PROGRESS REPORT FOR PREDICTION OF CYCLONE PATH

COMPLETED WORK:

1) Literature Review
2) Parameter Analysis
3) Model Selection
4) Exploration Of IBTrACS Dataset
5) Visualization Of Different Past Cyclones
6) Basic Model Train For IBTrACS Data

A. IBTrACS DATASET:

The IBTrACS dataset provides global tropical cyclone track data, including storm positions (latitude and longitude), wind speed, pressure, timestamps, storm IDs, and basin information. It consolidates historical cyclone observations from multiple agencies, making it a comprehensive source for analyzing cyclone frequency, intensity, and trajectories. The dataset is widely used for cyclone modeling, prediction, and climate research.

LINK:
https://www.ncei.noaa.gov/access/metadata/landing-page/bin/iso?id=gov.noaa.ncdc:C01552

SIZE: 26.2 MB

ATTRIBUTES : Over 140 features, including:

- Storm Information: Storm ID, name, season, basin, subbasin, nature.
- Location & Movement: Latitude, longitude, distance to land, storm speed, direction.
- Intensity: Wind speed, central pressure, cyclone category from various agencies.
- Structural Features: Wind radii (R34, R50, R64), eye radius, POCI, ROCI, RMW.
- Agency Observations: Gusts, sea height, tracking IDs, and other agency-specific data.

B. EXPLORATION OF IBTrACS DATA:

- DATASET OVERVIEW: Checked the structure, total number of observations, and types of attributes (numeric, categorical, datetime).
- MISSING VALUES: Identified and handled missing data in coordinates, wind speed, and pressure using filling or interpolation.
- STATISTICAL ANALYSIS: Calculated mean, min, max, and standard deviation for key numeric attributes like wind speed, pressure, latitude, and longitude; detected outliers.

- TEMPORAL ANALYSIS: Analyzed cyclone frequency per year and per season to identify trends and peak activity periods.
- CORRELATION ANALYSIS: Observed relationships such as negative correlation between wind speed and pressure, and patterns between cyclone intensity and location/basin.
- KEY INSIGHTS: Cyclones mostly occur in specific latitudinal bands; certain basins show higher frequency and intensity; anomalies were identified to guide preprocessing and feature engineering.

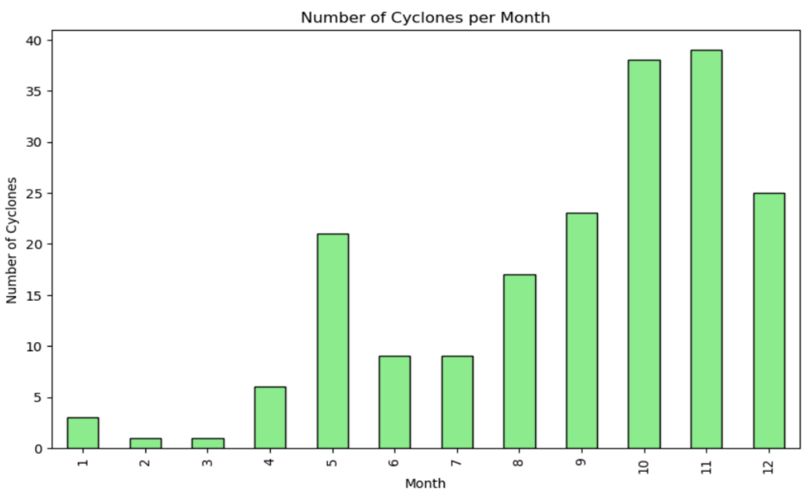
SNAPSHOTS:

- CYCLONES PER YEAR:

