 13). *Solar Storms and Flares - NASA Science*. Nasa.gov. https://science.nasa.gov/sun/solar-storms-and-flares/

National Oceanic and Atmospheric Administration. (n.d.). *Solar Flares (Radio Blackouts) | NOAA / NWS Space Weather Prediction Center*. Www.swpc.noaa.gov. https://www.swpc.noaa.gov/phenomena/solar-flares-radio-blackouts

Newman, T. S., Hall, C. W., Farris, L., Singh, T., Pogorelov, N., Benson, B., Raza, S., & Trital, P. (2025). Solar Flare Forecasting Using Machine Learning and SDO/HMI Data: A Multiple Machine Learning Model and Data Curation Technique Comparison Study. *The Astrophysical Journal Supplement Series*, *280*(2), 54. https://doi.org/10.3847/1538-4365/adf8e0