ECS 171 WINTER 2025

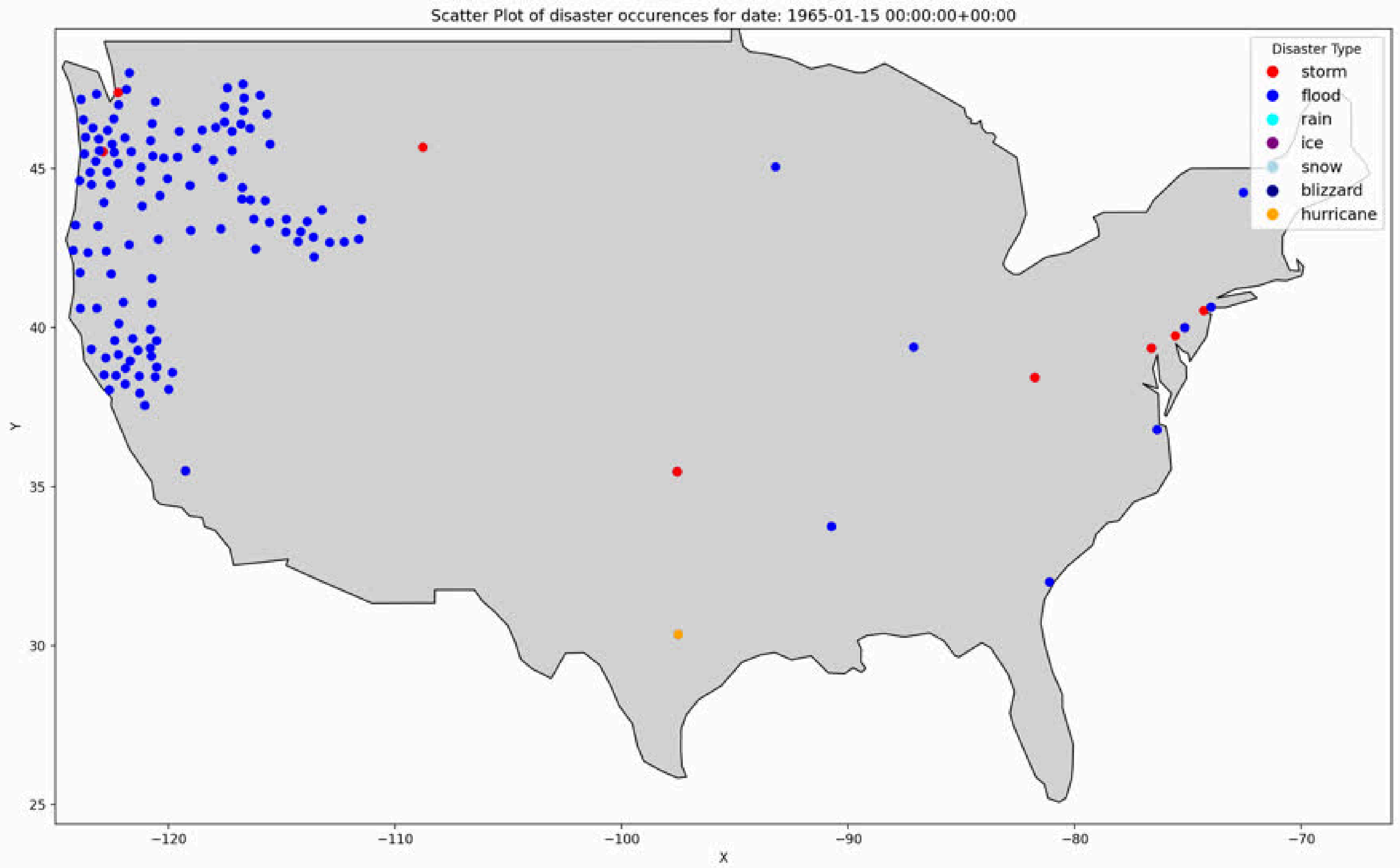
NATURAL DISASTER PREDICTION

By : Ayush Saha, Gavin Lang, Foram Shah, Eileen Huynh, Cassie Hopkin, Meigan Lu, Veronica Hangsan, Manav Gurnani, Jovan Radovic

PROPOSAL

In the state of unpredictable events that has caused widespread destruction to homes and infrastructure, there is a need for systems that can mitigate their impact. This project aims to develop predictive models that forecast location and time of water-based disasters. By analyzing past disaster occurrences, we seek to provide visual insights that can help the others proactively prepare, respond to these catastrophic events, and allocate funding and resources effectively.

DATASET



- FEMA Dataset on Disaster Declarations Summaries
- all federally declared disasters 1953-2025
- focus on water-based disasters: storm, flood, blizzard, hurricane
- EDA: removed instances of hurricane with evacuation location specified

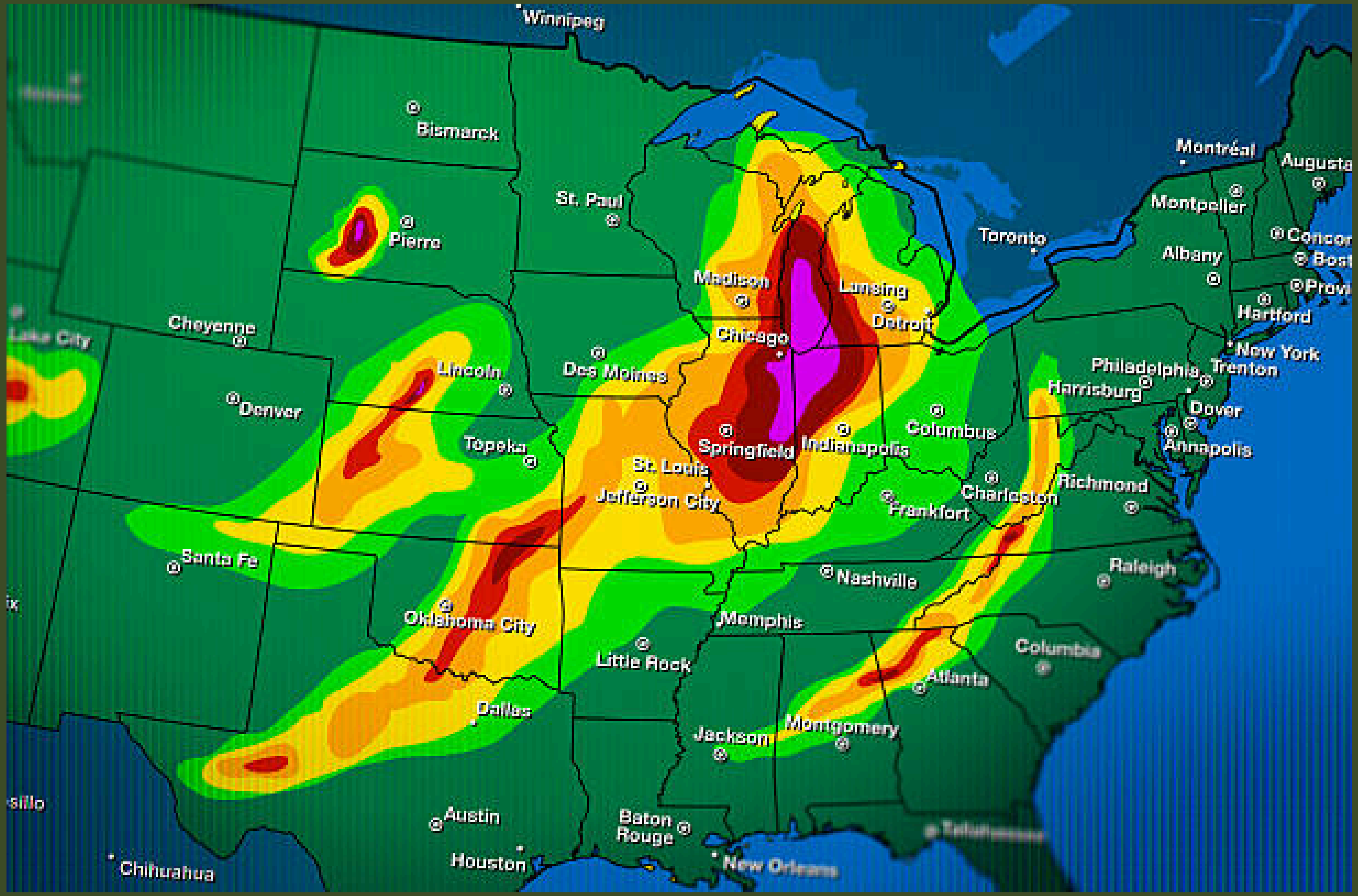
LITERATURE REVIEW

Three Main Perspectives on Predictive Natural Disaster Modeling

- The “Satellite Imagery” view
- The “Weather Data Modeling” view
- The “Spatiotemporal Point Processes” view

