Solar Irradiance Forecasting using INSAT-3DS Satellite Data

Abhay Chaudhary

Overview

This project forecasts surface-level Global Horizontal Irradiance (GHI) using spatio-temporal features from INSAT-3DS satellite imagery. The goal is to generate accurate 15-minute interval predictions for September 1–7, 2024, based solely on satellite data and past GHI values.

Dataset

- **Input:** NetCDF (.nc) files from INSAT-3DS
- **Channels used:** `IMG_MIR`, `IMG_SWIR`, `IMG_TIR1`, `Sun_Elevation`, `Sat_Elevation`
- **Target:** Observed GHI values from CSV
- **Temporal Resolution:** 15-minute intervals
- **Spatial Resolution:** 5×5 grid centered on observation point

Model Evolution

Stage	What I Tried	Result	Next Step
XGBoost (Tabular)	Statistical features (mean, std, min, max)	RMSE ≈ 210; MAPE exploded due to GHI=0	Moved to spatio-temporal model
ConvLSTM (1 sample/day)	Treated each day as 1 sample	Too few samples; overfit	Used sliding windows
Sliding ConvLSTM (Win=6)	Many samples with stride=1	SMAPE stuck at 54%	Discovered lag in prediction
Lag Analysis	Found 3-step (90 min) lag	Applied SHIFT=3 to label	SMAPE dropped to 13%, RMSE to 80

Final Model Details

- **Architecture:** 2-layer ConvLSTM → Flatten → Dense
- **Window Size:** 6 time steps (1.5 hrs)
- **SHIFT:** 3 (model predicts GHI at last timestep of window)
- **Loss:** Huber
- **Scaler:** Satellite inputs normalized by 1023, GHI by 1100

- **Training Samples:** ∼150+
- **Validation Metrics:**
  - RMSE: **80.5 W/m²**
  - MAE: **51.9 W/m²**
  - Daylight SMAPE: **12.9%**

Inference Strategy

- Ran stride=1 prediction on each day from Sept 1–7
- Each window predicted the GHI for its last timestep
- Model outputs de-normalized using MAX_GHI = 1100
- Final CSV:
  - 672 rows (7 days × 96 time steps)
  - 15-min cadence, no gaps, values clipped [0, 1100]

File Structure

```
.
model.ipynb                    # Full pipeline
convLSTM_shift3_best.h5        # Trained weights
sept_forecast.csv              # Final predictions
Sample Dataset - ML Assignment - Sheet1.csv
README.md / this LaTeX file
dataset/
    Jun_Output/
    Jul_Output/
    Aug_Output/
    Sept_Output/
```

Challenges and Fixes

- **Exploding MAPE:** GHI=0 at night →
- **Small dataset:** Solved with sliding window (stride=1) over 24-hour sequence
- **Lag issue:** Predictions were late by 90 minutes → fixed with SHIFT=3
- **Gaps in inference:** Interpolated missing timestamps to get 672 total

Conclusion

Through progressive experimentation—starting with XGBoost and converging on ConvLSTM with aligned labels—I was able to build a robust spatio-temporal model for GHI forecasting. The final pipeline is accurate, generalizable, and passes all consistency checks for deployment.