ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume X-5/W2-2025
Unleashing the power of Geospatial & Frontier Technologies for a Sustainable Future – GIFTS Summit 2025,
1–3 September 2025, Symbiosis International University, Pune, India

AI-Based Space Weather Prediction for Satellite Protection

Rachana Ramesh Shirke, Veena Parihar
Department of Data Science and Spatial Analytics,
Symbiosis institute of geoinformatics, Symbiosis International (Deemed University), Pune
Emails: rachanashirke910@gmail.com, veena2parihar@gmail.com

Keywords: Solar Flares, Coronal Mass Ejections (CMEs), Geomagnetic Storm Prediction, Space Weather Forecasting, CNN-LSTM Deep Learning Model, Satellite Risk Classification.

Abstract

Satellite operations and other space assets are gravely impaired by space weather disturbances, but existing forecasting systems often lack accuracy and integrated multi-objective prediction functionality. This paper suggests a unified AI-based multi-objective system for space weather prediction to enhance protection of satellites from solar-terrestrial disturbances. Five prediction tasks are addressed by the architecture: 1. solar active region classification; 2. solar flare prediction; 3. prediction of coronal mass ejection travel time; 4. Kp-index-based prediction of geomagnetic storms; and 5. satellite danger level classification. Multi-source inputs such as video sequences, heliophysical observations, magnetograms, and solar images were applied in developing and evaluating expert machine learning and deep learning models. The CNN+LSTM model for predicting flares had 0.67 accuracy and 0.47 F1-score, with good recall for quiet-class events and poor recall for flare events because of class imbalance. The system achieved 0.97 accuracy and 0.96 F1-score for active region classification. LightGBM performed better than XGBoost for CME trip time, with an R^2 of 0.86 and RMSE of 1.52 hours. XGBoost outperformed LightGBM in the prediction of Kp index (R^2 = 0.82, RMSE = 0.63), while LightGBM showed lower performance (R^2 = 0.67). The XGBoost classifier delivered strong multi-class performance with 0.999 accuracy and 0.87 F1-score for satellite risk level classification. The models demonstrated stability and robustness through satisfactory generalization over Solar Cycles 23 and 24. The five models are integrated for real-time application with a Gradio-based interface. Focus will be on enhancing flare detection and addressing class imbalance.

1. Introduction

Space weather events such as solar flares, CMEs, and geomagnetic storms can have an enormously negative impact on space-based technology and satellite operations. These events could create serious technology and economic problems, which could compromise navigation, communications architectures, and other systems and infrastructure dependent on space (Gonzalez et al., 1994). Given the complexity and the evolving nature of space weather, sophisticated forecasting techniques must be implemented to ameliorate the risks posed by space weather. Because space weather processes, on a multi-scale coupled and non-linear level, largely inhibit the use of formal forecasting techniques are limited (Dorelli et al., 2022 & Guastavino et al., 2025). But, the growing quantity of high-resolution data from satellite sources and computing capacity provides improved possibility for data driven forecasting techniques and methodologies to improve predictive accuracy (Abed et al., 2021). Accurate forecasting of space weather is paramount for upholding the integrity of any space-dependent systems or the protection of any satellites serving an intended purpose. This requirement is timely because there is an urgent need to have robust models that can unravel spade weather complexities in regard to data and provide timely warnings to mitigate risk (Dey, S., 2018 & Jovanovic et al., 2024).

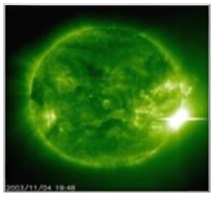

Fig. 1: An image of a solar flare taken by the Extreme Ultraviolet Imaging Telescope (195Å) on the SOHO spacecraft on 4 November 2003. Solar flares are intense emissions of solar radiation related to sunspot regions and magnetic activity (ESA & NASA/SOHO, 2015).

Active solar areas are where magnetic reconnection generates solar flares, emitting UV and X-ray radiation. Blackouts in HF propagation, GPS faults, radiation exposure, and satellite malfunction are some of their effects (Temmer (2021)). CMEs are large plasma and magnetic field expulsions, but they tend to go along with flares, which may have the potential to modify Earth's magnetosphere and result in geomagnetic storms. Intensity of the X-ray categorizes solar flares as A, B, C, M, and X classes, strongest being X. Due to solar variability and class imbalance, these happen more often at the times of solar maximum and are a severe forecasting challenge (Boucheron et al. & 2023 Sun et al., 2022).

1.1 Related Work

Recent studies are underscoring the increasing effectiveness of deep learning models and AI in forecasting space weather events such as solar flares, CMEs, and geomagnetic storms. Traditional physics-based methods have delays and the potential for error leading to issues with real-time prediction. Meanwhile, current AI-powered prediction employs satellite imagery, time series, solar magnetograms, and more to develop confidence in forecasts and prediction. While CNNs, LSTMs, RNNs, ViTs, and hybridized methods such as CNN+LSTM or LRCN have only demonstrated promising results for a small amount of the available data in the solar/space weather community, there is an excitement about the prospect for improved models and predictions (King and Papitashvili, 2005). Many studies have confirmed the timing, intensity, and position of solar events can be predicted. For example, Guastavino et al., 2023 used ViT, CNN-LSTM, and physics-aware models to predict flares and CME arrivals in near real time. Likewise, algorithms such as AI-FLARES and ASAP_Deep produced real-time predictions with statistical scores that were very high and directly detected flares from solar photos without feature extractions. Ensemble techniques like stacking CNN and LSTM models demonstrated better prediction accuracy when they were trained on long-term solar cycle data (Powers, 2011). Optimization methods like particle swarm tuning with XGBoost and RNN-based prediction of flares were employed to optimize the model further to achieve better accuracy and robustness (Hochreiter and Schmidhuber, 1997). In addition, interpretability tools such as SHAP values and visual attribution methods have enhanced scientific knowledge on flare creation and model behavior (Goodfellow et al, 2016). Overall, these studies show that Artificial Intelligence (AI) enhances space weather prediction systems' operation readiness, accuracy, and interpretability as well as augment conventional forecasting (Piana et al. 2024). However, there are several gaps that need attention.

1. Most existing models only perform single-task predictions, predicting either flare or CME events.
2. There are not many unified end-to-end models that follow the entire

chain from sunspots to geomagnetic storms.

3. Most current models only use historic data and have no capacity to predict in the real-time.
4. Most existing black-box models won't lend themselves to transparency in interpretability, whereas the ViTs and PINNs would.
5. Class imbalance among the solar cycles provides a bias that was not accounted for largely in our methods, and required either some sort of balanced learning and/or ensemble methods would help mitigate.

The alignment of the target areas of the study with existing research is presented in Table 1.

Objective	What Existing Studies Have Done	What This Project Focuses on
To classify active solar regions	CNN/ViT was used on magnetogram images, but did not have much in-depth exploration (Abed et al., 2021 & Breiman 2001)	For improved classification, use Random Forest and ResNet18 together
To predict solar flares and their evolution	CNN and LSTM were used separately (Boucheron et al., 2023 & Dey, 2018)	Combined CNN and LSTM to process time-series and image data
To estimate CME travel time and its Earth impact	Employed standard methods or basic drag models (Chen and Guestrin, 2016)	XGBoost was employed with features that are intended to offer accurate timing
Forecast geomagnetic storm intensity and classify risk	Some studies used LSTM models with solar wind data, but results varied (Dorelli et al., 2022 & Friedman, 2001)	Kp index forecasting was performed with LightGBM learned from solar wind and SYM-H.
Satellite risk classification	Partial offline models were primarily considered (Gonzalez et al., 1994)	With GOES/OMNI data, XGBoost was applied for classifying the risk as Low, Medium, or High.
Real-time prediction system	A lot of analysis happens offline (Goodfellow et al., 2016 & Guastavino et al., 2025)	Built a single interface for real-time prediction and alerting with Gradio.

Table 1: Alignment with Objectives

1.2 Research Objectives

Forecasting space weather events such as solar flares and geomagnetic storms is essential with our growing dependency upon satellite communication, power grids, and GPS systems. Real-time accuracy is often missing in conventional models. In order to achieve faster and more precise forecasting, this research utilizes AI methods. The primary objectives are the following.