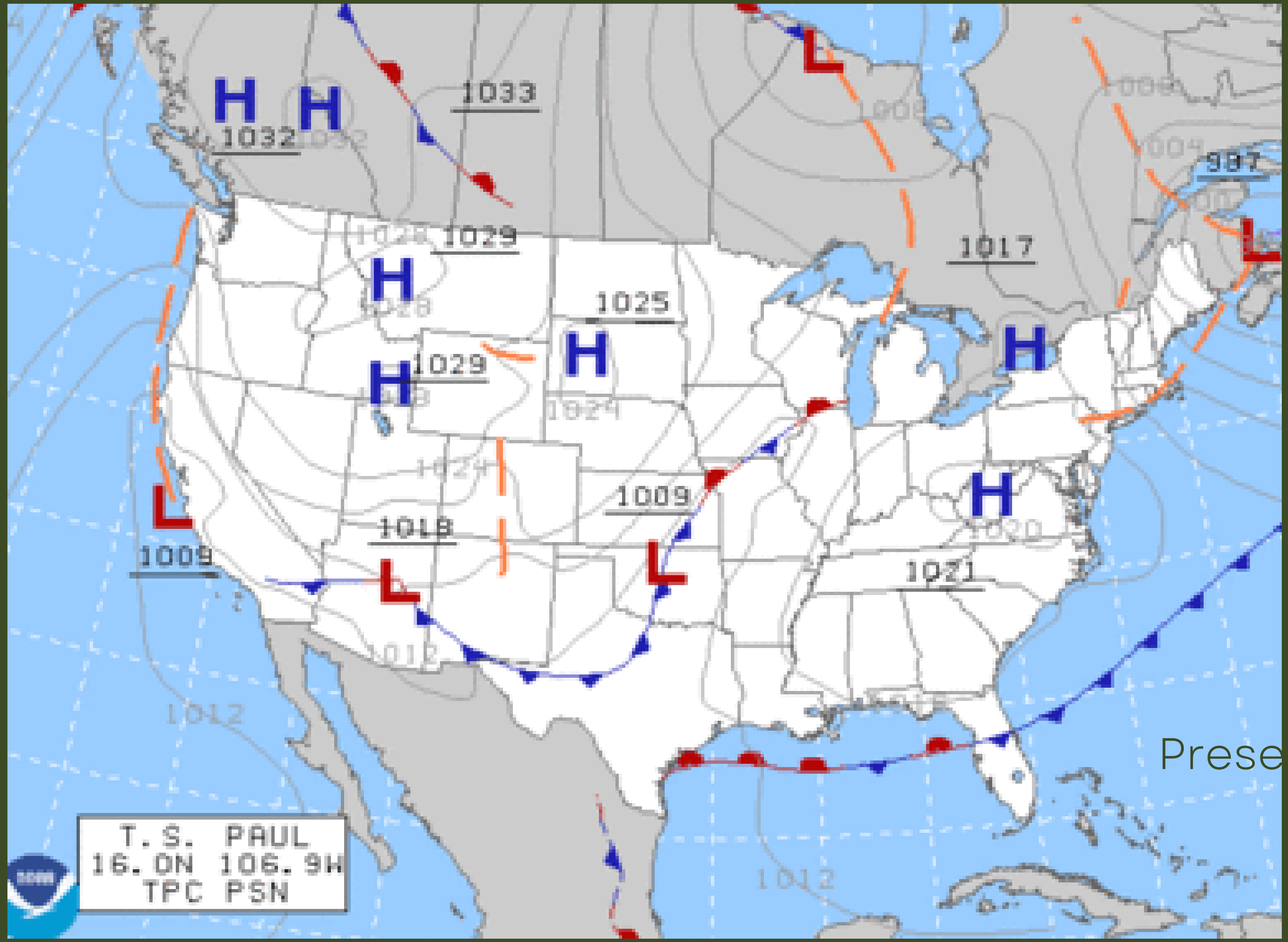
LITERATURE REVIEW

The “Satellite Imagery” View



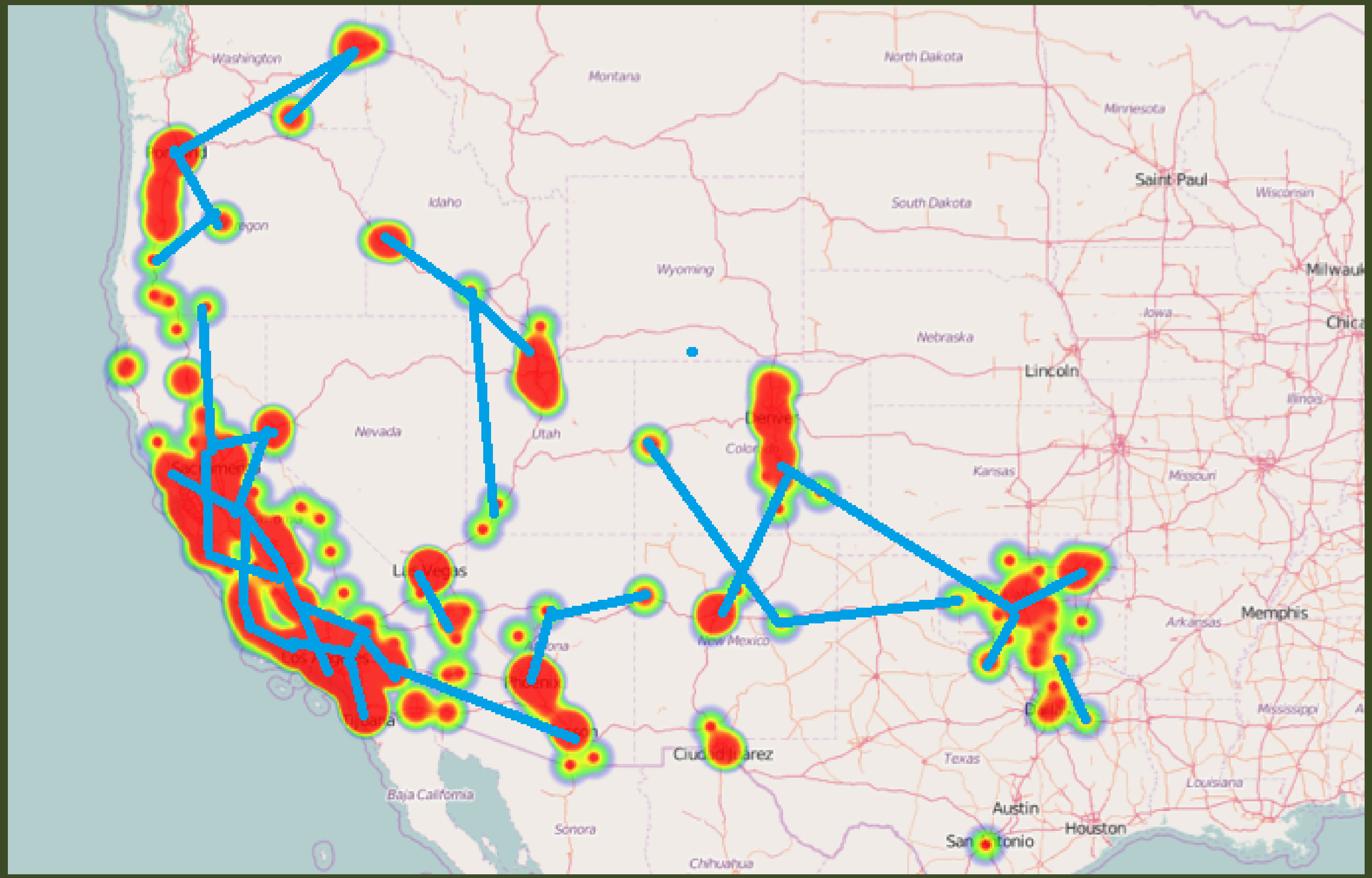
LITERATURE REVIEW

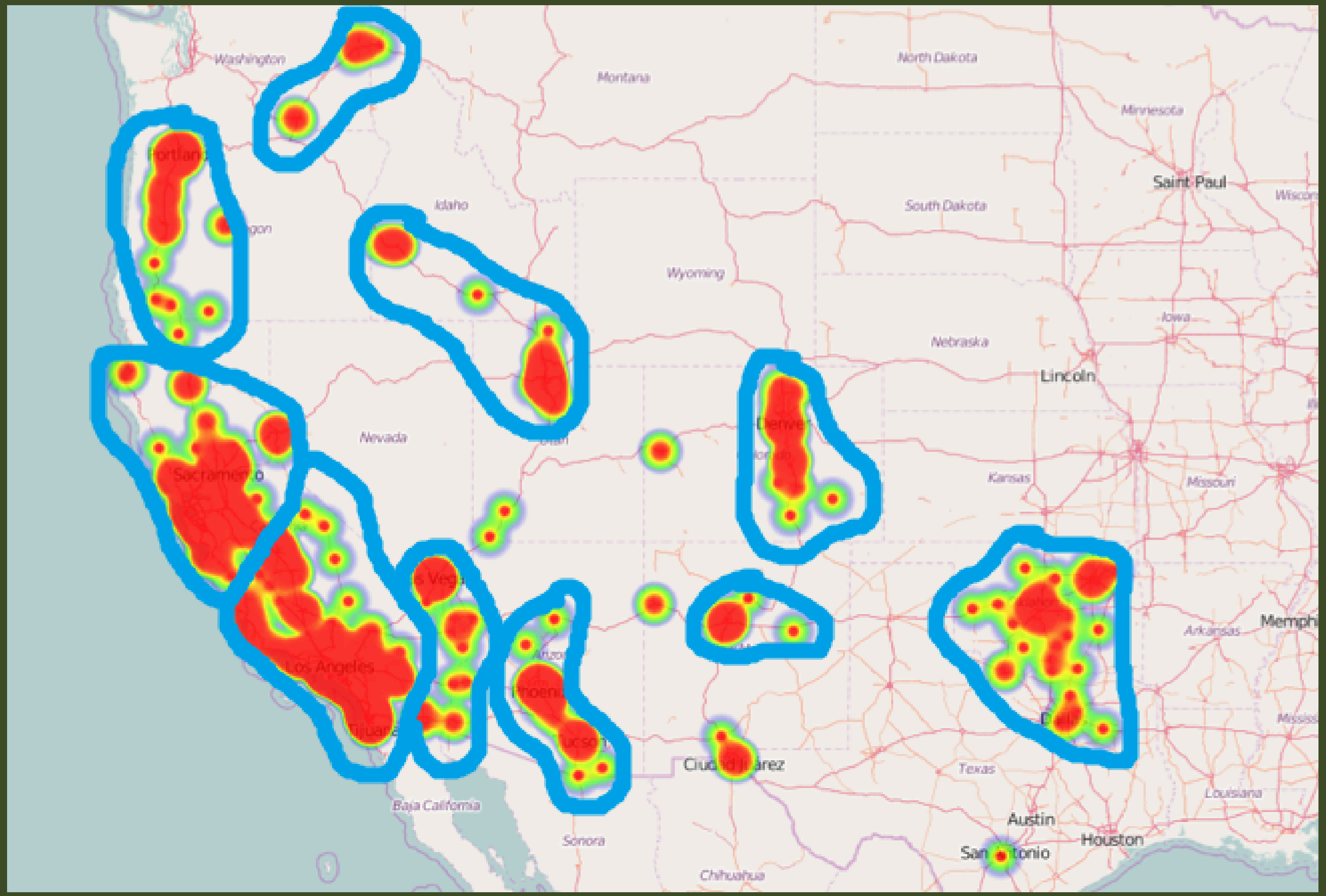
The “Weather Data Modeling” View



LITERATURE REVIEW

The "Spatiotemporal Point Processes" View

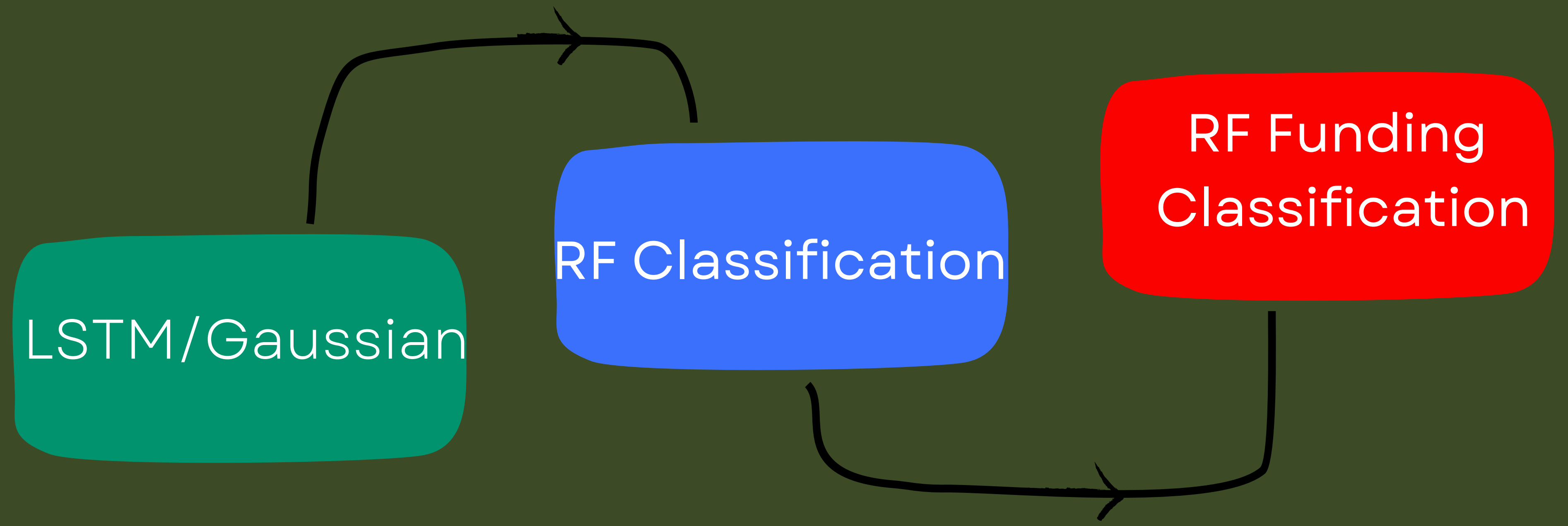




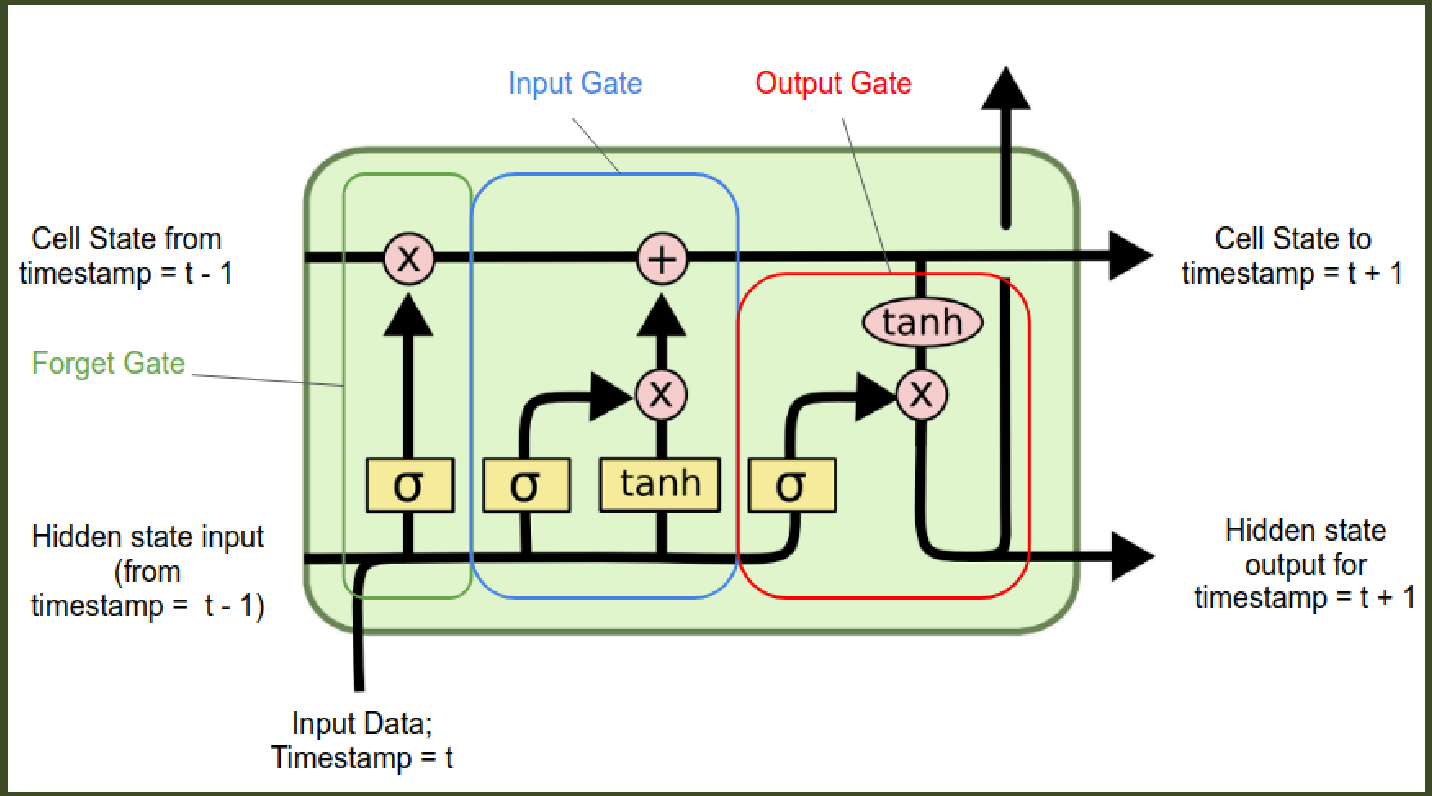
FLOW IDEA

We trained models for three main goals:

- Predicting the likelihood of a water based disaster occuring based on location and time
  - we created two models for this goal: an LSTM and Gaussian Regression Model
- Predicting the type of disaster for a given location and time (season)
- Predicting the funding category based on type of disaster, location, and time



LONG SHORT-TERM MEMORY MODEL



A single LSTM cell

General equation for each gate:

$ft = \sigma(Wf * [ht\text{-}1, xt] + bf)$

Background

- The **goal** of this model is to predict the likelihood of any disaster occurring given a specific time and location.
- RNNs are good at processing sequential data.
- LSTMs are well fit for disaster prediction because they can capture long-term dependencies in temporal data.
- Each LSTM cell has several gates that contribute to what an LSTM chooses to remember or forget.