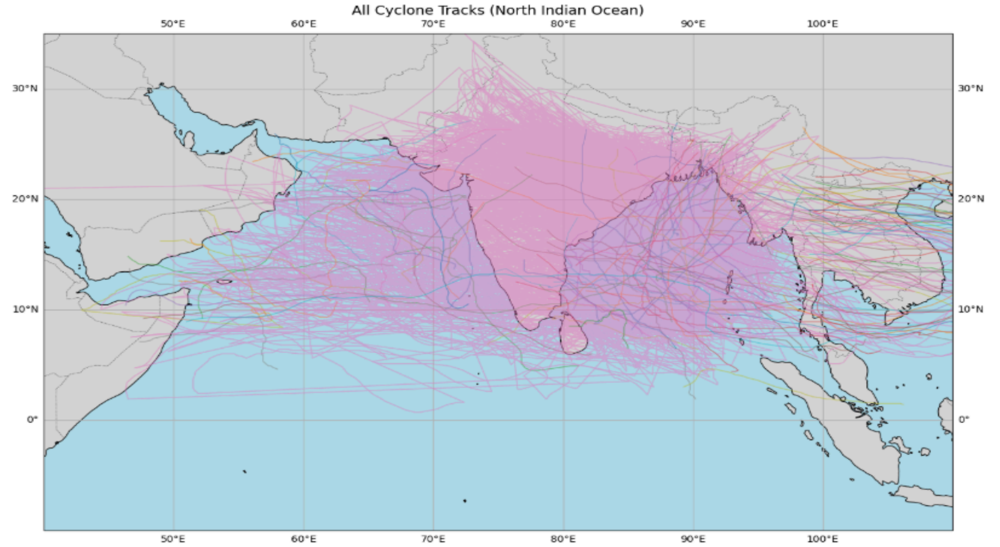
```
[16]: cyclones_per_year = df_clean.groupby("SEASON")["NAME"].nunique()
      print("Cyclone count per year:")
      print(cyclones_per_year.tail(10))  # last 10 years
```

```
Cyclone count per year:
SEASON
2015    5
2016    5
2017    5
2018    8
2019    8
2020    7
2021    6
2022    4
2023    7
2024    4
Name: NAME, dtype: int64
```

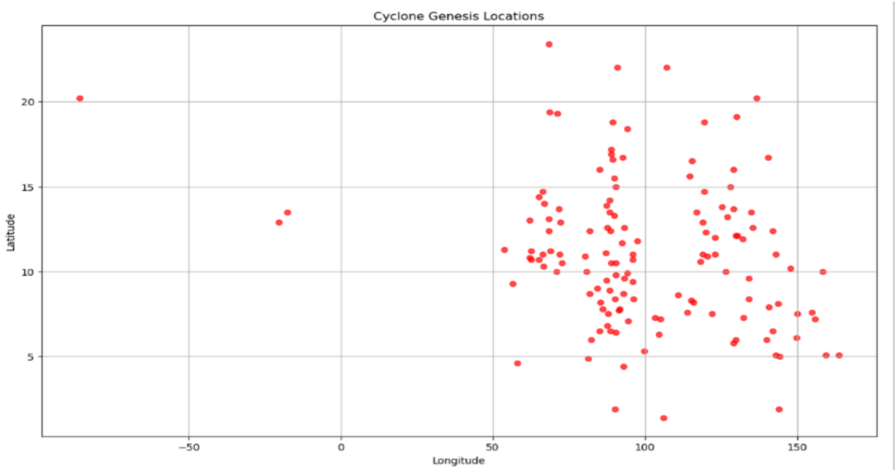
- NUMBERS OF CYCLONE PER MONTH:



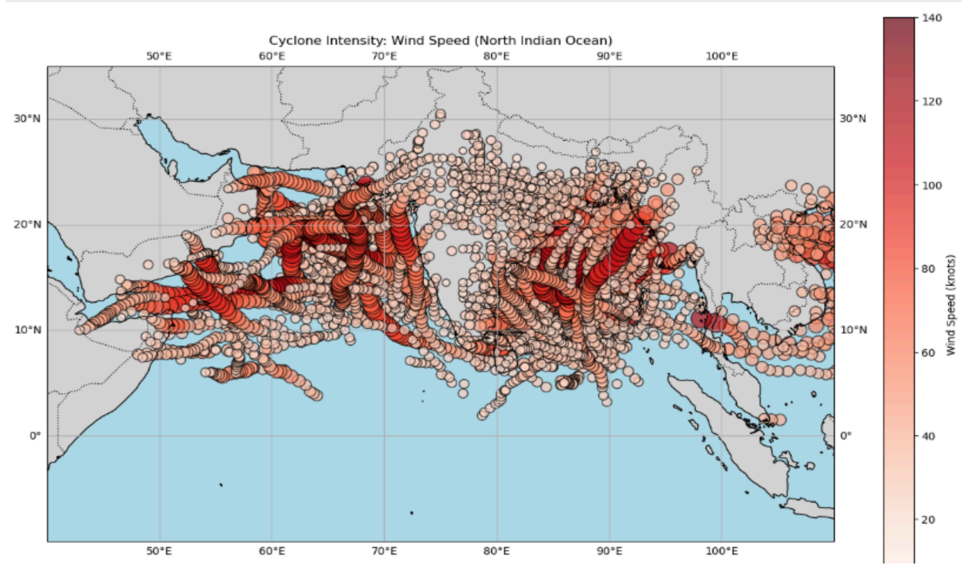
- ALL CYCLONE TRACK (NORTH INDIAN OCEAN):



