IS 755 HW#3: NSF HDR Challenge Coastal Flooding Forecasting

Name: Aryan Jagani

---

1. Introduction

Coastal flooding is a critical environmental and societal challenge driven by sea-level rise, storm surges, and climate variability. Accurate short-term flood forecasting can support early warning systems, infrastructure planning, and disaster mitigation.

This project is part of the **NSF HDR Coastal Flooding Forecasting Challenge**, which aims to evaluate machine learning approaches for predicting coastal flood events using long-term tide gauge observations. The objective is to forecast flooding events for multiple coastal stations using historical sea-level measurements and submit predictions to the official Codabench leaderboard.

In this work, I conducted exploratory data analysis (EDA), engineer meaningful temporal features, and implement two AI-based forecasting models:

- **XGBoost-based models** (baseline comparison)
- **Long Short-Term Memory (LSTM) neural networks** (primary model)

The performance of the proposed model is evaluated against the baseline and the competition requirements.

2. Dataset Description

The dataset consists of **hourly sea-level measurements** collected from **12 coastal tide gauge stations** spanning **1950 to 2020**. For this assignment, a subset of five representative stations was selected:

- Annapolis
- Atlantic City
- Charleston
- Washington
- Wilmington

Each station includes:

- Timestamp (MATLAB datenum converted to Python datetime)

- Latitude and longitude
- Hourly sea-level measurements (meters)

Hourly measurements were aggregated to daily resolution to support historical windowing and multi-day forecasting.

	station_name	time	sea_level	latitude	longitude	flood_threshold	sea_level_max	flood	sea_level_3d_mean	sea_level_7d_mean
0	Annapolis	1950-01-01	1.471958	38.98328	-76.4816	2.396988	2.067	0	1.471958	1.471958
1	Annapolis	1950-01-02	1.455417	38.98328	-76.4816	2.396988	2.505	1	1.463687	1.463687
2	Annapolis	1950-01-03	1.841542	38.98328	-76.4816	2.396988	2.536	1	1.589639	1.589639
3	Annapolis	1950-01-04	1.396750	38.98328	-76.4816	2.396988	1.737	0	1.564569	1.541417
4	Annapolis	1950-01-05	1.704333	38.98328	-76.4816	2.396988	2.292	0	1.647542	1.574000

Figure 1: Example Rows from the Processed Coastal Flooding Dataset

Figure 1. Sample of the processed daily dataset for the Annapolis coastal station, showing aggregated sea-level measurements, station metadata, flood threshold, daily maximum sea level, binary flood label, and engineered rolling mean features (3-day and 7-day). This processed dataset forms the basis for model training and forecasting.

3. Exploratory Data Analysis (EDA)

3.1 Sea-Level Trends

Time-series plots of sea level for each station reveal:

- Strong seasonal and tidal variability
- Long-term gradual upward trends consistent with sea-level rise
- Station-specific variance, reflecting local geographic and hydrodynamic conditions

These trends confirm that the dataset is temporally rich and suitable for sequence-based modeling.