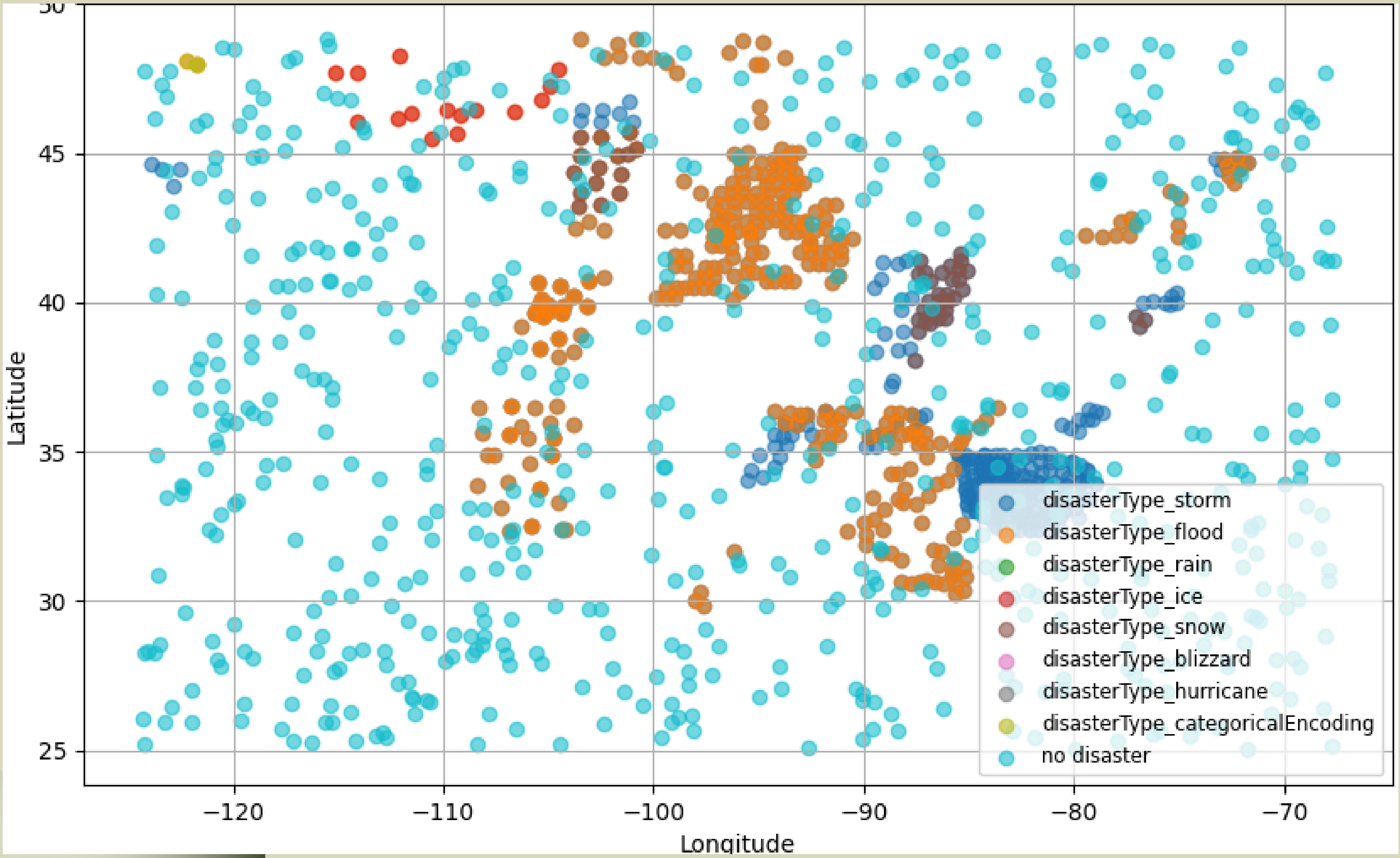
TAKEAWAYS

Used cKDTree to generate no-disaster data points. For a specific day, the LSTM seems to generally predict high disaster probability from mid-west to east coast.

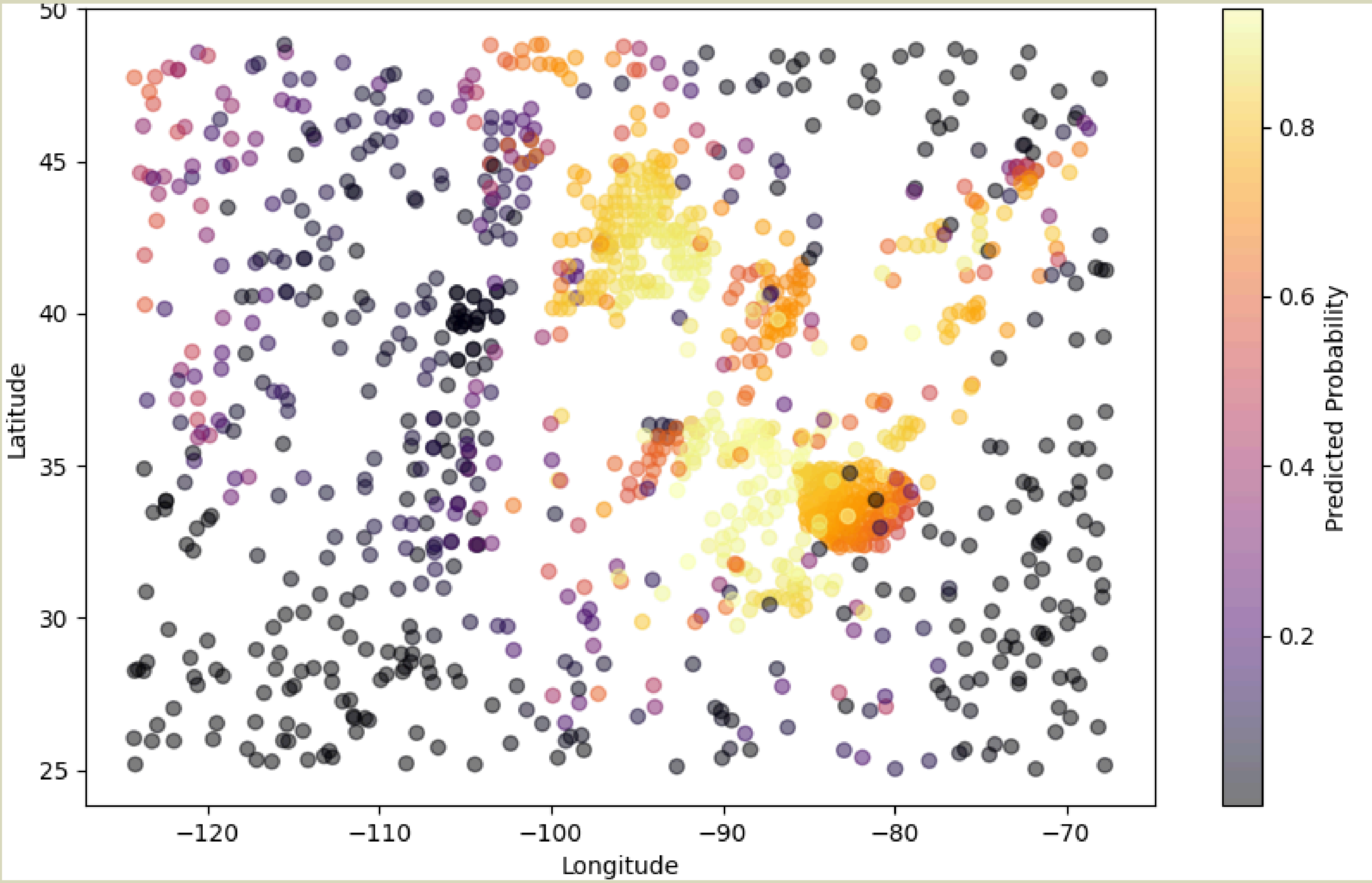
Table 1.1: Deep LSTM Performance Metrics

Model	Accuracy	Precision	Recall	F1-Score
Deep LSTM (5 layers)	0.838	0.8404	0.8382	0.8377