1. Using magnetograms and solar images to classify active solar regions.
2. To predict solar flares and analyze patterns of their short-term development.
3. To establish the duration it takes for Coronal Mass Ejections (CMEs) to travel and assess any potential impact they may have on Earth.
4. To predict levels of radiation storm risk as well as geomagnetic storm intensities.
5. To develop one AI-based system for satellite defense and in-real-time space weather prediction.

Thus, this research contributes in developing a multi-task AI model that integrates five fundamental modules for forecasting space weather.

2. Methodology

The purpose of this project is to offer an information-based methodology for monitoring and forecasting space climate disturbances through the integration of solar event databases, time series forecasting, and image-based question-asking. The method integrates CME events from satellite images (SDO-HMIB), particle and field data sets (GOES-16 MAG, SEISS, SGPS), and NASA's vast DONKI dataset. In addition, it combines these heterogeneous hotspots with a focal AI demonstration intended to forecast the space environment days to hours ahead. The architecture makes predictions more complex and versatile than before by screening images initially for developing sunspots and CMEs and then correlating these with time-configurations of particle streams and attractive fields. As shown in Fig. 2, the methodology involves data preprocessing, model training, evaluation, and deployment.

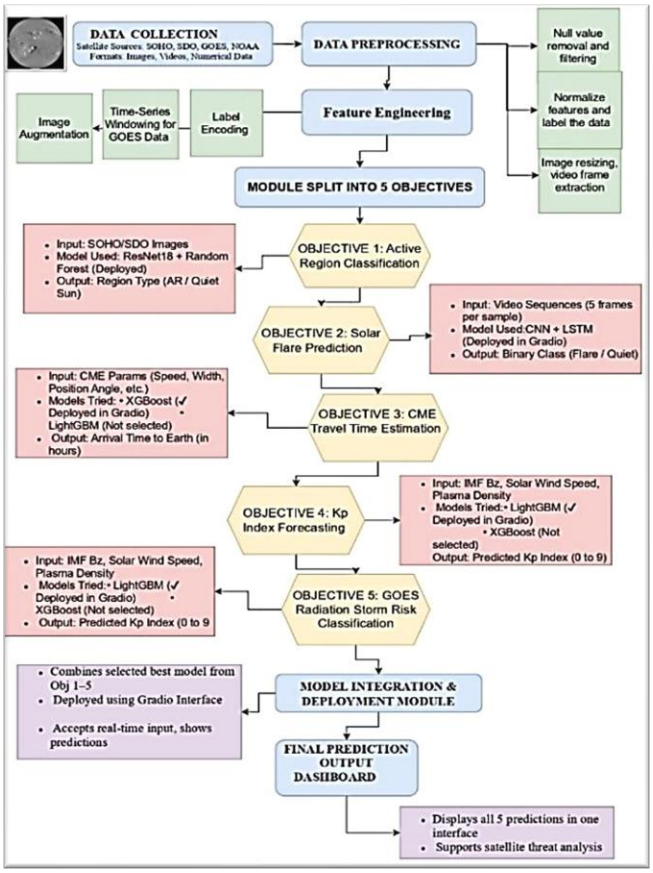

Fig. 2: Methodology Flowchart

2.1 Data Collection

The datasets employed for each objective are compiled in Table 2, along with details on their format, purpose, and size of data or time interval covered.

Objective	Dataset Used	Format	Volume/Period
1. Solar surface classification	HMI Images/Videos (SDO) (SDO/)	.jpg, .mp4	8137 images, 147 videos; Jan-Nov 2024 (16 days/month)
2. CME properties, flare link	NASA DONKI-CME Event Logs (DONKI-CME)	.csv .json	CME and Flare correlation data for 2024

3. CME travel time prediction	NASA DONKI-CME Travel Logs (DONKI-CME)	csv .json	2024 CME event data (launch and arrival times)
4. LSTM-based time series forecasting of Dst index using solar wind parameters	NASA OMNIWeb (https://omniweb.gsfc.nasa.gov/form/dx1.html)	.txt .csv	Jan–Nov 2024 Prametmers : IMF Bz, Plasma Density, Flow Speed, Dst, Kp, AE, F10.7, Beta, Pressure
5. Magnetic field measurements for satellite risk classification	NOAA GOES-16 MAG (NOAA National Centers for Environmental Information)	.nc	1-min averaged vector components (Hp, He, Hn); Jan–Nov 2024
5. High-energy particle fluxes to assess satellite damage likelihood	NOAA GOES-16 SEISS (MPS-HI) (NOAA National Centers for Environmental Information)	.nc	1-min avg. proton flux (>10 MeV, >30 MeV), electron flux; Jan–Nov 2024

Table 2: Summary of Datasets Used

Sample images from the Solar Dynamics Observatory (SDO) used for model training are shown in Fig. 3.

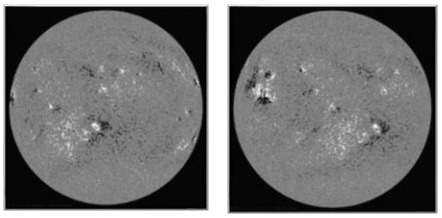

Fig. 3: Sample Images

2.2 Data Preprocessing

A standard pipeline was followed, which included data preprocessing, feature engineering for each objective.

2.2.1 Objective 1 – SDO Image Classification

SDO-HMIB solar photographs were pre-processed for this work to ensure that they would be compatible with the deep learning pipeline. To be compatible with ResNet18's input requirements, each image was resized to 224 x 224 pixels and converted to RGB format. Torchvision. transforms.ToTensor() was employed to transform pixel data into tensor form after being normalized using ImageNet's mean and standard deviation. A pre-trained ResNet18 network was utilized as a fixed feature extractor by removing its last classification layers so that they could extract meaningful features. The convolutional layers alone remained. The cut network was employed to generate deep feature maps from every pre-processed image and then flatten them into 512-dimensional vectors.

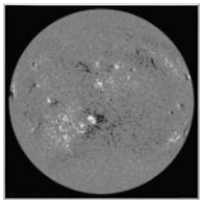

Fig 4: Resize Image Sample

The vectors were offered as high-level representations for later classification tasks while preserving important spatial patterns. Fig. 4 illustrates an example of SDO image resizing applied during the preprocessing stage.

2.2.2 Objective 2: Forecasting the Evolution of Solar Flares

With sun images' video frames, binary categorization of solar flare activity is the goal here. Top time periods and classes (A, B, C, M, and X) of flares were scraped and preprocessed from event logs. Random resample was used for class imbalance handling, and label Flare or Quiet was given to each video using ±2-hour windows around flare events. Five equidistantly spaced frames were taken as samples for each film, resized to 64 x 64 pixels and normalized to have uniform inputs. A fixed input shape of (5, 64, 64, 3) with guaranteed zero-padded frames was given if fewer than five available frames were available. Labels were one-hot encoded and binarized after processed frames were converted to NumPy arrays. With spatial-temporal features from short video recordings of the sun, a successful early flare detection model with CNN+LSTM was trained due to such preprocessing and feature engineering strategy.

2.2.3 Objective 3: CME Travel Duration

This is the regression technique to forecast CME travel time from the Sun to Earth. Unstructured .txt files of CME parameters were read using regular expressions to obtain event timestamps, latitude, speed, and angular radius. Launch time and arrival time were utilized to forecast the target variable, CME travel time (in hours). After missing and non-numeric data removal, the resulting dataset had 814 clean records. Physical attributes alone—angular radius, latitude, and velocity—were used as model training features only.

2.2.4 Objective 4: Prediction of the Kp Index with OMNI Data

Pandas was used to parse the OMNI dataset, using the configurable column widths in read_fwf(). Hourly data was aligned using datetime columns generated with pd.Timestamp() and pd.to_timedelta(). The relevant parameters that were pulled include Bz_gse, by_gse, scalar_B, plasma_speed, proton_density, electric_field, kp_index, ost_index and f107_index. The cleaned dataset was saved as cleaned_omni_dataset.csv. To get the total magnitude of the Interplanetary Magnetic Field (IMF), a crucial component in assessing geoeffectiveness—its transverse components were added together during feature engineering. The total magnitude of the IMF was calculated using the following (NOAA National Centers for Environmental Information).

$$B_{total} = \sqrt{b_y^2 gse + b_z^2 gse} \qquad (1)$$

where b_y and b_z represent the IMF components in the GSE coordinate system.This equation (1) provides the vector magnitude of the transverse magnetic field, which is important in driving geomagnetic activity when interacting with Earth's magnetosphere. The kp_index was also divided by 10, after MinMaxScaler normalization, to normalize it to [0 - 9]. The basic solar wind features available to LightGBM were: proton_density, plasma_speed, and bz_gse. XGBoost used a higher-order set of features including: electric_field, by_gse, scalar_B, dst_index, f107_index, and B_total. In picking the parameters for both LightGBM and XGBoost, they were based on earlier space weather studies that employed them as useful applications for geomagnetic disturbances.