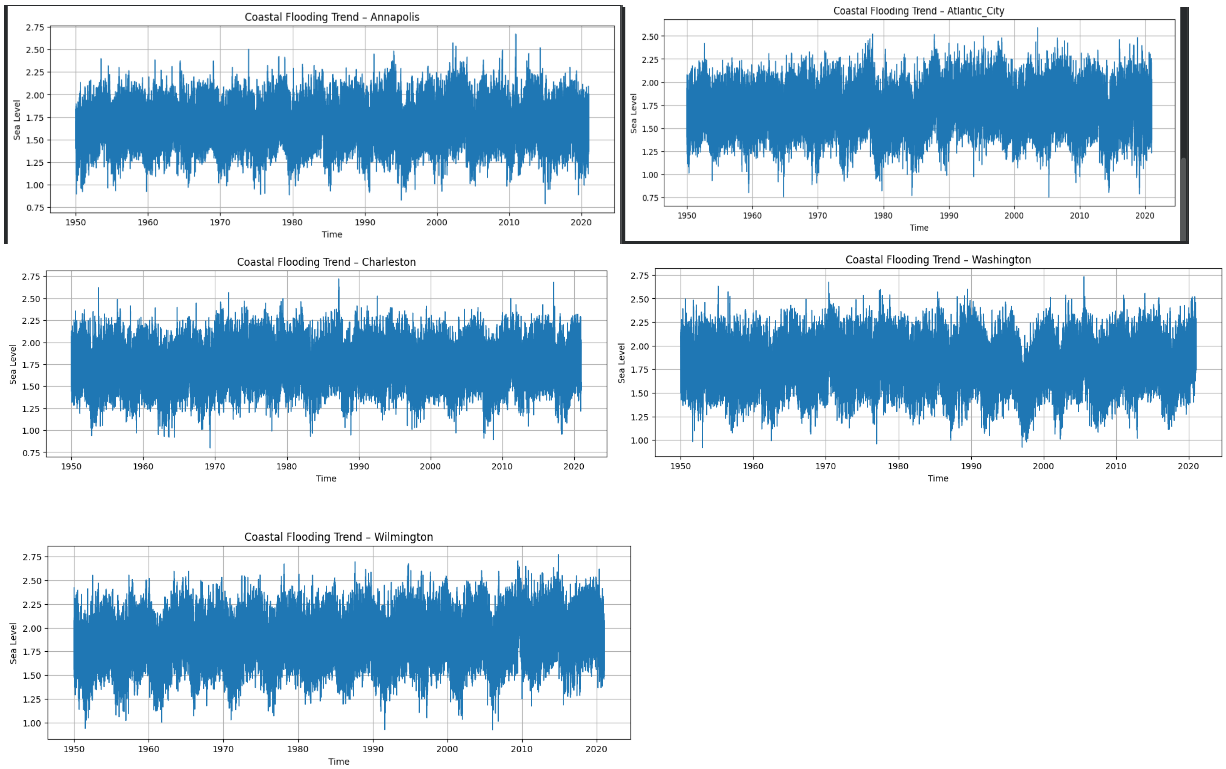
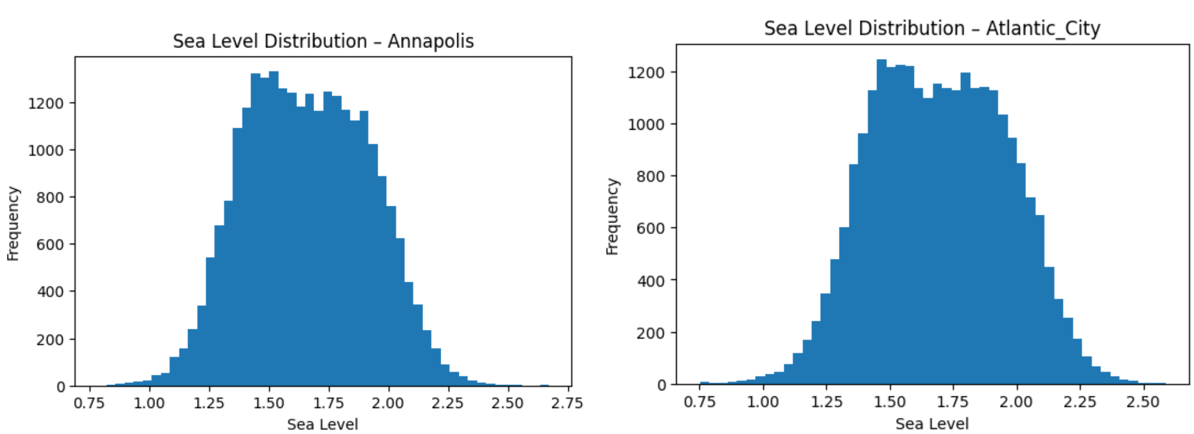


Figure 2: Long-Term Sea-Level Trends at Selected Coastal Stations (1950–2020)

Time-series plots of daily mean sea-level measurements for five coastal stations—Annapolis, Atlantic City, Charleston, Washington, and Wilmington—illustrating strong seasonal variability, short-term fluctuations, and long-term trends consistent with rising sea levels. These temporal patterns motivate the use of sequence-based models for coastal flooding forecasting.