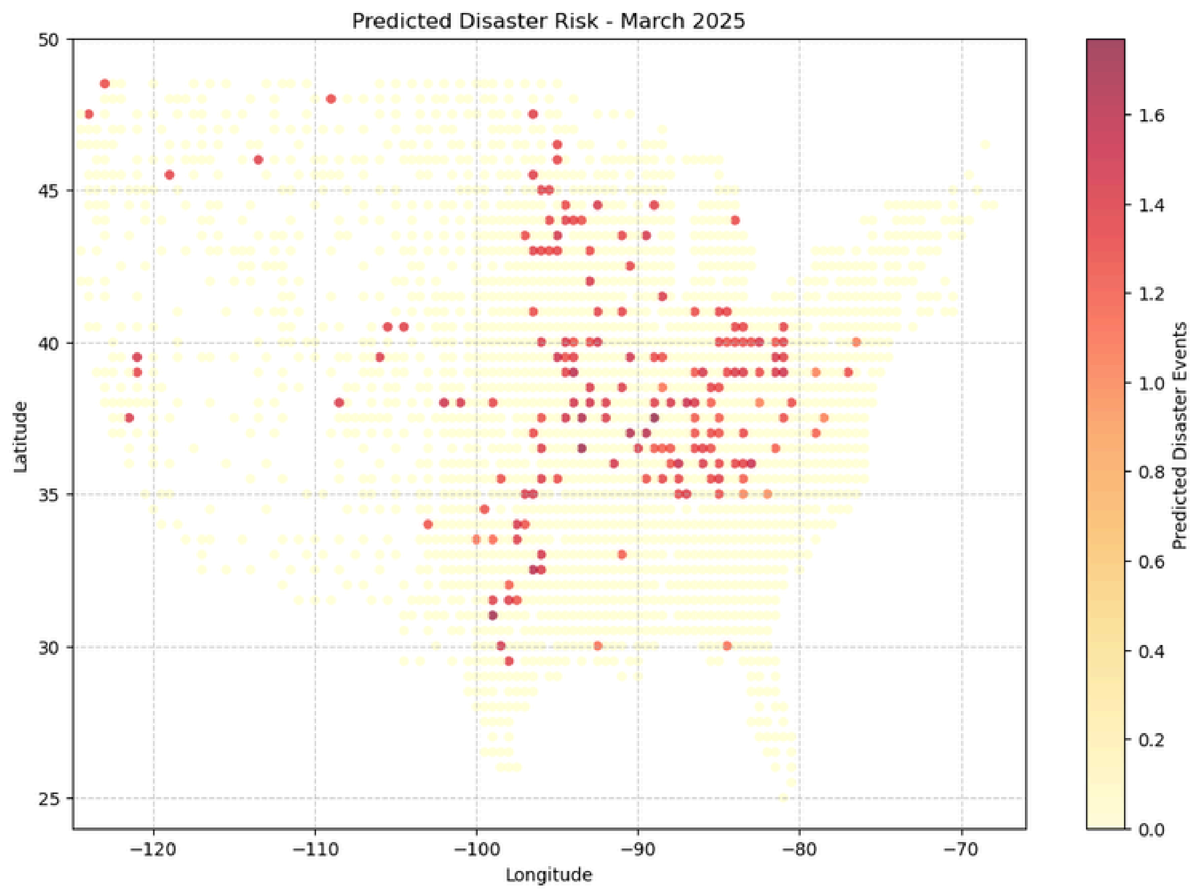
Ground truth data for 1 year period



Predictions for same datapoints from left graph



GAUSSIAN REGRESSION MODEL



Background

- The **Gaussian Regression model** is well-suited for disaster prediction
  - Excels at modeling spatial and temporal correlations while also quantifying prediction uncertainty
- The model defines a prior distribution over functions and updates it with historical data
- Models cyclical seasonality and encode seasonal indicators to learn from past events using 1-12 month historical disaster frequencies across geographical grid cells

TAKEAWAYS

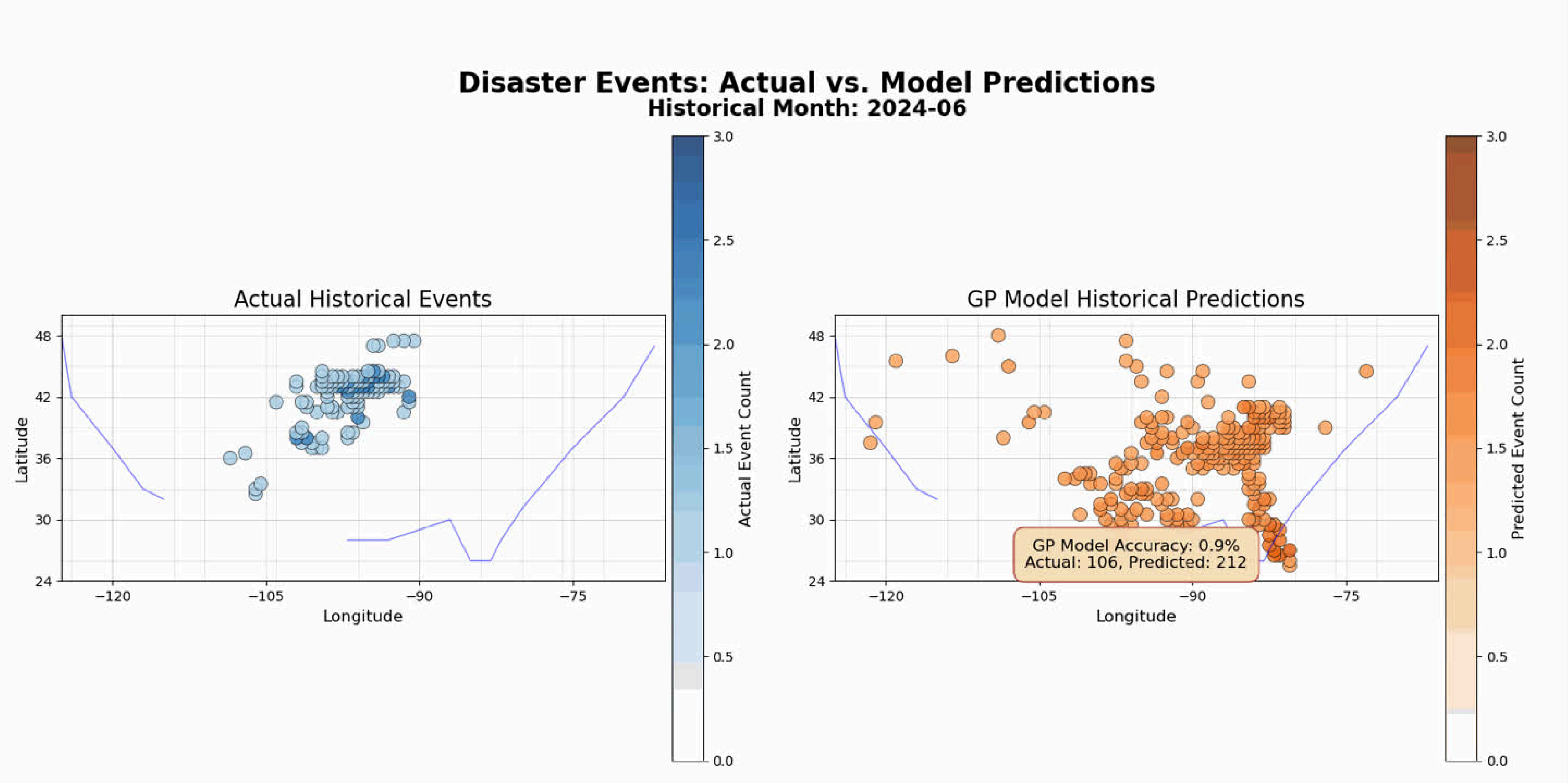
Table 2.1: GPR Performance Metrics

Model	Accuracy	Precision	Recall	F1-Score	ROC-AUC
Gaussian Process Regression	0.795	0.7684	0.7944	0.7763	0.8726

Table 2.2: GPR Model Training Metrics

Model	Train RMSE	Test RMSE	Test R^2 Score	Average Uncertainty (std)
Gaussian Process Regression	0.5509	0.6452	0.4933	0.3156

We also used composite kernel functions to capture localized patterns and broader regional trends in disaster occurrences



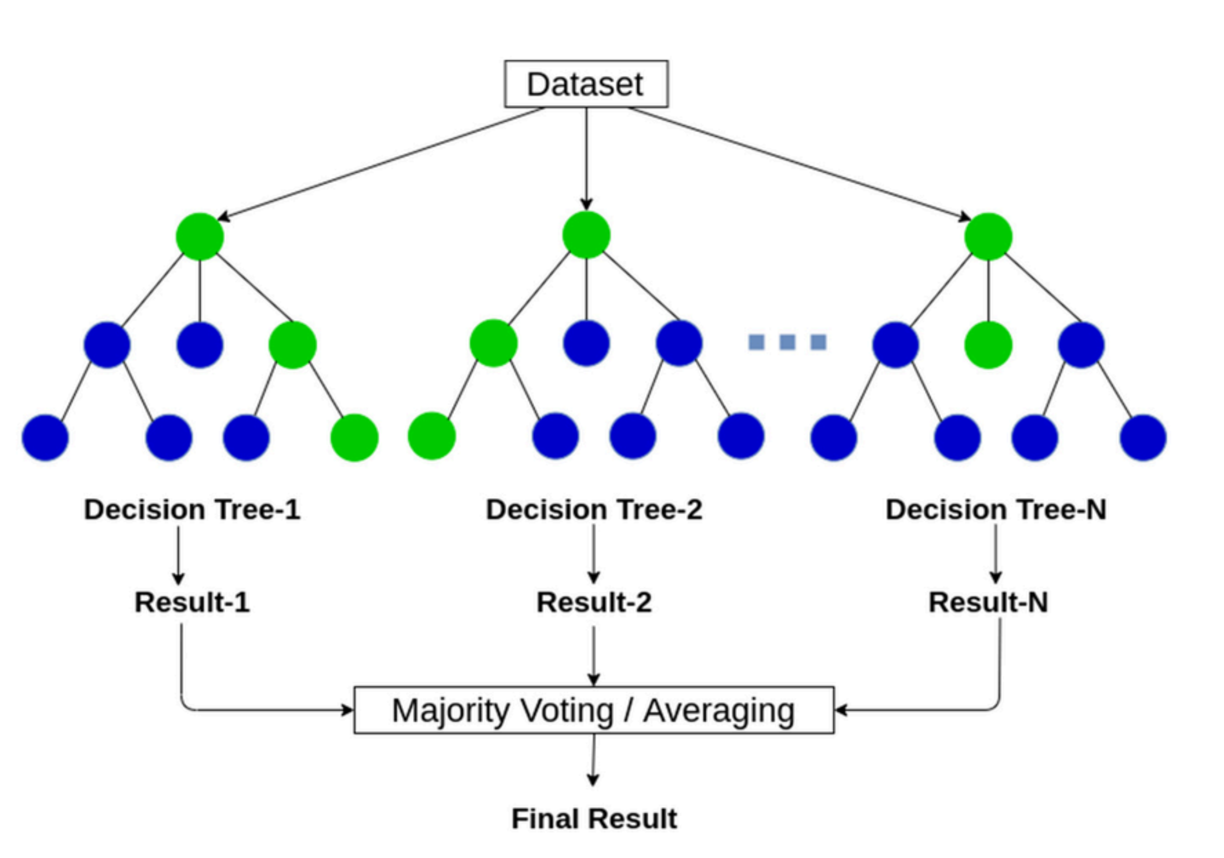
COMPARISON

LSTM Model performed better, having higher accuracy and recall.