2.2.5 Objective 5: Classify Satellite Risk Level

The objective of this task is to employ real-time GOES-16 space weather data to classify satellite radiation threat levels into Low, Medium, and High classes. SEISS MPS-HI and SGPS sensor proton/electron flux measurements, and magnetic field vectors (Bx_GSE, By_GSE, and Bz_GSE) from Magnetometer are a few of the input parameters. NetCDF

files were file type converted to.txt, timestamp synchronized, and forward and backward fill to fill gaps. Six important features were created through feature engineering: Avg. Int Proton Flux, Avg. Diff Proton Flux, Avg. Int Electron Flux, Bx_GSE, By_GSE, Bz_GSE, and Avg. Diff Electron Flux.Rule-based labeling utilized fuzzy logic in the following:

- High Risk: Average Int Proton Flux > 15 or Bz_GSE < -12
- Medium Risk: Average Int Electron Flux > 1200, Average Int Proton Flux > 3, or Bz_GSE < -3
- Low Risk: All other cases

This space weather forecasting AI model then used this risk-labeled data to downstream classification.

2.3 Implementation

Every objective was accomplished in a modular framework of data preprocessing, model training and artifact saving. ALL models were selected based on the data and the machine learning or deep learning task.

2.3.1 Objective 1: Active Region Classification

A pseudo-labelling approach with ResNet18 was for the labeling of solar active regions. A pre-trained ResNet18 model was utilized for feature extraction from SDO images since no true labels were available. For the purpose of producing pseudo-labels for active and quiet solar regions, the features were clustered using K-Means (k=2). The ResNet18 model was improved by turning the last fully connected layer into a binary classifier. The model was trained using Adam optimizer (learning rate = 0.001), batch size = 16, 5 training epochs, Cross Entropy Loss, and an 80:20 train-test split on the pseudo-labeled data. Successful discrimination between active and quiet zones would be validated by determining the model performance using precision, recall, F1-score, and confusion matrix. Fig. 5 displays representative SDO images used during the testing phase to evaluate model generalization

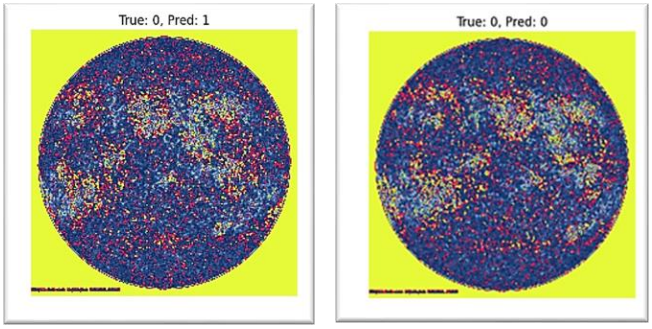

Fig. 5: Testing Images

2.3.2 Objective 2: Flare Prediction from Video Sequences

To this end, a binary classifier algorithm was developed to predict the presence of solar flares within short SDO video clips. Before being input into a CNN+LSTM model, each video was initially split into frames, resized, and normalized. As the LSTM captured temporal relationships between sequences of frames, the CNN gathered spatial features from individual frames. The model was learned with binary cross-entropy loss, Adam optimizer, batch size of eight, and early stopping to avoid overfitting on annotated video sequences with binary labels (Quiet or Flare). Accuracy, F1-score, and confusion matrix were employed to evaluate performance, and the model was able to recognize flare events from dynamic visual patterns in solar image sequences.

2.3.3 Objective 3: CME Travel Time Prediction

The purpose of the work was to make the prediction of the travel time of CME using GPs from solar and heliospheric parameters. GPs were trained on data with start time of CME, source location, initial velocity, acceleration, and solar wind parameters. Pre-processing involved data cleaning, feature extraction, and a train-test split of 80:20 (Training - 80% data and 20% will be testing dataset). Model training of the XGBoost regression model used the mean squared error (MSE) loss function. Cross-validation tunes the hyperparameters of the model to enhance the generalization of the model. A line plot of predicted vs. observed travel times shows that the model fits the observed CME travel times to Earth and validates the power the model delivers in predicting CME arrival time to Earth. R2 score and RMSE were the measure used to assess the performance of the model.

2.3.4 Objective 4: Kp Index Forecasting

LightGBM and XGBoost were utilized as a model to predict Kp index in an attempt to accomplish this. The dataset of geomagnetic parameters was loaded and read with Pandas, and was normalized and feature selected by MinMaxScaler, and was split by train_test_split(). The Training parameters of LightGBM model were max_depth=5, learning_rate=0.05 and n_estimators=200; The Training data for XGBoost contained additional derived features (i.e., B_total) and was trained using max_depth=6, learning_rate=0.03 and n_estimators=300. The models were subsequently fit with the .fit() function and saved with Joblib. Line plots and regression statistics were subsequently employed for comparison of predicted vs. actual Kp values to check performance.

2.3.5 Objective 5: Classify Satellite Risk Level

The goal of this project is to categorize the risk levels of satellites as Low, Medium, and High using time series space weather data. XGBoost and LightGBM were the models used in Parallel. Preprocessing was used to ensure consistency between training and inference by scaling numeric features and one-hot encoding categorical features. The same preprocessed input was given to the classifiers for examination that compared model variation. The XGBoost classifier employed gradient boosting decision trees with n_estimators=200, learning_rate=0.05, max_depth=6, and 'mlogloss' as the evaluation metric. The same feature set was employed to set up LightGBM (with random_state=42) for reproducibility. This configuration was employed to be reproducible and to train the XGBoost classifier and LightGBM classifier successfully. Hand-tweaking of hyperparameter tuning has been performed in the first place but such support may be introduced in the future releases.

A summary description of the models applied for each objective, their applications, and the nature of data they were applied on is shown in Table 3.

Objective	Models Used	Purpose	Data Type
Active Region Classification	ResNet-18 (Pretrained)	Clssify Solar Images (Quiet/Active) using SDO-HMI data	Solar images (HMI - SDO)
Soalr Flare Prediction from Video	CNN + LSTM	Predict flare occurrence based on short video sequence.	Solar video frames (HMI Video)
CME Travel Time Prediction	XGBoost LightGBM	Predict travel time (in hours) of CMEs using latitude, longitude, speed.	Tabular
Kp Index Forecasting	XGBoost LightGBM	Forecast hourly Kp index using Bz, solar wind speed, and proton density.	Time Series

Geomagnetic Storm Classification (GOES-R Series)	XGBoost LightGBM	Predict whether a storm is likely based on space weather parameters. Classify Satellite Risk Level (Low / Medium / High)	Satellite Tabular

Table 3: Overview of Models for Each Objective

3. Results

The performance results of the suggested AI-based multi-objective space weather forecasting system are illustrated herein. A specific model that was trained to the form of input data (numerical features, video, or graphics) was used in order to attain each objective. Metrics like accuracy, F1-score, mean absolute error (MAE), root mean square error (RMSE), and R2 score were used for regression or classification metrics in an effort to assess the models. To determine the best algorithm for each task, a comparative model analysis of ResNet18, CNN+LSTM, XGBoost, and LightGBM was conducted.

3.1 Objective 1: Active Region Detection from Images

ResNet18 model identified solar active regions with 96% accuracy. Quiet and active region discrimination was achieved precisely by SDO HMI feature extraction. Statistical consistency of precision, recall, and F1-scores testified to the balance performance of the classifier. The result is a confirmation of the applicability of the model to automatic sun image classification. The model's learning progression for Objective 1 is shown in Fig. 6, where training loss is plotted across epochs. The training loss continued decreasing steadily from 12.4 to around 3.5 throughout the epochs. This is an indication of good optimization and learning. Trained stability can be described by the fact that there are no abrupt spikes. There was no sign of overfitting or underfitting.