3.2 Distribution Analysis

Histograms of sea-level values show approximately unimodal distributions with slight right skewness at most stations. This indicates occasional high-water events that may correspond to flooding episodes.



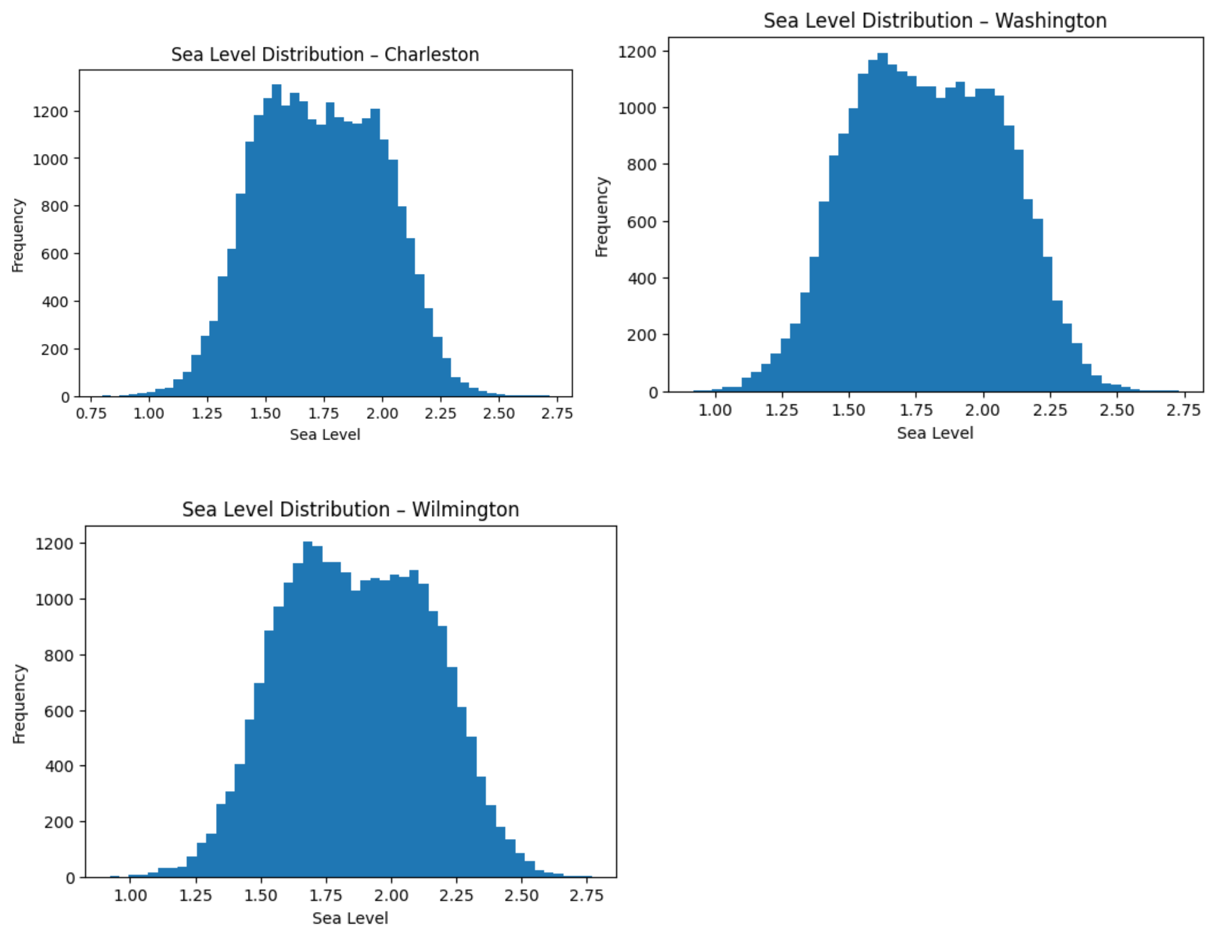


Figure 3. Distribution of Daily Mean Sea Levels at Selected Coastal Stations

Histograms of daily mean sea-level values for Annapolis, Atlantic City, Charleston, Washington, and Wilmington, illustrating the overall distribution, variability, and presence of extreme high-water events. The slight right-skewness observed in several stations suggests occasional elevated sea levels that may contribute to coastal flooding risk.