- Higher recall is most important as we prefer to minimize false negative classifications

Model	Accuracy	Precision	Recall	F1-Score	AUC-ROC
Deep LSTM (5 layers)	0.838	0.8404	0.8382	0.8377	0.9109
Gaussian Process Regression	0.795	0.7684	0.7944	0.7763	0.8726

RANDOM FOREST CLASSIFICATION



Background

- Decision trees make reliable predictions by considering all various factors and then choose the prediction with the majority vote
- Each tree can be sampled on a random sample of the dataset and each time, only a random variety of features are taken into consideration, handles unbalanced data
- After finding likelihood of disaster, wanted to learn about potential regions and season would be predicted and correlated
- Disasters and this dataset are complex and based on various factors (location, season, previous patterns)

TAKEAWAYS

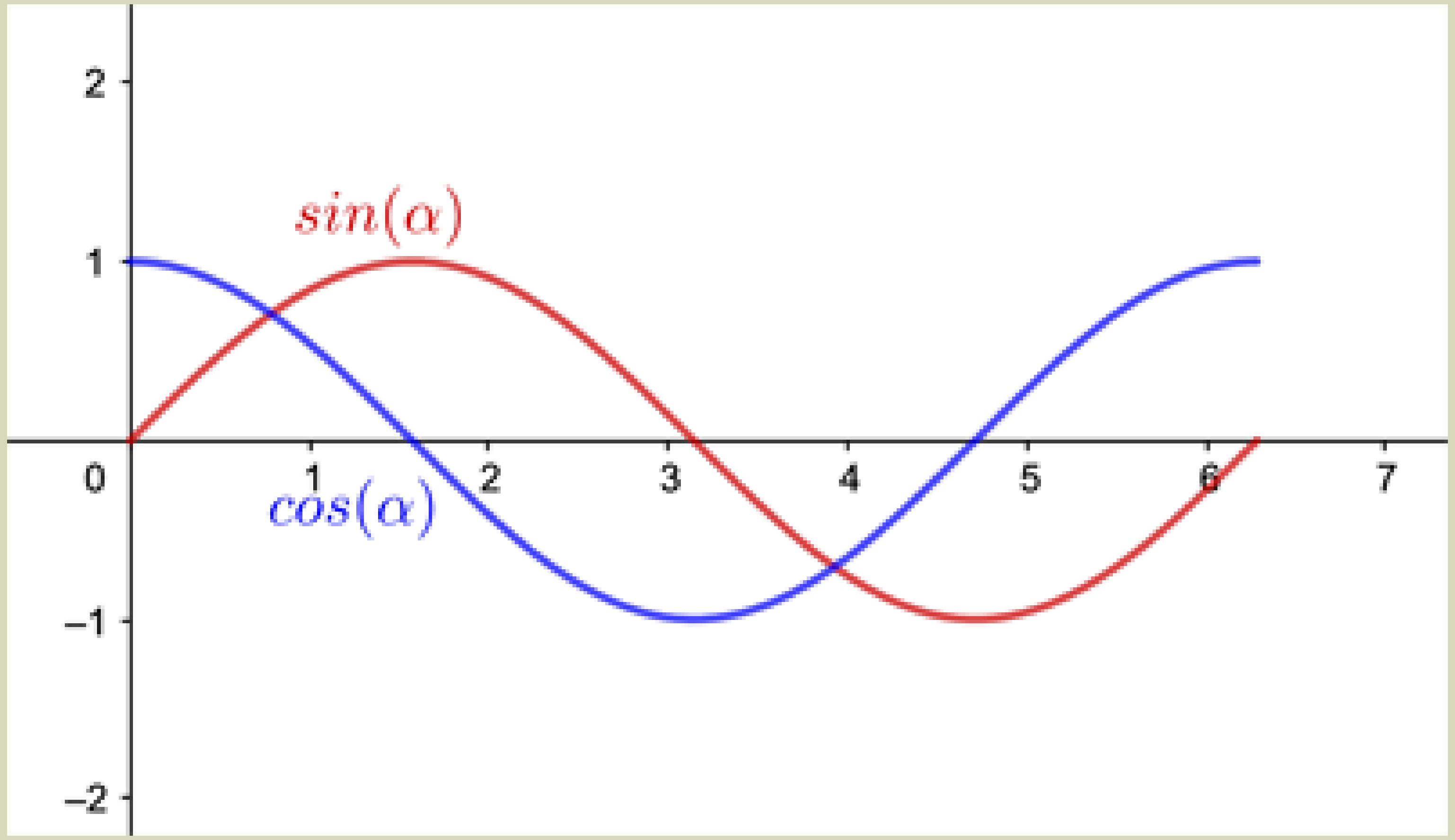
The month is implemented with the sine/cosine functionality since months are cyclical

Stratified K-fold Cross Validation for even distribution of various classes within training and testing.

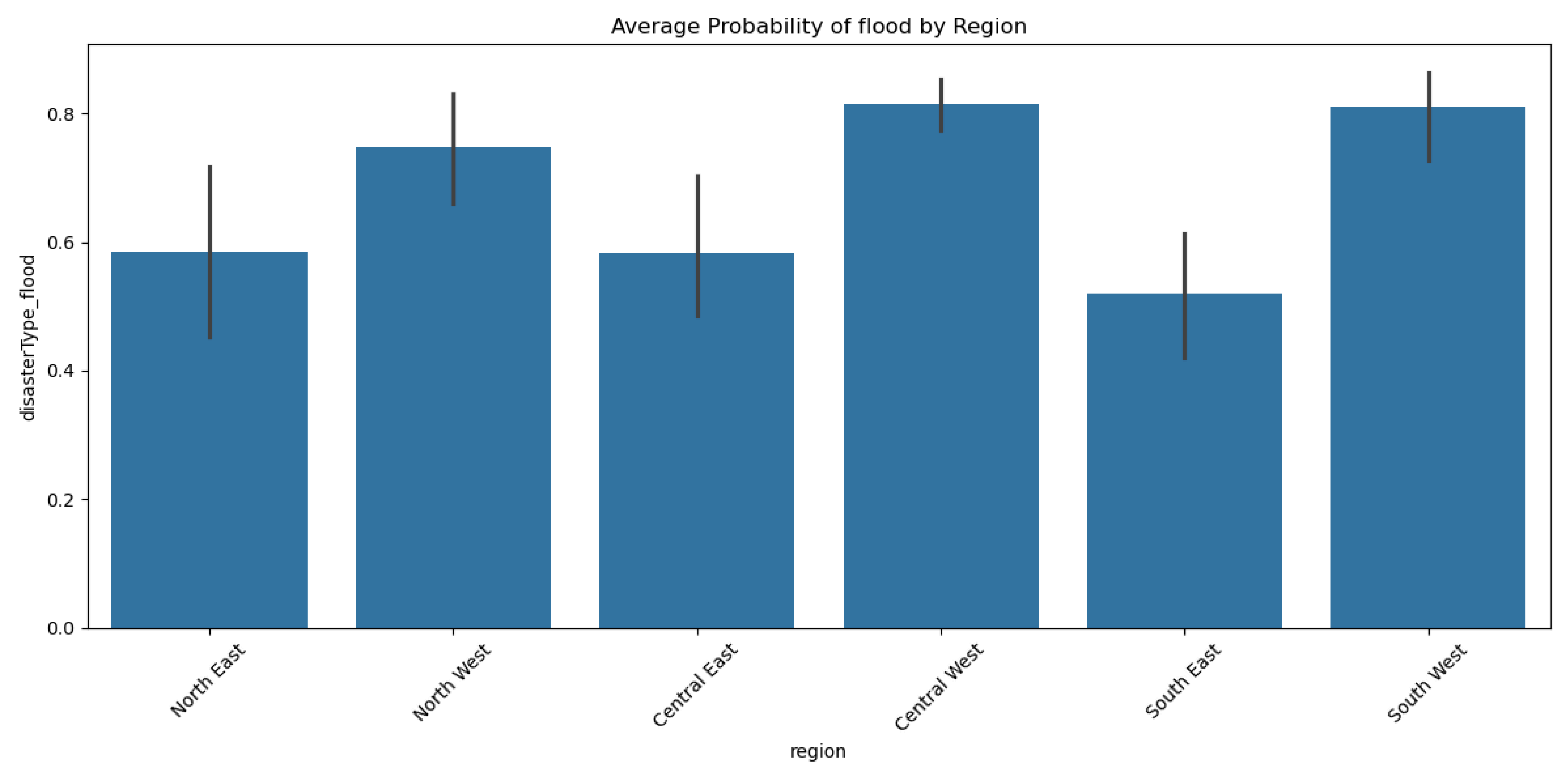
Hyperparameters: how deep the trees can grow, how many trees are built, and how features are selected during training.