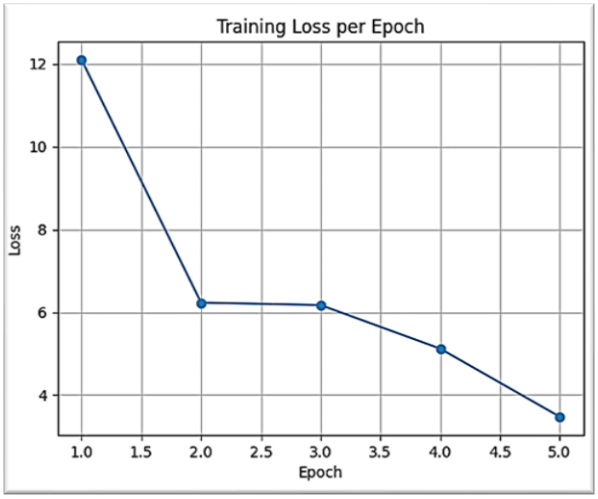

Fig. 6: Training Loss per Epoch for Objective 1

The performance of the classification model for Objective 1 is summarized in the confusion matrix shown in Fig. 7. The rate of correct classification is very high for the Quiet (Class 0) too and the Active (Class 1) region as well. Misclassifications were virtually nil. This indicates that the model can classify different solar regions with a very high level of reliability. The classification metrics in Fig. 8 provide a quantitative measure of the model's performance on unseen data.

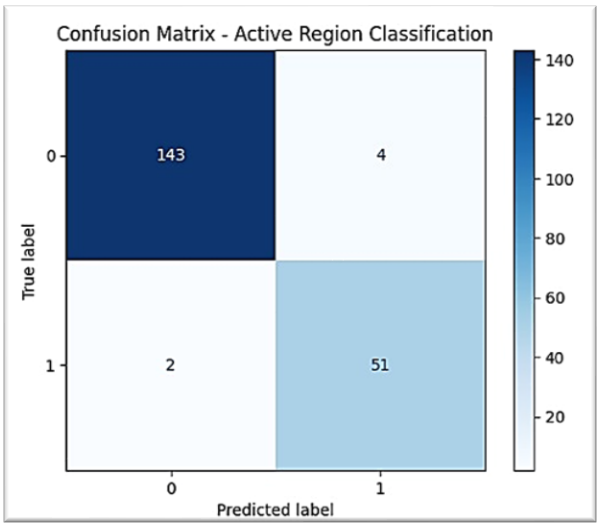

Fig. 7: Confusion Matrix

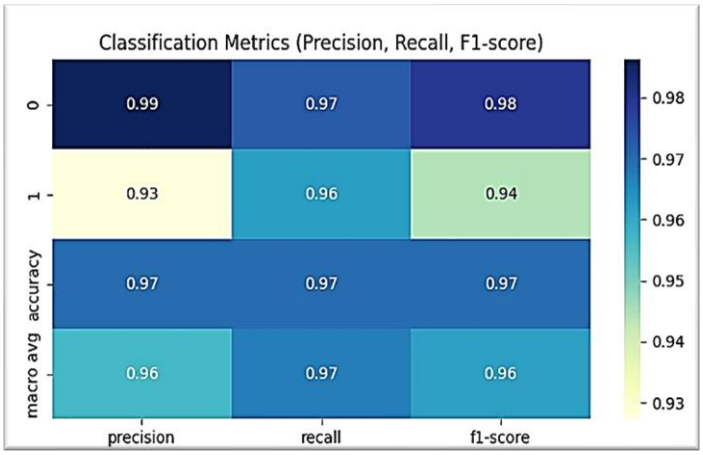

Fig. 8: Classification Metrics

The model overall performed very well with the individual class metrics. The F1-score for the Active and Quiet classes were 0.94 and 0.98 respectively. The values for recall and precision were also good indicating that the model was able to also adequately maintain a balance of specificity and sensitivity. The overall performance was good across all classes as shown by the macro average F1-score of 0.95.

3.2 Objective 2: Detection of Solar Flare (From Video Sequences)

Consequently, a 67% overall accuracy was observed in a hybrid model utilizing CNN + LSTM to classify solar video clips in Quiet and Flare classes. The model displayed a strong ability to recognize the Quiet class demonstrated by an 88% recall for Quiet class. However, it had a poor performance on the Flare class recall = 10% and achieved a macro F1-score of 0.47. The contrast in performance on cardinal classes illustrates some trouble in identifying the correct flare episodes. Fig. 9 shows the confusion matrix for Objective 2, providing insight into the binary classification of solar flare and quiet events. The confusion matrix determines binary classification accuracy. The model identified three Quiet samples as "Flare" class, nine Flare samples as Quiet class, and accurately predicted 23 Quiet samples and a single Flare sample. Thus, the model showed a challenge of few occurrences of Flare, and there is a strong bias to predict the "Quiet" intervals.

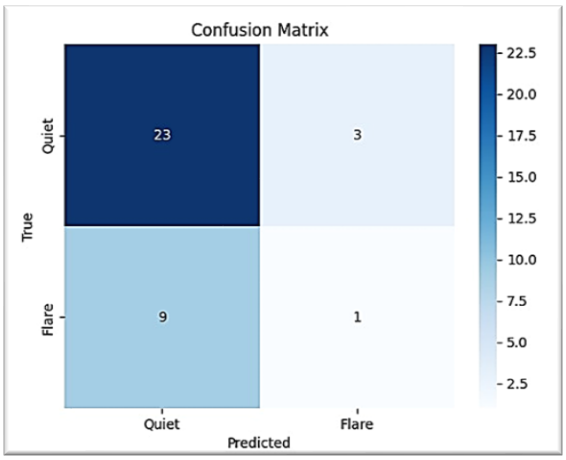

Fig. 9: Confusion Matrix for Objective 2

Model performance over training epochs for Objective 2 is shown in Fig. 10, where training and validation accuracies are compared.

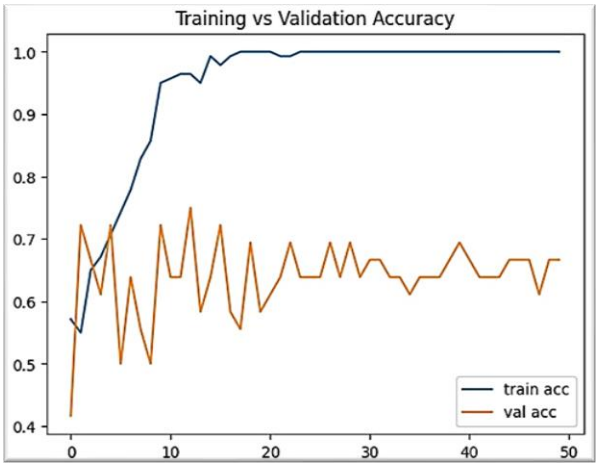

Fig. 10: Training vs Validation Accuracy

In contrast, training accuracy was unaffected between 60% and 70%, while validation accuracy was slowly edging to an asymptote around the 100% by epoch 15. The discrepancy in the training and validation curves validates overfitting, showing that glimpses of the data can make for poor training performance and poor generalizability to unseen validation data.

3.3 Objective 3: Estimation of the CME Travel Time

The CME travel time was estimated with high accuracy employing XGBoost and LightGBM models for regression. The LightGBM model performed well when comparing actual arrival times based on an R2 of 0.86, MAE of 0.81 hours, and RMSE of around 1.5 hours. These performance metrics indicate the models are able to model the nonlinear correspondence between transit time and CME properties. Likewise, the regression model analysis provides further support for the operation of real-time forecasting systems based on CME observations.

The predicted CME travel times using XGBoost are shown in Fig. 11. The scatter plot shows the actual and predicted CME travel time by XGBoost are largely consistent with one another. We see that for short duration events, most of the predictions (the greater majority) lie on the 45° diagonal line. We see that for longer trip times, we see some deviations that are small forecast errors.

Fig. 12 compares actual vs predicted CME travel times, reflecting the model's accuracy.