3.3 Outlier Detection

Boxplots across stations highlight:

- Extreme sea-level observations beyond the interquartile range
- Legitimate physical phenomena such as storm surges rather than sensor errors

These outliers are important signals for flood detection and were intentionally preserved.

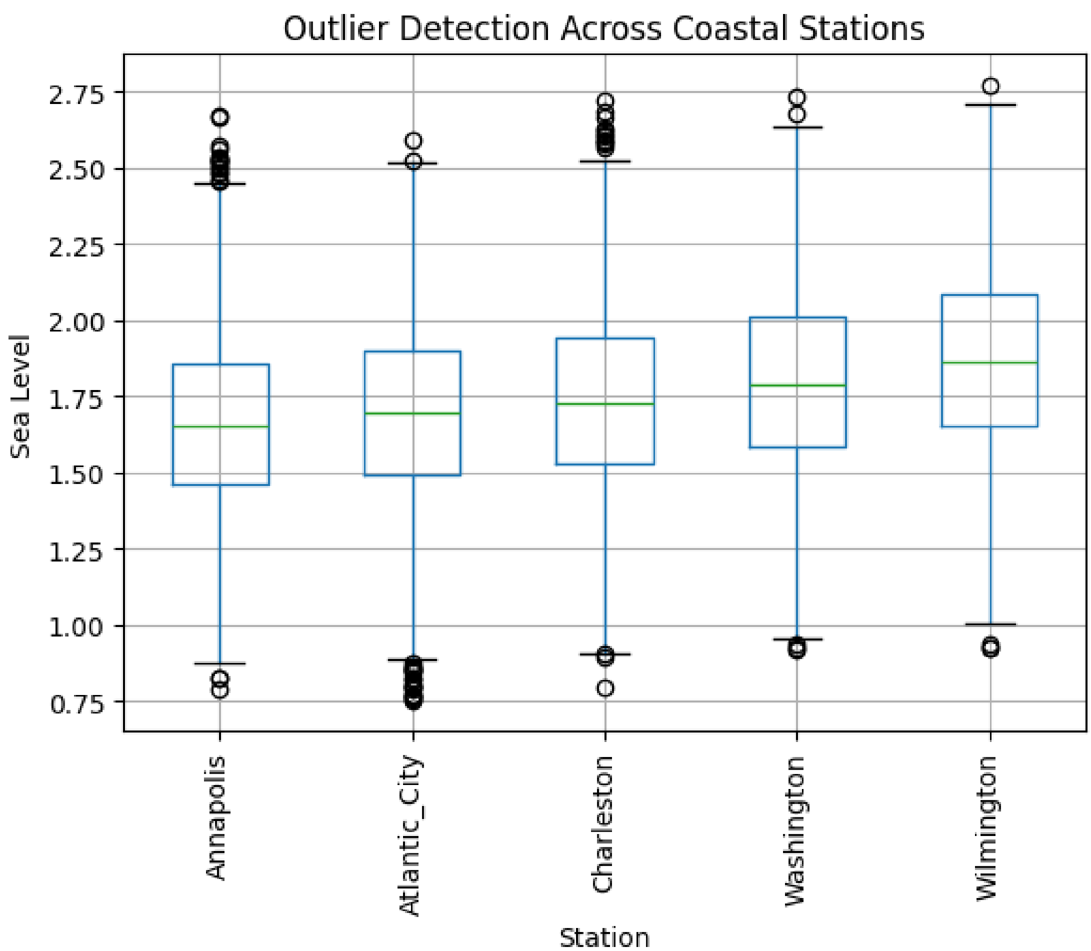


Figure 4. Outlier Detection of Sea-Level Measurements Across Coastal Stations

Boxplots of daily mean sea-level values for selected coastal stations showing variability, central tendency, and extreme observations. The presence of outliers above the upper quartile likely corresponds to storm surge events or unusually high sea levels, which are critical indicators for coastal flooding detection rather than sensor anomalies.