Table 2.1 Random Forest Classification Metrics

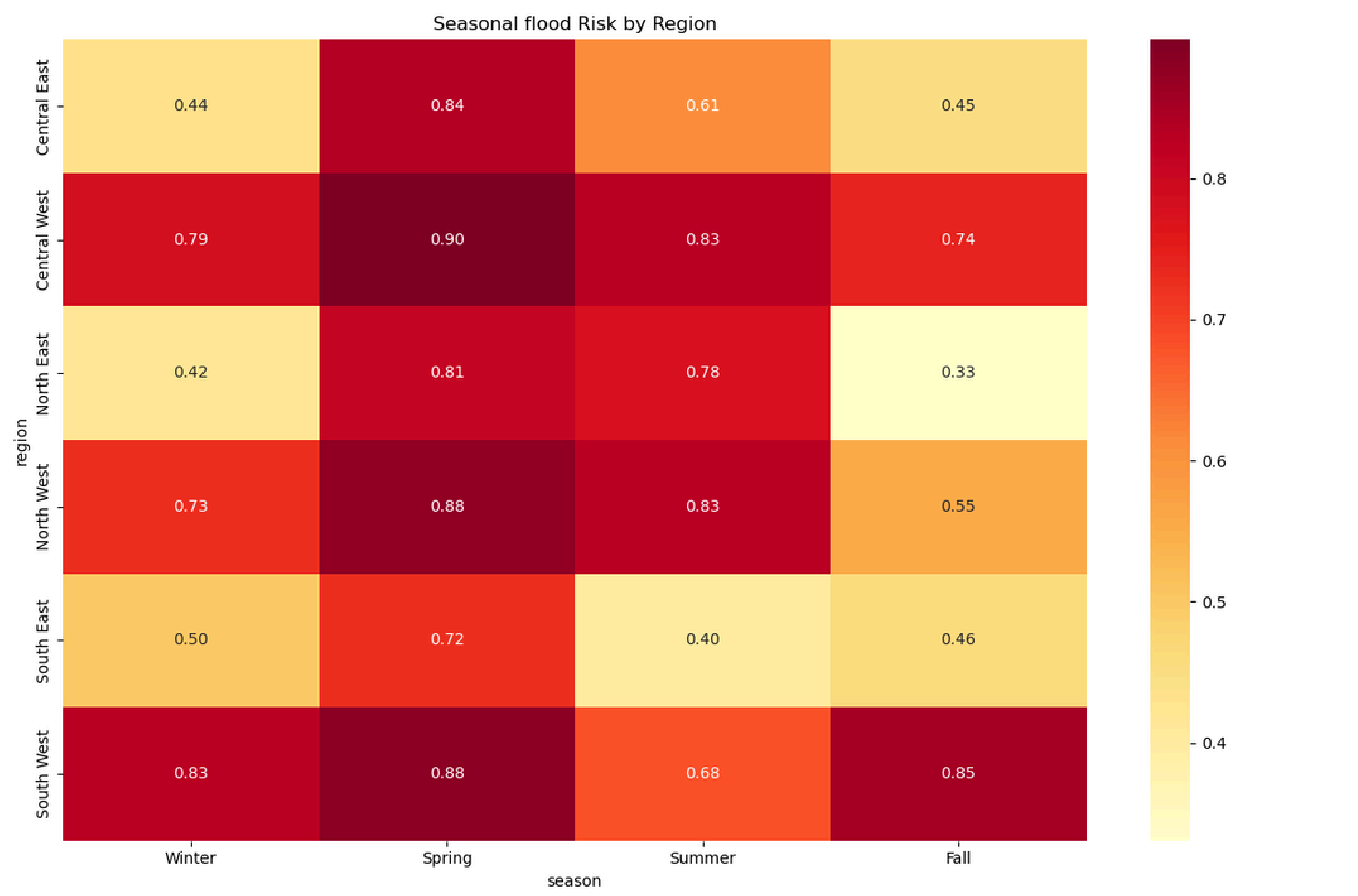
	Storm	Flood	Rain	Ice	Blizzard	Hurricane
Accuracy	0.85	0.87	0.92	0.94	0.96	0.94
Precision	0.92	0.89	0.19	0.24	0.23	0.76
Recall	0.86	0.84	0.32	0.44	0.34	0.99
F1- Score	0.89	0.87	0.24	0.31	0.27	0.86



RESULTS



Average Probability of flood by Region
disasterType_flood
0.0
0.2
0.4
0.6
0.8
North East
North West
Central East
Central West
South East
South West
region

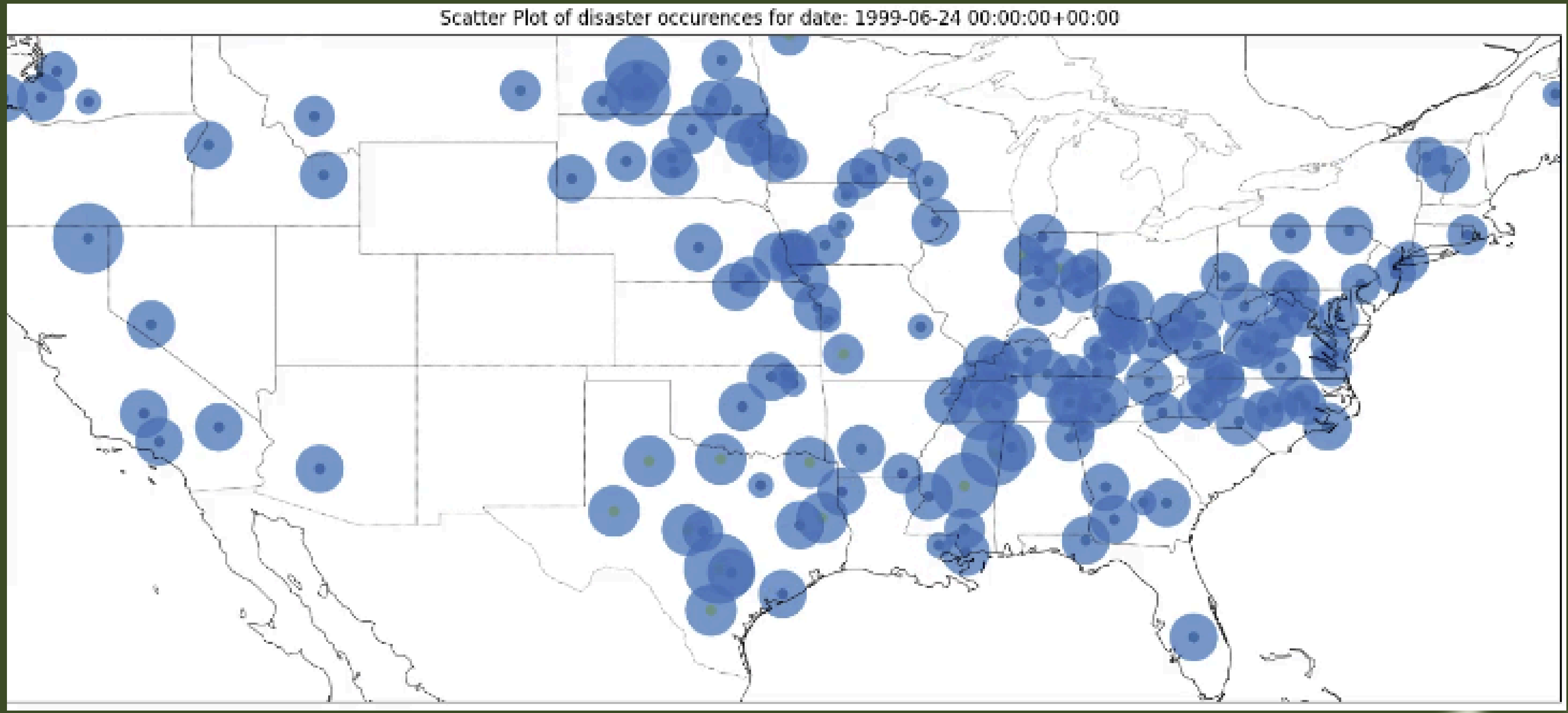


Seasonal flood Risk by Region
region
Central East
Central West
North East
North West
South East
South West
0.44 0.84 0.61 0.45
0.79 0.90 0.83 0.74
0.42 0.81 0.78 0.33
0.73 0.88 0.83 0.55
0.50 0.72 0.40 0.46
0.83 0.88 0.68 0.85
Winter
Spring
Summer
Fall
season
0.4
0.5
0.6
0.7
0.8

FUNDING

- Dataset includes information on types of funding provided for a given disaster event
- Can we predict the types of funding needed for a disaster event?
    - Aid in resource allocation and recovery efforts
- Useful as proxy for severity of an event
    - More funding/variety of funding necessary --> more severe weather event

ihProgramDeclared	IH Program Declared	boolean	Denotes whether the Individuals and Households program was declared for this disaster. For more information on the program, please visit https://www.fema.gov/assistance/individual/program
iaProgramDeclared	IA Program Declared	boolean	Denotes whether the Individual Assistance program was declared for this disaster. For more information on the program, please visit https://www.fema.gov/assistance/individual/program
paProgramDeclared	PA Program Declared	boolean	Denotes whether the Public Assistance program was declared for this disaster. For more information on the program, please visit https://www.fema.gov/assistance/public/program-overview
hmProgramDeclared	HM Program Declared	boolean	Denotes whether the Hazard Mitigation program was declared for this disaster. For more information on the program, please visit https://www.fema.gov/grants/mitigation