- CYCLONE GENESIS LOCATION:



- CYCLONE INTENSITY:

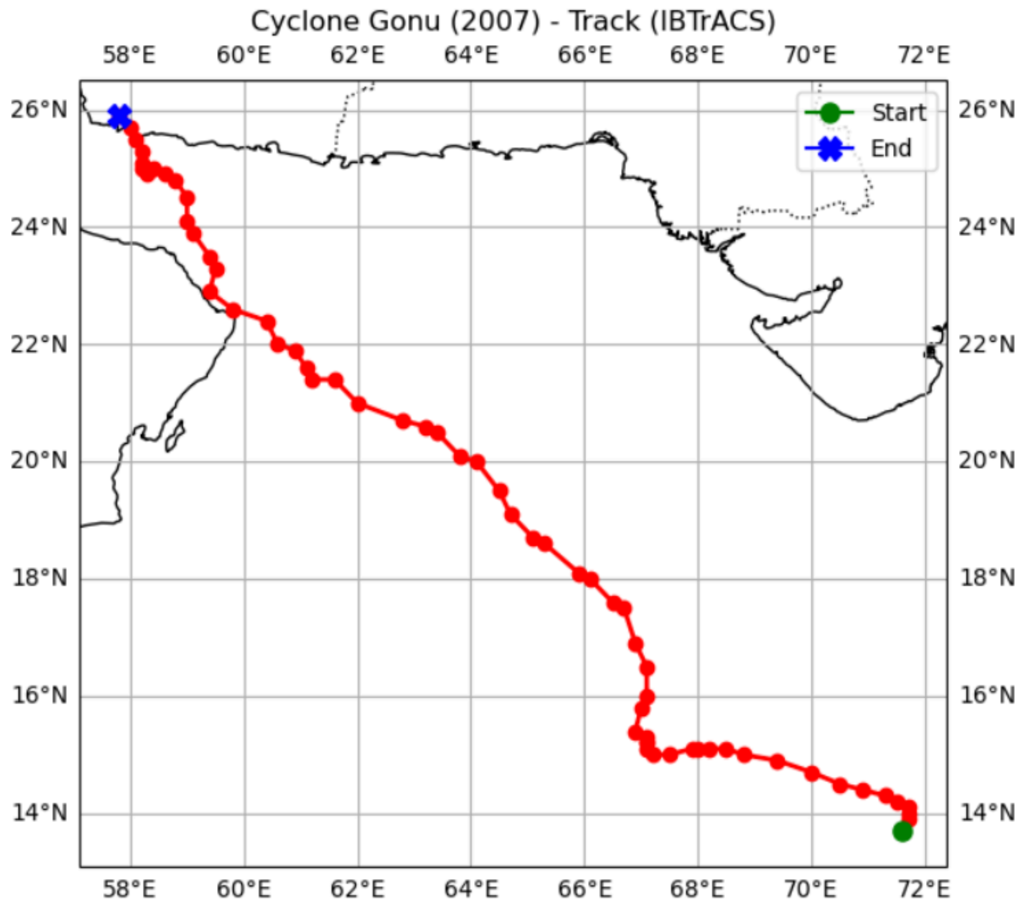


C. VISUALIZATION OF DIFFERENT PAST CYCLONES:

Individual Cyclone Visualization

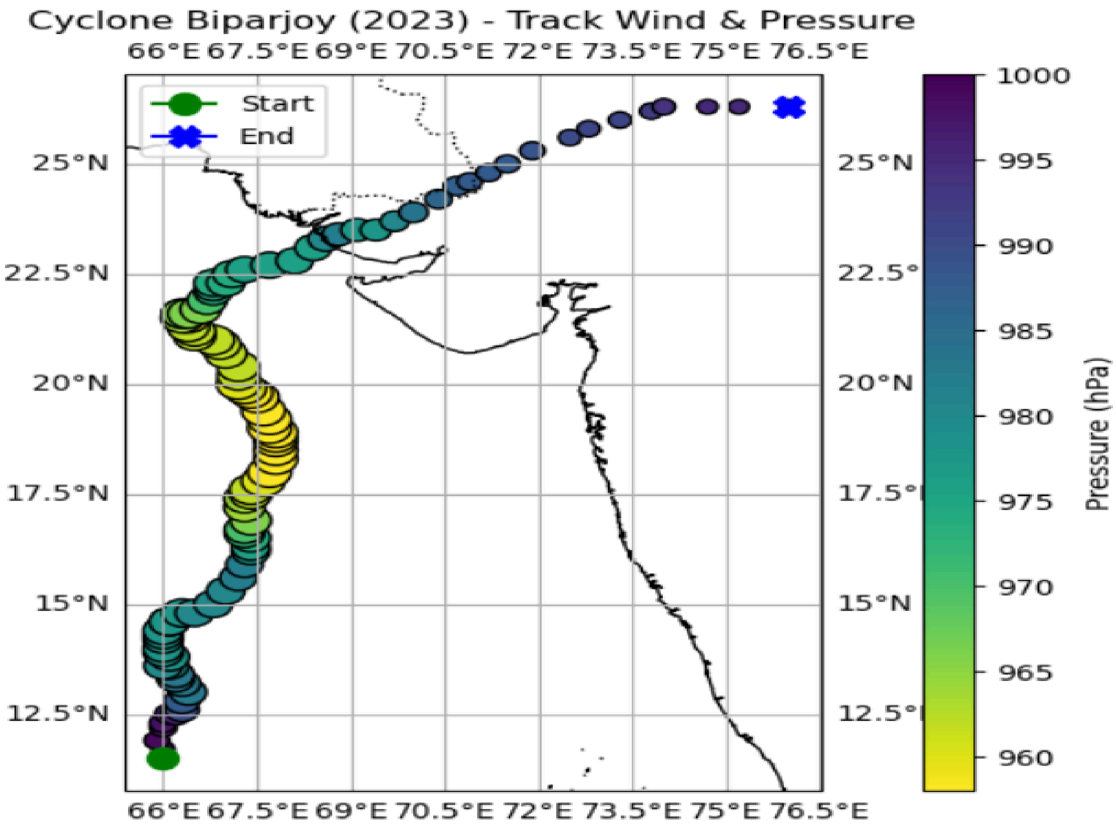
1) Cyclone Gonu:

- Plotted the track (latitude vs longitude) for Gonu using IBTrACS data.
- Color-coded by wind speed to show intensity changes along the track.
- Observed path direction, speed, and landfall points.



2) Cyclone Biparjoy:

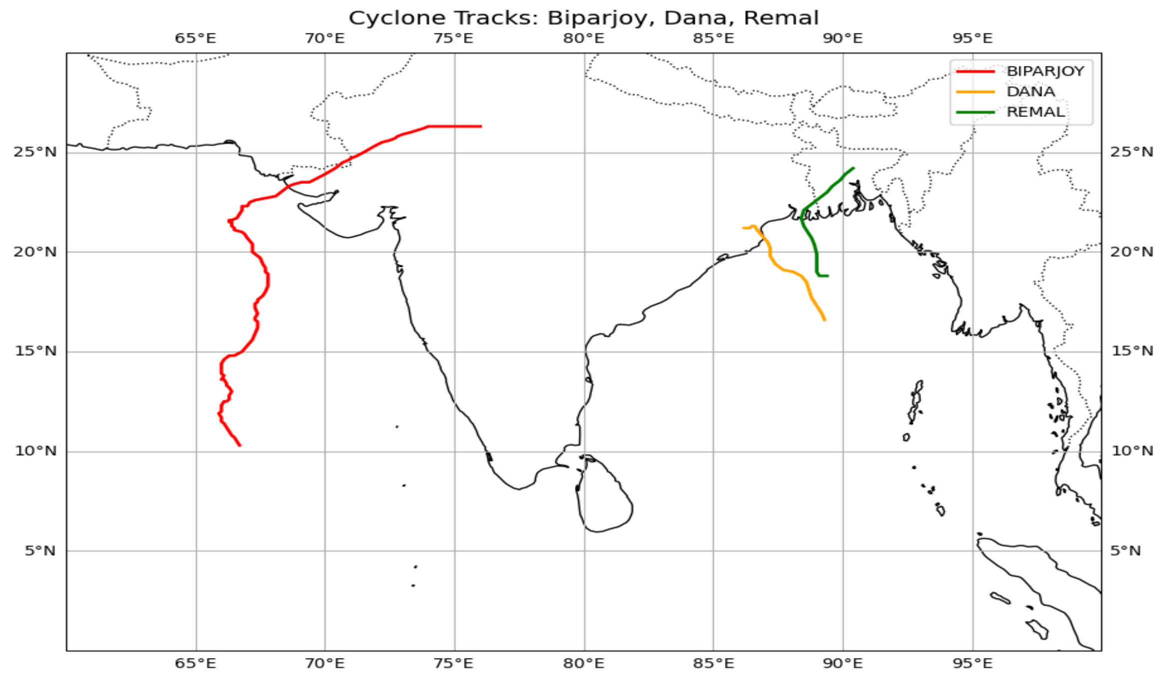
- Tracked its movement individually, highlighting positions at different timestamps.
- Wind speed and pressure visualized along the track.
- Key features like peak intensity and distance to land were highlighted.