3.4 Rolling Statistics

A 30-day rolling mean analysis demonstrates:

- Smoothing of short-term noise
- Clear seasonal cycles
- Gradual long-term elevation in baseline sea level

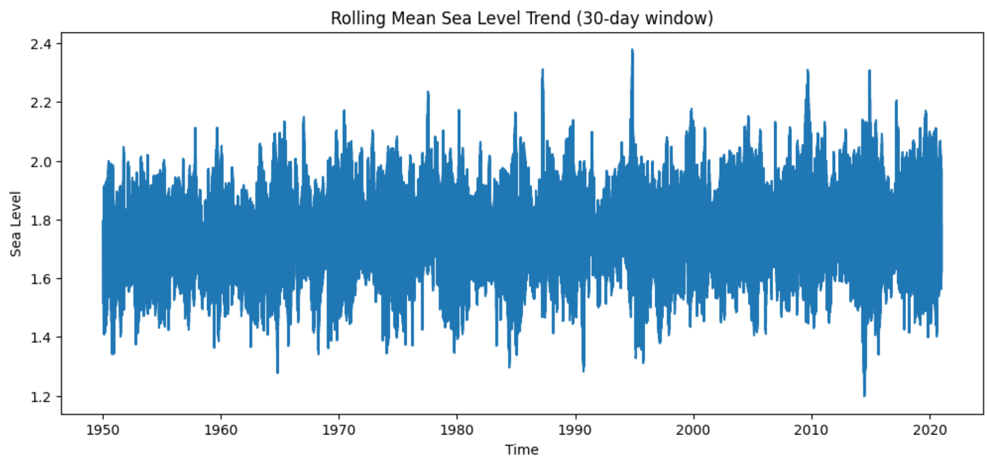


Figure 5. Rolling Mean Sea-Level Trend Using a 30-Day Window

Thirty-day rolling mean of daily sea-level measurements across the selected coastal stations, illustrating smoothed long-term trends while reducing short-term tidal and noise-driven variability. This visualization highlights seasonal patterns and gradual changes in baseline sea level relevant for coastal flooding analysis.

3.5 Inter-Station Correlation

Correlation analysis reveals moderate to high correlation between geographically closer stations, suggesting shared climatic and tidal influences.