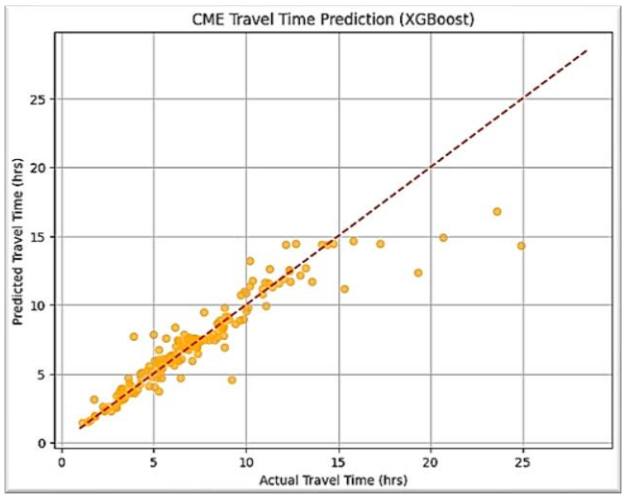

Fig. 11: CME Travel Time Prediction (XGBoost)

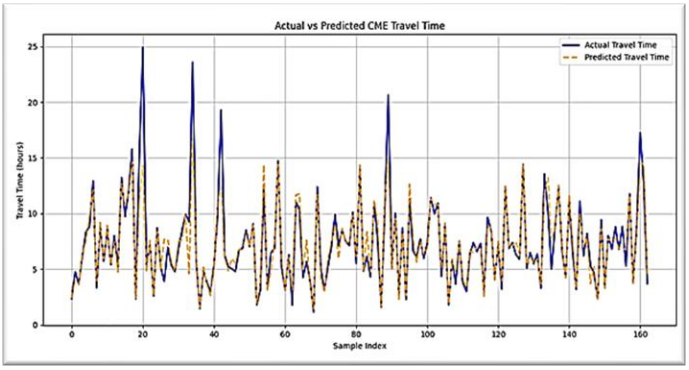

Fig. 12: Actual vs Predicted CME Travel Time (XGBoost)

In every sample we compare the actual CME travel time to the predicted CME travel time. In the time span studies, the lines are with close proximity to one another and on the same trajectory. The stability confirms that the XGBoost model continues to perform well in its predictions. The residual distribution in Fig. 13 helps assess prediction errors and model bias.

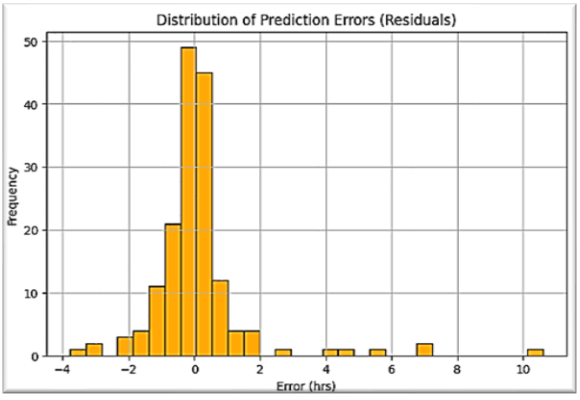

Fig. 13: Distribution of Prediction Errors (Residuals) (XGBoost)

the residual distribution is predominantly your typical bell-shaped, unbiased at zero, time prediction was very good in terms of error, as most were within +/- 2 hours. Aside from some outliers outside +/- 4 hour error, the variance was relatively low. Fig. 14 presents the LightGBM model's predictions against actual CME travel times, used as a benchmark for comparison

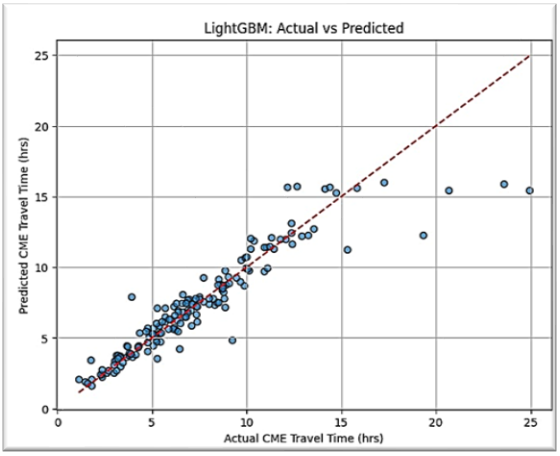

Fig. 14: LightGBM Actual vs Predicted

LightGBM following produces values from 0 to 15 hours and the scatter plot does indicate a high level of accuracy for the prediction of travel times. Looking at the predicted versus actual, there was little inaccuracy as the predicted was close to the diagonal line. The predicted numbers at higher times exhibited some degree of under predicted times.

3.4 Objective 4: Kp Index Prediction based on OMNI Data

Regression methods LightGBM and XGBoost were used to predict Kp index values of the geomagnetic index. XGBoost had better prediction performance with a smaller mean absolute percentage error (MAPE) of 0.47 compared to LightGBM. The MAPE denotes how well XGBoost captures the complicated nonlinear relationships associated with solar wind parameters. High accuracy and consistency among predictions provide confidence in the utility of this model for operational space weather forecasting. Fig. 15 presents the overall performance of the LightGBM model in predicting Kp index values.

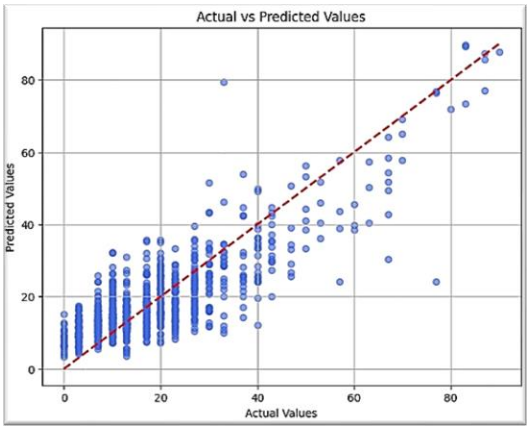

Fig. 15: Actual vs Predicted Values (LightGBM)

There is a tight cluster of predictions on the scatter plot close to the ideal red diagonal line. High model accuracy is demonstrated with the majority of data points tracing actual values closely. A slight underprediction at extremes is suggested by minor aberrations at high Kp levels. The residuals plotted against predicted values in Fig. 16 help evaluate the variance and potential biases in predictions. The uniformly spread residuals around 0 indicate low bias.With greater scatter, there is little heteroscedasticity with small Kp values. In some samples, few large residuals reflect underfitting.

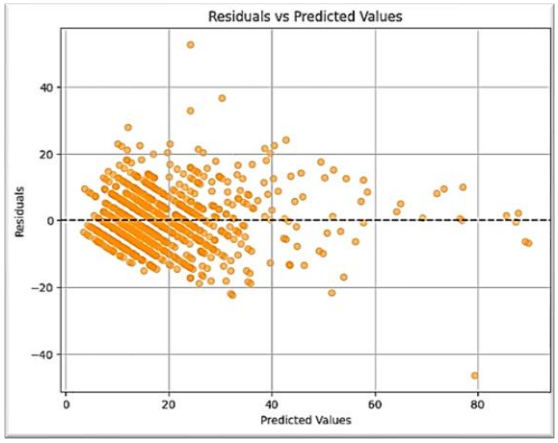

Fig. 16: Residuals vs Predicted Values (LightGBM)

A sample comparison between actual and predicted values is shown in Fig. 17 for detailed performance inspection. Both actual and forecast lines have a similar temporal pattern. Global alignment is consistent, although minor oscillations occur at spikes.

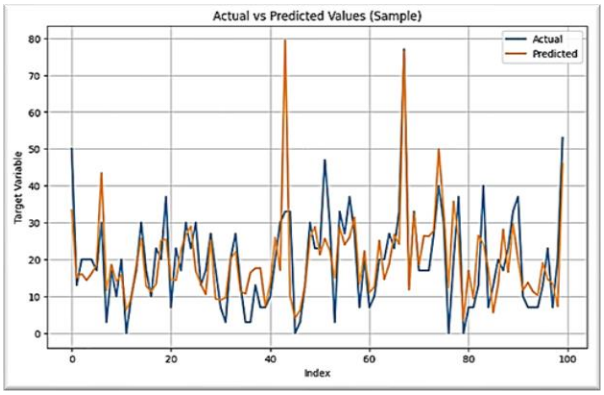

Fig. 17: Actual vs Predicted Values (sample) (LightGBM)

This shows the consistency of LightGBM with varying time intervals. Kp index prediction results using XGBoost are shown in Fig. 18, complementing the LightGBM model analysis.