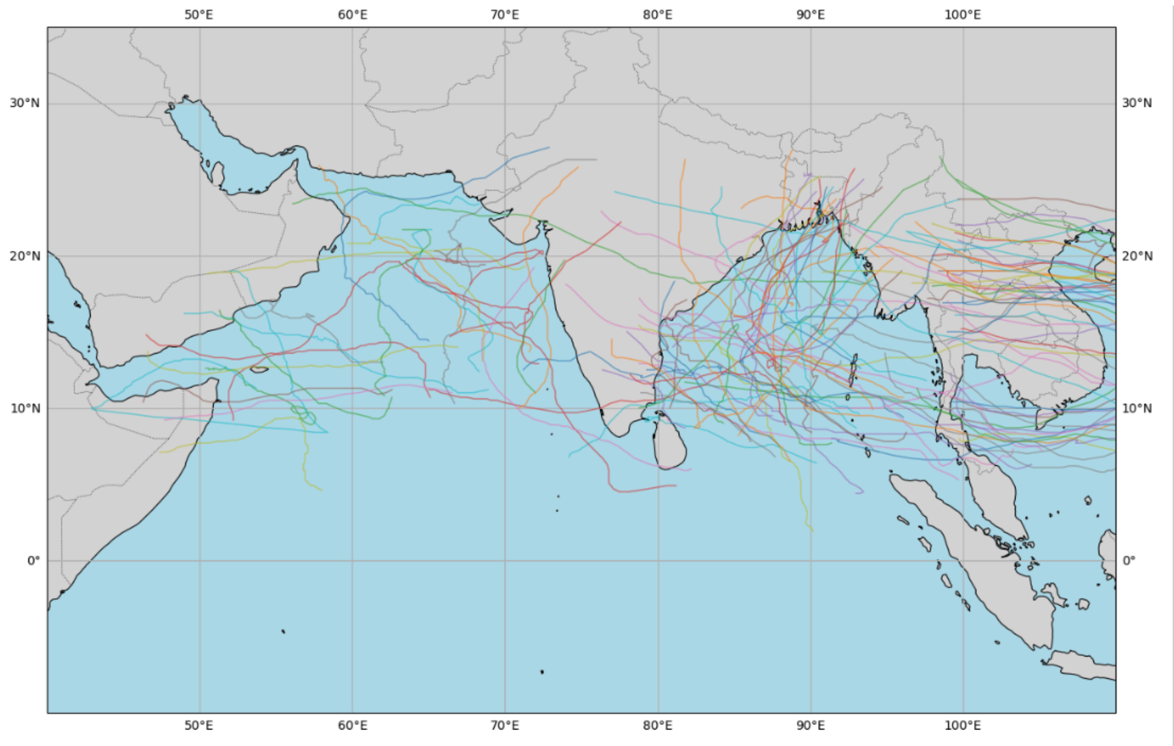
Combined Cyclone Visualization

3) Cyclones Biparjoy, Dana, and Remal:

- Plotted all three cyclone tracks on a single map to compare paths and intensities.
- Used different colors or markers for each cyclone for clear distinction.
- Showed overlapping regions and differences in track directions, speed, and landfall locations.