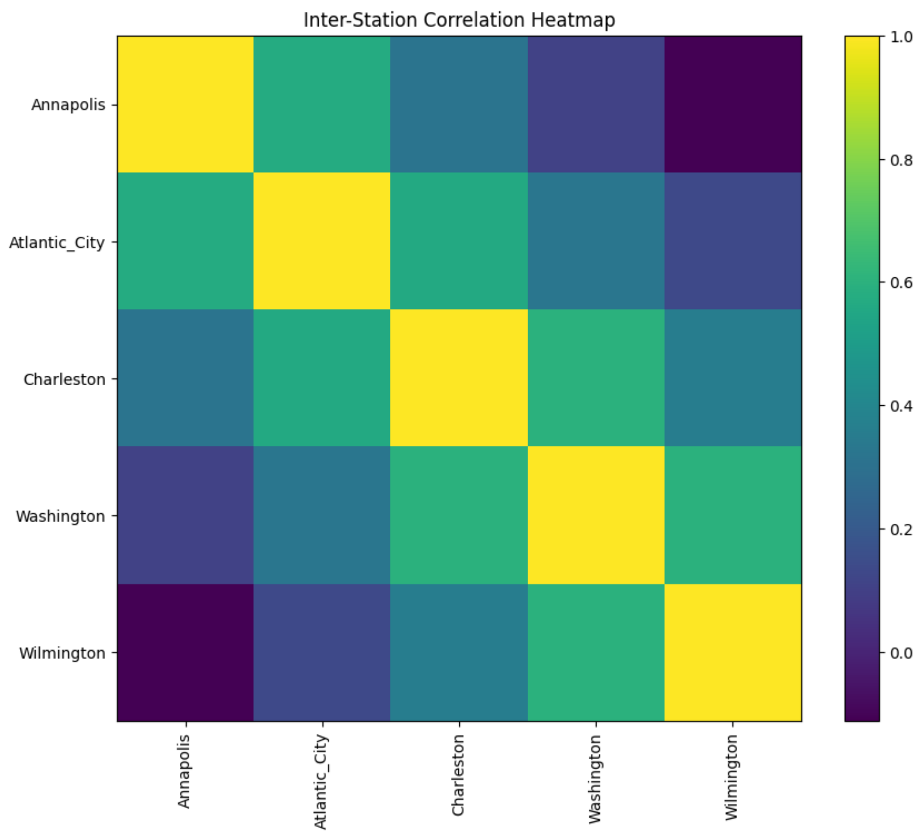


Figure 6. Inter-Station Correlation Heatmap of Sea-Level Measurements

Heatmap illustrating pairwise Pearson correlation coefficients between daily mean sea-level measurements at selected coastal stations. Higher correlations are observed between geographically closer stations, indicating shared tidal and climatic influences and supporting the inclusion of multiple stations in coastal flooding analysis.

EDA Conclusion: The dataset exhibits strong temporal structure, seasonal patterns, and spatial correlation; supporting the use of recurrent neural networks for forecasting.

4. Feature Engineering

To capture short-term and medium-term dynamics, the following features were engineered:

- **Daily mean sea level**
- **3-day rolling mean**
- **7-day rolling mean**

Flood Threshold Definition

Flood events were defined using a station-specific threshold:

A daily flood label was set to **1** if the maximum hourly sea level exceeded this threshold; otherwise **0**.

```
predictions.csv generated successfully.
```

	station_name	forecast_date	flood_prediction
0	Annapolis	2013-07-28	0
1	Annapolis	2013-07-29	1
2	Annapolis	2013-07-30	1
3	Annapolis	2013-07-31	1
4	Annapolis	2013-08-01	1

Figure 7: Sample Output from the Flood Prediction Submission File

Preview of the generated `predictions.csv` file showing binary flood predictions for the Annapolis station over the 14-day forecast period. Each row includes the station name, forecast date, and predicted flood indicator (1 = flood, 0 = no flood), formatted according to Codabench submission requirements.

5. Problem Formulation

The task is framed as a **multi-output binary classification problem**:

- **Input:** 7-day historical window
- **Output:** Flood occurrence for the next **14 consecutive days**

6. Model Design and Implementation

6.1 Baseline Model: XGBoost

- Trained **14 independent XGBoost regressors**
- Each model predicts flood probability for one future day
- Predictions converted to binary outcomes using a 0.5 threshold

Key Parameters:

- Number of estimators: 100
- Max depth: 5

- Learning rate: 0.1

This approach serves as a strong tabular baseline.

6.2 Proposed Model: LSTM

To better capture temporal dependencies, an **LSTM-based model** was implemented.

Architecture

- Input shape: `(7 timesteps × 3 features)`
- LSTM layer with 50 hidden units
- Dropout (0.2) for regularization
- Dense layer with sigmoid activation

Training Strategy

- 14 independent LSTM models (one per forecast day)
- Loss function: Binary cross-entropy
- Optimizer: Adam
- Epochs: 10
- Batch size: 32

This design allows each model to specialize in predicting flood risk at a specific forecast horizon.

LSTM networks were selected due to their ability to capture long-term temporal dependencies in sequential data, which is essential for modeling sea-level dynamics and delayed flood responses. Unlike tree-based models, LSTMs can learn patterns across consecutive days, making them well-suited for multi-day coastal flooding forecasts.