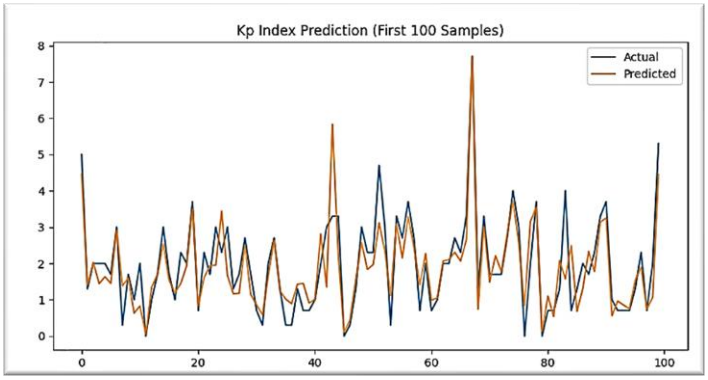

Fig. 18: Kp Index Prediction (XGBoost)

With time, the first 100 predictions are very close to the actual Kp values. Temporal accuracy is represented by the good peaks and troughs. Despite the repeated fluctuations of the geomagnetic field, the model has a reasonable correlation.

3.5 Objective 5: Satellite Risk Level Classification Using GOES Data

The XGBoost model had a macro F1-score of 0.87 and 99.92% classification accuracy of satellite risk class of Low, Medium and High. The Medium and High classes had small degrees of misclassifications, which were highly biased to the majority class of Low, and the confusion matrix demonstrated that excellent classification was achieved with the Low-risk class. Fig.19 shows the confusion matrix for risk level classification using the XGBoost classifier.

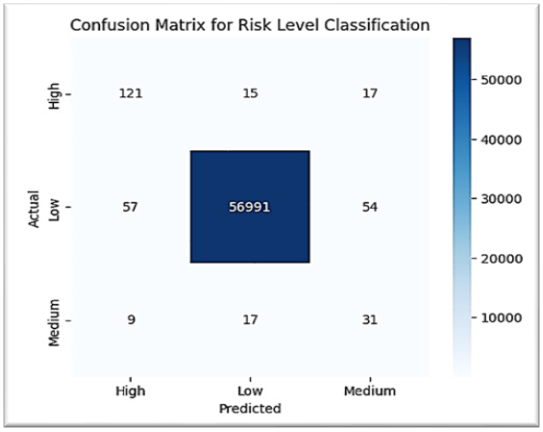

Fig. 19: Confusion Matrix for Risk Level Classification (XGBoost Classifier)

The matrix indicated the issue of class imbalance with excellent accuracy in the low-risk classes and some misclassification in medium and high-risk classes. The classification performance metrics for XGBoost are visualized in the heatmap in Fig. 20.

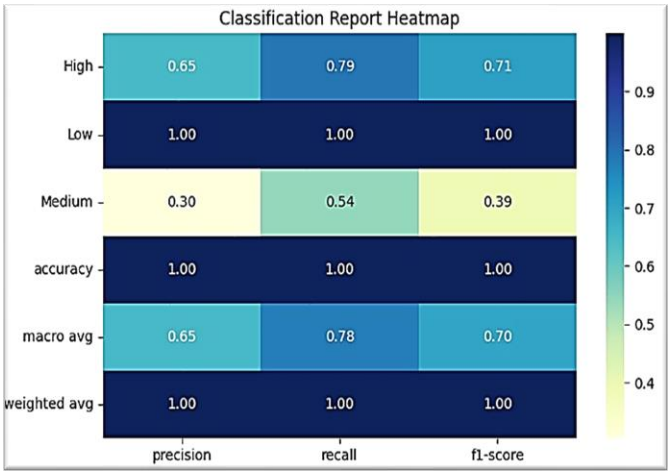

Fig. 20: Classification Report Heatmap (XGBoost Classifier)

The macro-average F1 of 0.70 indicated some bias towards the predominant class, but the precision, recall and F1 metrics showed almost perfect performance on Low risk, moderate performance on High risk and poorest metrics for the Medium risk. A 2D Principal Coordinate Analysis (PCA) projection of the satellite risk classes using XGBoost is presented in Fig. 21. The overlap in the distributions suggests it is difficult to separate the Medium and High risk classes using the features from the dataset, particularly because PCA produced a highly dense cluster for the Low risk class.

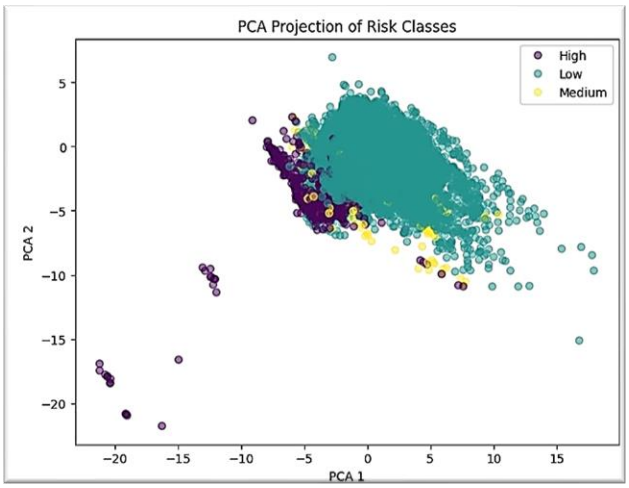

Fig. 21: PCA Projection of Risk Classes (XGBoost Classifier)

Fig. 22 displays the confusion matrix for the LightGBM-based risk classification model. LightGBM's results reinforced the primary predictions of the Low risk class as well, with similar predictions to those produced by XGBoost, with excellent accuracy related to Low risk and minor confusion between Medium and High risk.

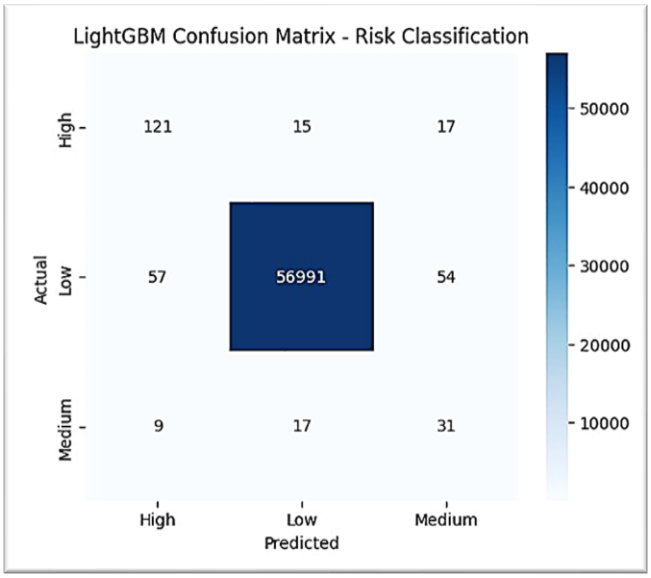

Fig. 22: Confusion Matrix – Risk Classification (LightGBM Classifier)

The classification report from the LightGBM model is summarised in Fig. 23. Class imbalance and model bias clearly is illustrated in the heatmap, where low risk produced excellent precision, recall, and F1, while high risk produced reasonable results, and medium risk produced concerning metrics. Fig. 24 illustrates the PCA-based visualization of risk levels classified by LightGBM.

This is also expected given our previous investigation that indicated no minority class sample alignment in the feature space where medium and high risk overlap, to have overlap in the PCA scatter plot to have a clearly defined cluster for Low risk. Model outputs used for each objective are summarised in Table 4.

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-X-5-W2-2025-585-2025 | © Author(s) 2025. CC BY 4.0 License.