- Enabled visual comparison of intensity patterns and geographic coverage



4) All Cyclone Tracks

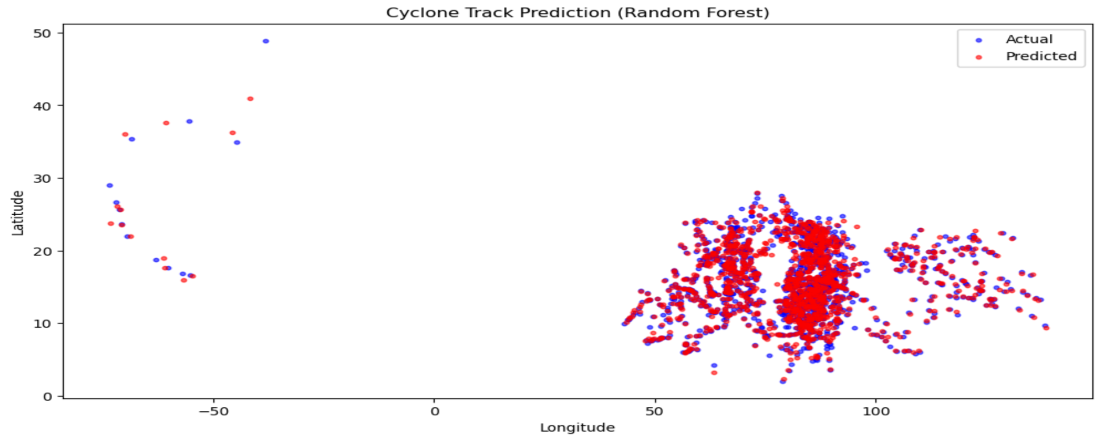


D. MODEL TRAIN FOR IBTrACS DATA

Trained 2 basic models - Random Forest and LSTM, just for IBTrACs data.

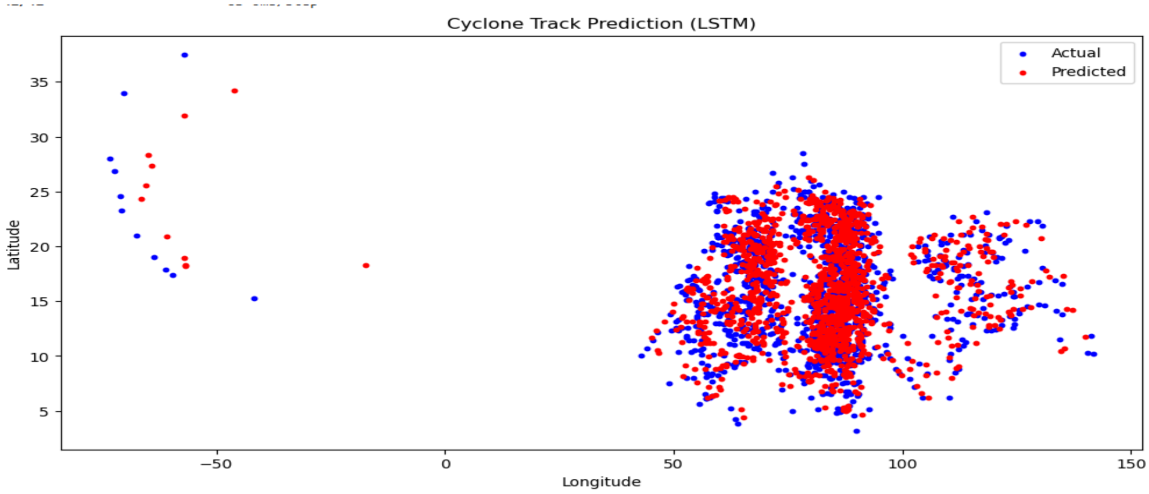
1) Random Forest (RF):