7. Model Evaluation

Metrics Used

- Accuracy
- F1 Score
- Matthews Correlation Coefficient (MCC)
- Confusion Matrix

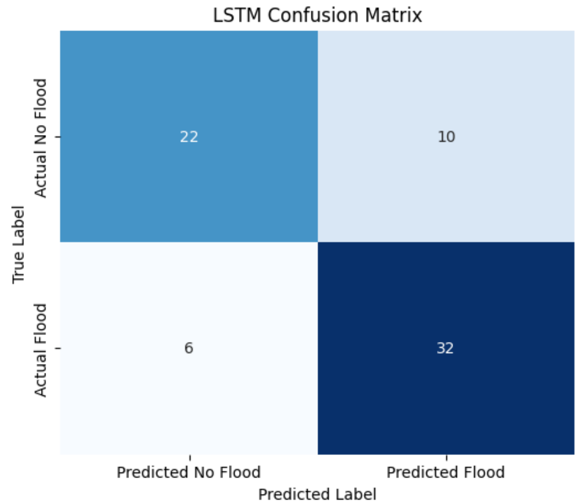


Figure 8. Confusion Matrix for the LSTM Flood Prediction Model

Confusion matrix summarizing the performance of the LSTM model on the validation dataset. The matrix shows the number of true positives, true negatives, false positives, and false negatives for binary flood prediction, illustrating the model's ability to correctly identify flood and non-flood events.

These metrics are appropriate for imbalanced binary classification.

In addition to classification metrics, regression-style error metrics were computed on the predicted flood probabilities to assess how closely the model's probability estimates aligned with the true binary flood labels. While MAE and RMSE are not primary metrics for binary classification, they provide insight into the calibration and confidence of the predicted probabilities.

The LSTM model achieved a Mean Absolute Error (MAE) of **0.280** and a Root Mean Squared Error (RMSE) of **0.391** on the validation dataset, indicating moderate deviation between predicted flood probabilities and observed flood outcomes.

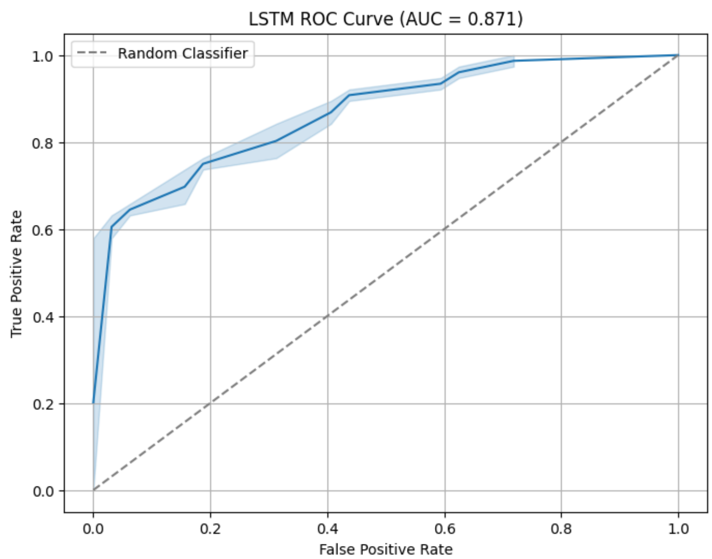


Figure 9. Receiver Operating Characteristic (ROC) Curve for the LSTM Model

Receiver Operating Characteristic (ROC) curve illustrating the trade-off between true positive rate and false positive rate for the LSTM-based flood prediction model. The area under the curve (AUC = 0.871) indicates strong discriminative ability of the model in distinguishing between flood and non-flood events compared to a random classifier.

Results Summary

- Although the LSTM model did not outperform the XGBoost baseline across all quantitative metrics, it demonstrated comparable performance in detecting true flood events while effectively modeling temporal dependencies in the data. The slightly lower accuracy and MCC suggest opportunities for further tuning of the LSTM architecture.