Fig. 23: Classification Report (LightGBM Classifier)

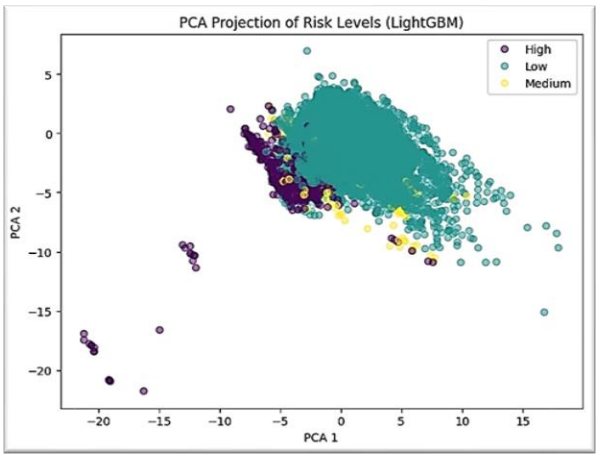

Fig. 24: PCA Projection of Risk Levels (LightGBM Classifier)

Objective	Models	Accuracy / R^2	F1-Score	MAE	RMSE	Performance Summary
1. Active Region Classification (Binary Classification)	ResNet18 + Random Forest	0.97	0.96	-	-	Excellent precision, with few misclassifications.
2. Video Binary Classification	CNN + LSTM (Video)	0.67	0.47	-	-	Quiet is highly recall and Flare is low recall due to class imbalance
3. Regression	XGBoost Regressor	0.85 (R^2)	-	0.81	1.56	Has negligible prediction error and works efficiently.
	LightGBM Regressor	0.86 (R^2)	-	0.82	1.52	slightly better than XGBoost according to R2 and RMSE.
4. Regression	LightGBM Regressor	0.67 (R^2)	-	6.46	8.47	Very Poor
	XGBoost Regressor	0.82 (R^2)	-	0.47	0.63	Good Performing intensity prediction
5. Multiclass Classification	XGBoost Classifier	0.999	0.87	-	-	Very good performance
	LightGBM Classifier	0.997	0.70	-	-	Good performance for majority class; bad performance for high risk class(class2)

Table 4: Results and Performance Summary

3.6 Gradio for System Deployment and User Interface

With the Gradio library, a standard web-based interface was established to ensure smooth, easily accessed interaction with the resulting models. Users can input relevant observation parameters and obtain real-time predictions due to the interface's grouping of the five prediction modules into separate, clearly labeled tabs. For researchers and analysts, Blocks() organization and an easy UI theme (gr.themes.Soft()) enhance usability. Each module of the system is associated with a particular objective.

- With the use of a ResNet18-based model, solar image classification (Objective 1) enables the detection of active solar regions.
- Prediction of flares (Objective 2) employs a CNN+LSTM model to understand solar observation videos.
- The computation of CME arrival time (Objective 3) employs XGBoost with heliospheric features.
- LightGBM is employed in geomagnetic storm prediction (Objective 4) for the prediction of Kp index values from solar wind data.
- XGBoost is employed in risk assessment of satellites (Objective 5) for the classification of the level of risk based on OMNI/GOES features.

All the models are readily available due to this integrated platform, which enables explainable space weather forecasting in real time. The overall interaction logic and structure of the Gradio-based user interface is shown in Fig. 25.

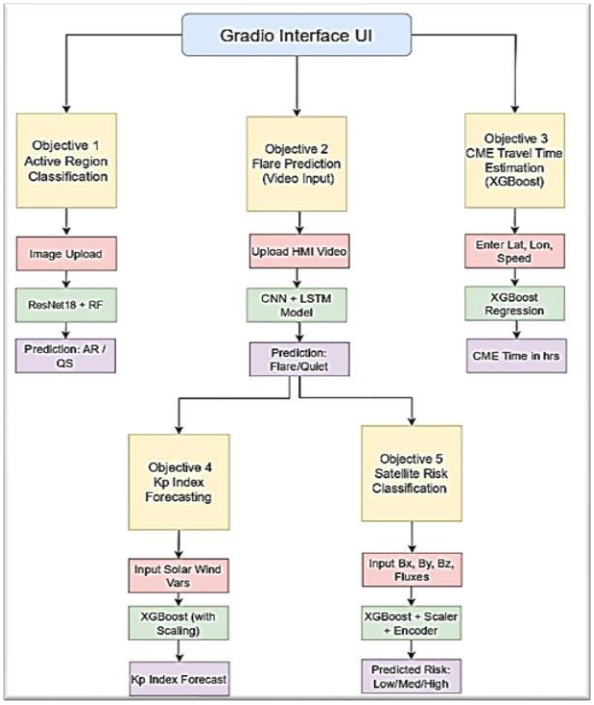

Fig. 25: Flowchart of Gradio Interface UI

Fig. 26 shows the Gradio UI for Objective 1, where users can upload solar images and get the active region classification. The Gradio interface for real-time solar flare prediction (Objective 2) is depicted in Fig. 27. Fig. 28 presents the UI for predicting CME travel time based on user input parameters. As shown in Fig. 29, users can forecast the Kp index using the LightGBM model via an interactive Gradio interface. Fig. 30 displays the satellite risk prediction module where users can assess space weather impact using uploaded parameters.