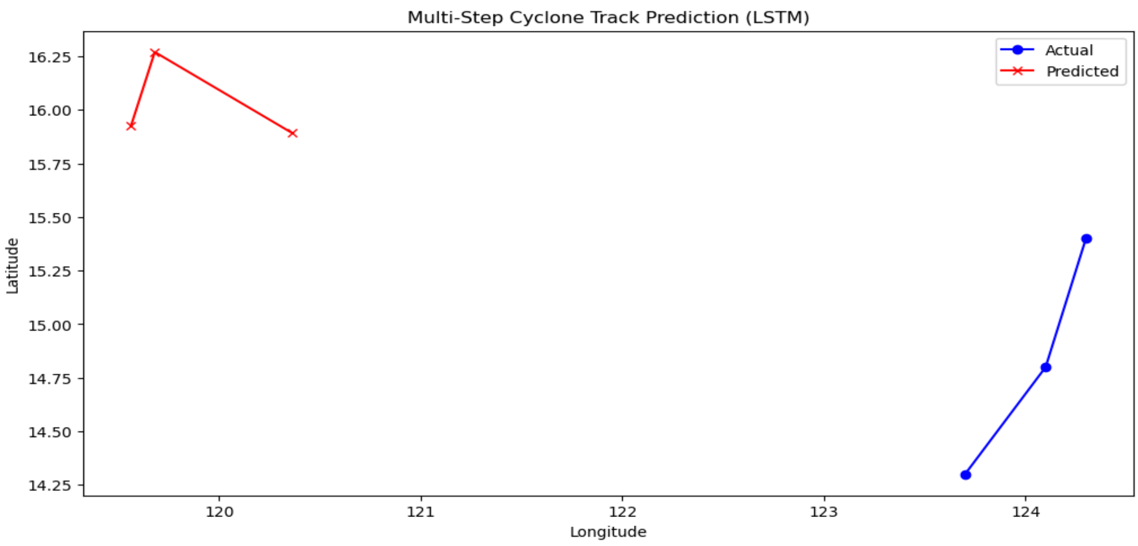
- Used for predicting cyclone intensity or next position based on current features like latitude, longitude, wind speed, and pressure.
- Handles non-linear relationships and is robust to outliers, but cannot capture sequential dependencies in cyclone tracks.



2) Long Short-Term Memory (LSTM):

- Designed to model temporal patterns in cyclone tracks for predicting future positions or intensity.
- Utilizes sequences of past observations, but without environmental data, predictions are often inaccurate.





Performance Observation:

- Both models trained on IBTrACS data alone showed significant deviation between predicted and actual tracks.
- Lack of key environmental variables (SST, wind shear, humidity) limits prediction accuracy.

Conclusion:

- RF is suitable for static, feature-based predictions.
- LSTM captures sequential trends but needs additional environmental data to improve accuracy.
- Hybrid approaches combining track history with reanalysis data are recommended for better cyclone forecasting.

FUTURE WORK:

- Using ERA5 dataset for more accurate cyclone predictions.
- Visualization of ERA5 dataset
- Data Preparation for actual p
- Training models using both IBTrACS track data and ERA5 features.
- After that, start training with basic models (RF, LSTM) and then move to hybrid models for better performance.

CURRENT REQUIREMENT:

- ERA5 Reanalysis Dataset for Atmospheric and Oceanic parameters.
- Climatic Indices - ENSO, IOD, MJO.