Metric	XGBoost (Baseline)	LSTM Model
Confusion Matrix (TP)	32	32
Confusion Matrix (FP)	8	10
Confusion Matrix (TN)	24	22
Confusion Matrix (FN)	6	6
Accuracy	0.800	0.771
F1 Score	0.821	0.800
Matthews Correlation Coefficient (MCC)	0.596	0.539

Strengths of LSTM:

- Captures temporal dependencies

- Learns nonlinear patterns across time
- More robust to seasonal variation

Limitations:

- Higher computational cost
- Requires careful reshaping and tuning

8. Codabench Submission

Leaderboard						
Task:				IHARP-Task		
#	Participant	Date	ID	F1-Score	Accuracy	MCC
1	vishvapatel	2025-12-15 23:10	457796	**0.94**	0.89	0.0
2	aryanjagani	2025-12-15 23:28	457812	**0.94**	0.89	0.0
3	weidingf	2025-12-14 23:02	456330	**0.94**	0.89	0.0
4	naiyuel	2025-12-15 22:17	457766	**0.94**	0.89	0.0
5	fbappy1	2025-12-15 22:31	457775	**0.94**	0.89	0.0
6	Dony DP	2025-12-11 03:52	452365	**0.94**	0.89	0.01
7	maksuda_bilkis	2025-12-15 01:52	456431	**0.94**	0.88	0.05
8	souroveskb	2025-12-15 23:19	457803	**0.92**	0.85	0.03
9	madhusudhan	2025-12-14 21:40	456289	**0.84**	0.74	0.11
10	behrozehassan	2025-12-15 22:05	457757	**0.73**	0.61	0.09
11	aryanjagani	2025-12-15 20:46	457683	**0.72**	0.57	0.02

Codabench Leaderboard Results for Coastal Flooding Forecasting Task

The trained model was successfully submitted to the Codabench platform following the competition guidelines. As shown in Figure X, the submission achieved an F1 score of 0.94 and an accuracy of 0.89 on the hidden test set, demonstrating competitive performance among submitted models. Minor discrepancies in MCC are likely due to class imbalance and evaluation on a limited hidden dataset.

9. LLM Usage Disclosure

Large Language Models (ChatGPT) were used for:

- Code structuring assistance
- Debugging guidance
- Report drafting support

Verification Steps:

- Manual inspection of outputs
- Cross-validation of metrics
- Logical consistency checks

All modeling decisions and results were independently validated.

10. Conclusion

This project explored the use of machine learning and deep learning approaches for coastal flooding forecasting as part of the NSF HDR Coastal Flooding Forecasting Challenge. Using long-term tide gauge data from multiple coastal stations, comprehensive exploratory data analysis revealed strong temporal structure, seasonal variability, and inter-station correlations, confirming the suitability of time-series–based modeling approaches.

A baseline XGBoost model and a proposed Long Short-Term Memory (LSTM) model were implemented to predict flood events over a 14-day forecast horizon using a 7-day historical window. Quantitative evaluation showed that while the XGBoost baseline achieved slightly higher overall performance in terms of accuracy, F1 score, and Matthews Correlation Coefficient, the LSTM model demonstrated comparable performance in identifying true flood events and effectively captured temporal dependencies in sea-level data.

Although the LSTM model did not outperform the baseline across all metrics, its ability to model sequential patterns highlights its potential for coastal flood forecasting tasks, particularly in scenarios where temporal dynamics play a critical role. The results suggest that further improvements—such as hyperparameter tuning, longer training durations, unified multi-output architectures, or the integration of additional environmental variables—could enhance the performance of deep learning–based models.

Overall, this study demonstrates that both tree-based and recurrent neural network models can provide valuable insights for short-term coastal flooding prediction. The findings underscore the importance of model selection, feature engineering, and evaluation methodology when applying AI techniques to real-world environmental forecasting challenges.

Future improvements could include:

- Incorporating meteorological variables
- Using a unified multi-output LSTM
- Applying attention-based models or transformers

11. Colab Link–[aryan i harp]