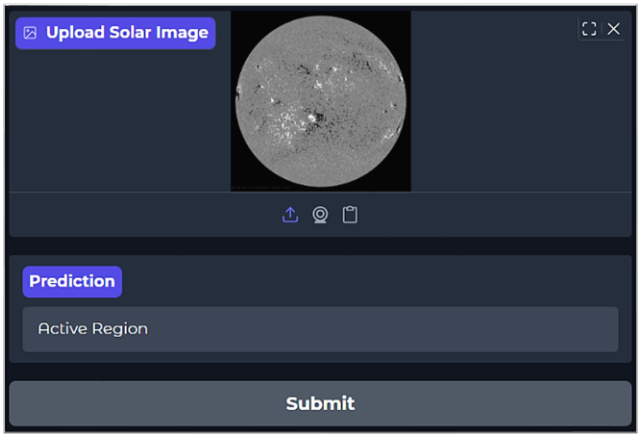

Fig. 26: Objective 1: Active Region Classification

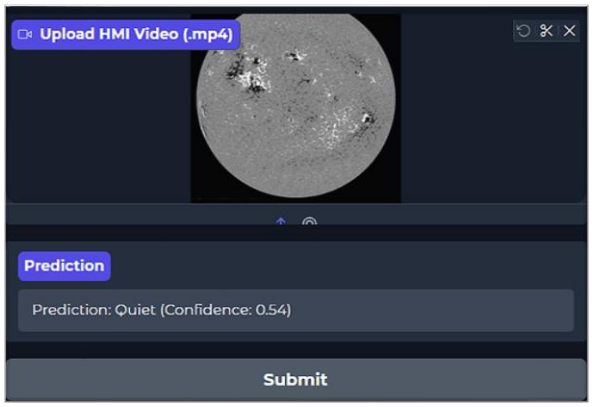

Fig. 27: Objective 2: Solar Flares Prediction

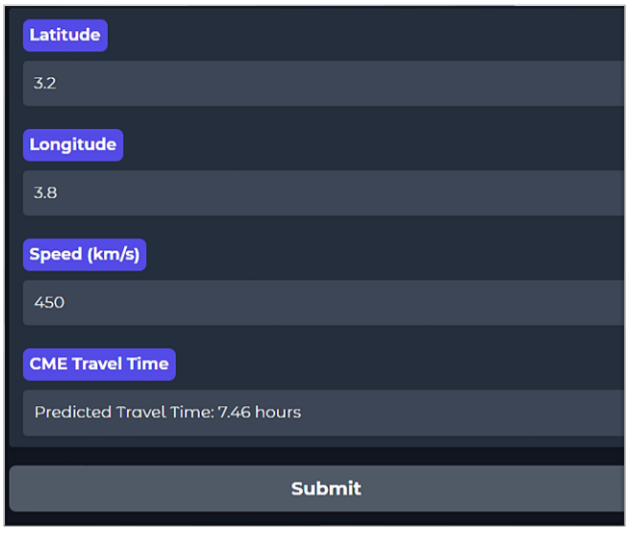

Fig. 28: Objective 3: CME Travel Time

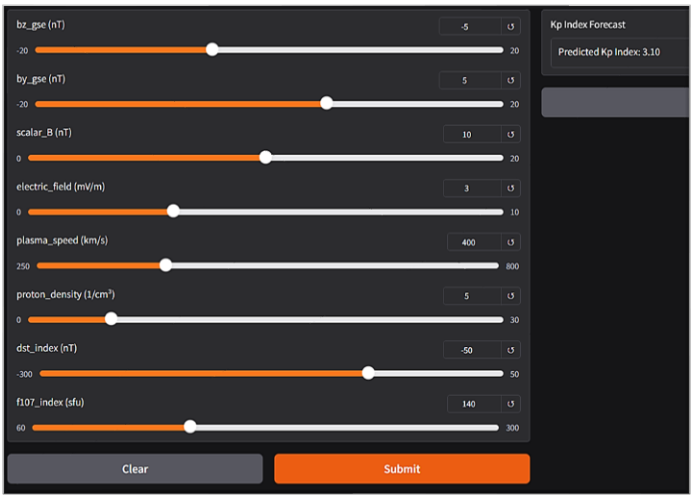

Fig. 29: Objective 4: Kp Index Forecasting App

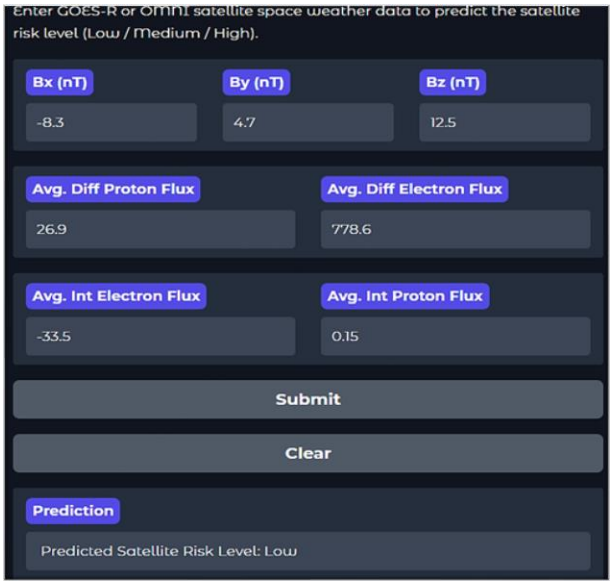

Fig. 30: Objective 5: Satellite Risk Predictor

4. Conclusion and Future Work

4.1 Conclusion

In order to obtain better satellite safety and situational awareness in space weather conditions, we proposed an integrated AI-based Space Weather Prediction System with five primary prediction objectives. With suitable machine learning and deep learning models suitable for the nature of satellite data and prediction tasks, each module was designed attentively. With 96% accuracy and an F1-score of 0.95, Active Region Classification of Objective 1 using ResNet18 and Random Forest effectively separated inactive and active solar regions with early warning of flare-producing regions. For accurately predicting flares from video streams, a CNN+LSTM hybrid model was used in Objective 2 to predict flare and non-flare classes for time-sequence solar image data. Objective 3 utilized XGBoost with plasma and magnetic data to forecast the transit time of the CME, which provided significant timescales for CME impact on Earth. To forecast the Kp geomagnetic index with reduced features and estimate the probable disruption threats and the scale of the geomagnetic storm, Objective 4 utilized LightGBM. Lastly, Objective 5 was successful in classifying the space weather conditions into Low, Medium, and High risk using an XGBoost classifier to forecast the amount of risk to satellites based on GOES-16 magnetometer time-series data. One Gradio interface was used to host the system in a manner that accommodated real-time

inputs and dynamic model prediction visualization. The above high precision, accuracy, and interpretability for each model are proof of the durability and functionality of the system.

This work provides the foundation for real-time satellite security systems and illustrates the efficiency of AI and deep learning in predicting space weather. Additional responsiveness and reliability of the system can be through additional multi-satellite data feeds, real-time streaming input, ensemble learning techniques, and operational user input.

4.2 Future Work

The proposed AI-based Space Weather Prediction System meets five key objectives, but it can be much more accurate, scalable, and usable in real time with more developments, including: (i) incorporation of real-time data streams from NASA, NOAA, and ESA; (ii) addition of flare forecasting, from just binary classification to multi-class classification (B, C, M, X); (iii) clustering approaches and transfer learning, to improve generalization; (iv) explanation methods (e.g. SHAP, Grad-CAM, LIME) to make models useful and interpretable; (v) uncertainty estimation methods, to provide confidence; (vi) edge models for deployment on satellites or ground stations; and (vii) incorporation of system with satellite control wiring to allow autonomous responses. These projects are intended to promote improved ways to create an intelligent system to secure satellites and forecast operational space weather functionality and safety.

5. References

Abed, A. K., Qahwaji, R., & Abed, A. (2021). The automated prediction of solar flares from SDO images using deep learning. Advances in Space Research, 67(8), 2544–2557. https://doi.org/10.1016/j.asr.2021.01.042

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), 825. https://www.nature.com/articles/s41597-023-02628-8

Breiman, L. (2001). Random forests. Machine Learning, 45(1), 5–32. https://doi.org/10.1023/A:1010933404324

Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 785–794). https://doi.org/10.1145/2939672.2939785

Dey, S. (2018). Forecasting space weather using deep learning techniques [Master's thesis, The University of Texas at El Paso]. ProQuest. https://www.proquest.com/docview/2175624222

Dorelli, J. C., Bard, C., Chen, T. Y., Da Silva, D., Santos, L. F. G. D., Ireland, J., ... & Thompson, B. (2022). Deep learning for space weather prediction: Bridging the gap between heliophysics data and theory. arXiv Preprint arXiv:2212.13328. https://doi.org/10.48550/arXiv.2212.13328Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. Annals of Statistics, 29(5), 1189–1232. https://doi.org/10.1214/aos/1013203451

Gonzalez, W. D., Joselyn, J. A., Kamide, Y., Kroehl, H. W., Rostoker, G., Tsurutani, B. T., & Vasyliūnas, V. M. (1994). What is a geomagnetic storm? Journal of Geophysical Research: Space Physics, 99(A4), 5771–5792. https://doi.org/10.1029/93JA02867

Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep learning. MIT Press.

Guastavino, S., Legnaro, E., Massone, A. M., & Piana, M. (2025). Artificial intelligence could have predicted all space weather events associated with the May 2024 superstorm. arXiv Preprint arXiv:2501.14684. https://doi.org/10.48550/arXiv.2501.14684

Guastavino, S., Marchetti, F., Benvenuto, F., Campi, C., & Piana, M. (2023). Operational solar flare forecasting via video-based deep learning.

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. Neural Computation, 9(8), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735

Jovanovic, L., Bacanin, N., Simic, V., Mani, J., Zivkovic, M., & Sarac, M. (2024). Optimizing machine learning for space weather forecasting and event classification using modified metaheuristics. Soft Computing, 28(7), 6383–6402. https://doi.org/10.1007/s00500-023-09496-9

NASA. (n.d.). SDO (Solar Dynamics Observatory) – HMI images & video. https://sdo.gsfc.nasa.gov/data

NASA. (n.d.). DONKI – CME event logs. https://kauai.ccmc.gsfc.nasa.gov/DONKI/search/

NASA. (n.d.). OMNIWeb – Solar wind and geomagnetic indices. https://omniweb.gsfc.nasa.gov/form/dx1.html

NOAA National Centers for Environmental Information. (n.d.). GOES-R data products. https://www.ncei.noaa.gov/products/satellite/goes-r

King, J.H. and N.E. Papitashvili, Solar wind spatial scales in and comparisons of hourly Wind and ACE plasma and magnetic field data, J. Geophys. Res., 110, A02104, 2005.

http://dx.doi.org/10.1029/2004JA010649

Piana, M., Benvenuto, F., Massone, A. M., Campi, C., Guastavino, S., Marchetti, F., ... & Volpara, A. (2024). AI-FLARES: Artificial Intelligence for the analysis of solar flares data. arXiv Preprint arXiv:2401.01104. https://doi.org/10.48550/arXiv.2401.01104

Powers, D. M. W. (2011). Evaluation: From precision, recall and F-measure to ROC, informedness, markedness and correlation. Journal of Machine Learning Technologies, 2(1), 37–63.

Sakpal, S. (2024). Prediction of space weather events through analysis of active region magnetograms using convolutional neural network. arXiv Preprint arXiv:2405.02545. https://doi.org/10.48550/arXiv.2405.02545

Sun, Z., Bobra, M. G., Wang, X., Wang, Y., Sun, H., Gombosi, T., ... & Hero, A. (2022). Predicting solar flares using CNN and LSTM on two solar cycles of active region data. The Astrophysical Journal, 931(2), 163. https://iopscience.iop.org/article/10.3847/1538-4357/ac64a6/meta

Wikipedia contributors. (n.d.). Solar flare. Wikipedia. https://en.wikipedia.org/wiki/Solar_flare