Visualization of Disaster Events, radius corresponding to funding provided

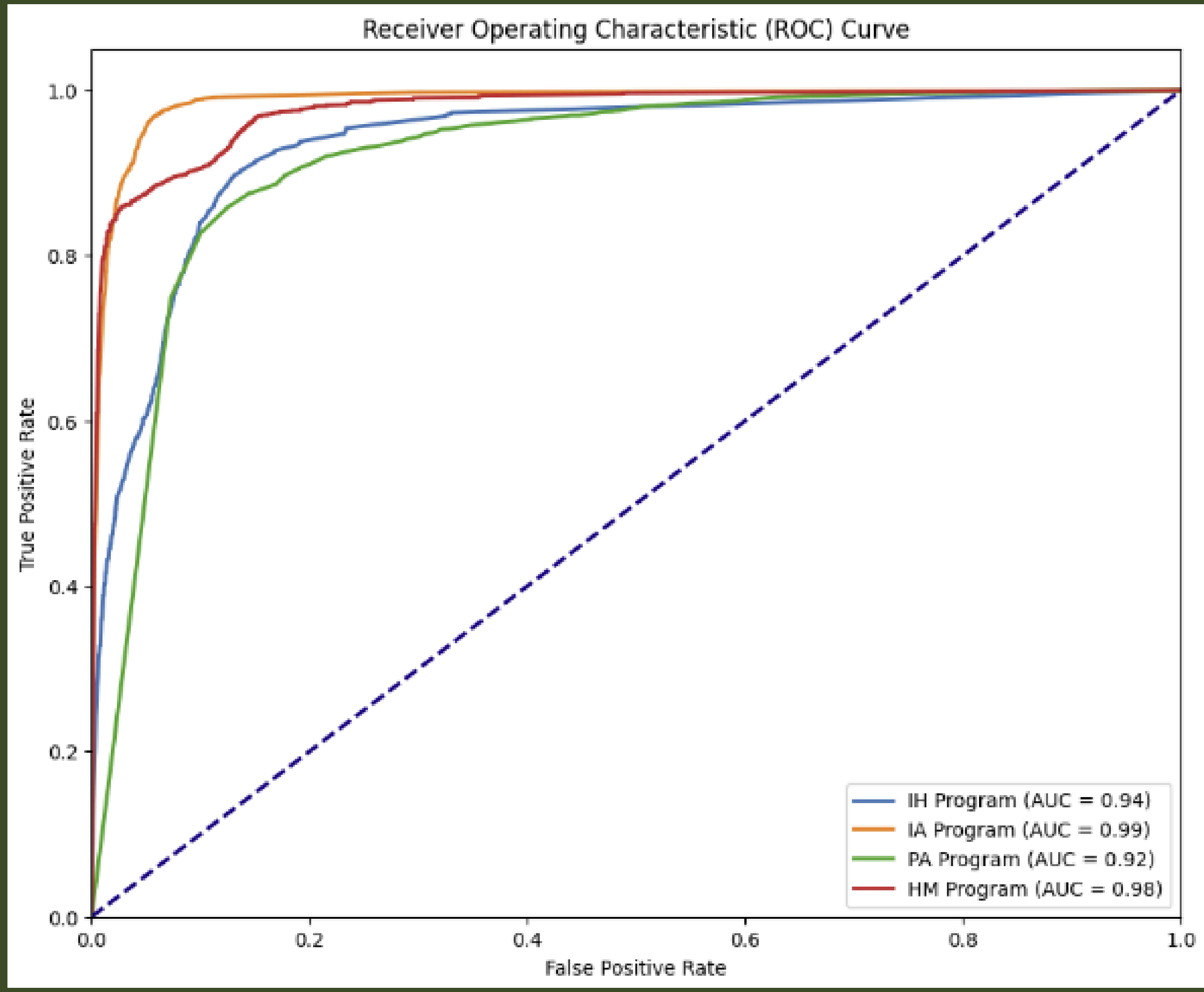
RFC CONT: FUNDING PREDICTION

- Objective: Given information available prior to a disaster event, predict the types of federal funding programs provided in recovery efforts
- Neural network, logistic regression, and Random Forest Classifier
- Best performance achieved with Random Forest Classifier approach
  - Inherently good at dealing with mixed categorical/continuous dataset
  - Effective for imbalanced datasets
- Grid Search hyperparameter tuning --> bootstrap: True, max_depth: None, min_samples_leaf: 1, min_samples_split: 5, n_estimators: 50

Table 4.1: Random Forest Classifier for Funding Prediction Metrics

Programs:	Public Assistance	Individual Assistance	Hazard Mitigation	Individuals & Household
Precision	0.97	0.90	0.90	0.65
Recall	0.95	0.96	0.92	0.75
F1- Score	0.96	0.93	0.91	0.69
Support	8709	2929	4659	1356
Overall Accuracy (Predicting ALL funding programs correctly)				0.76907

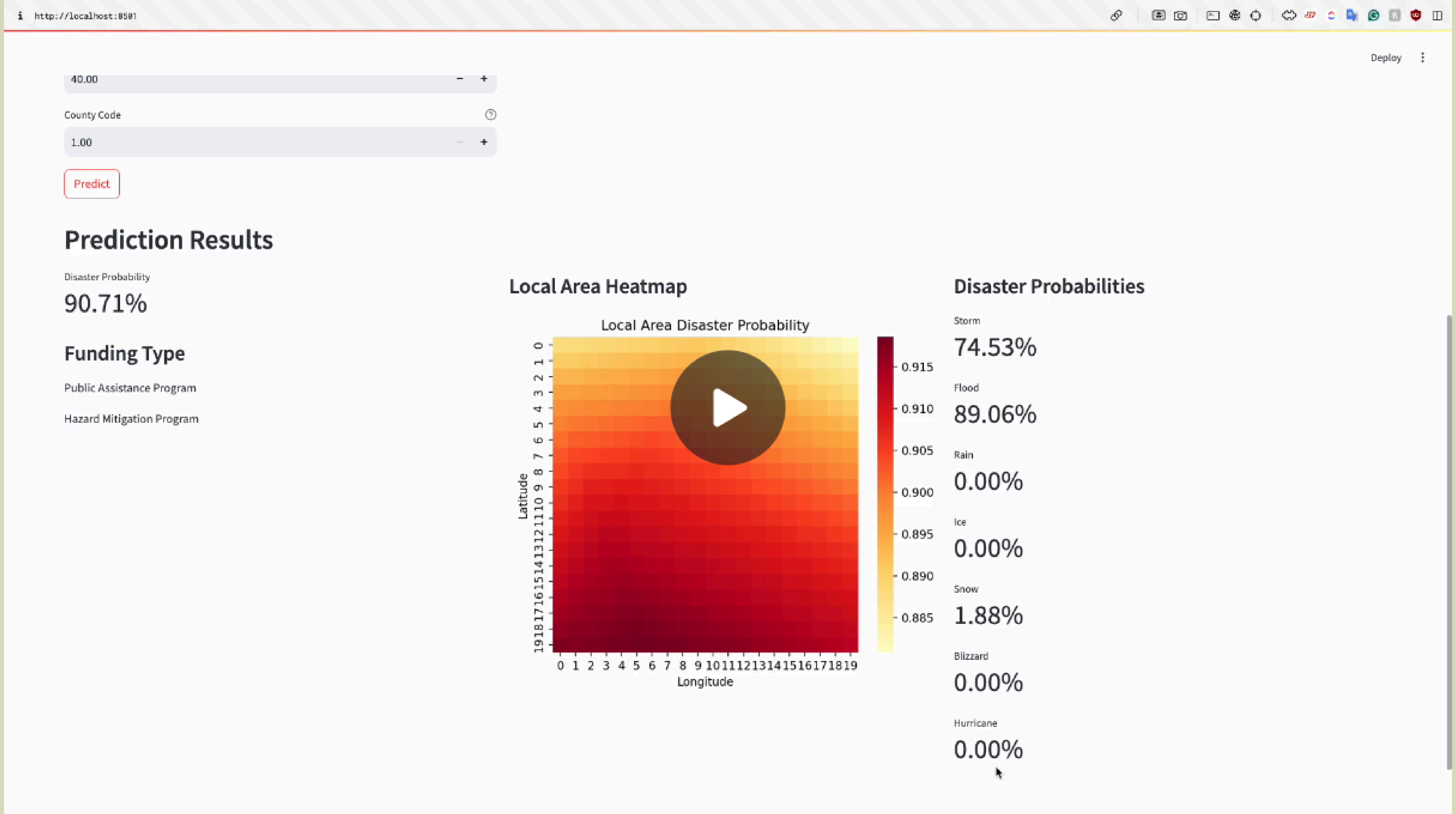
Model Performance Metrics



DISCUSSION AND CHALLENGES

- Spatial-temporal data is challenging to model
  - Could use models with structure to handle time-series data (RNNs, LSTMs) for all metrics
- Data structure is important for prediction on reactive models (RFC, Gaussian)
  - could employ grid search/hyperparameter tuning and data cleaning for bias
- could employ a hybrid model structure to combine two models (reactive and proactive model) to best utilize historical time-series data and short-term forecasting
- Not enough data provided for all disasters
- Some data was not disaster focused but rather evacuation/evacuee so had to do some pre-processing to work with accurate data results

CONCLUSION


http://localhost:8501
Deploy
40.00
County Code
1.00
Predict
Prediction Results
Disaster Probability
90.71%
Funding Type
Public Assistance Program
Hazard Mitigation Program
Local Area Heatmap
Local Area Disaster Probability
Latitude
Longitude
0.915
0.910
0.905
0.900
0.895
0.890
0.885
Disaster Probabilities
Storm
74.53%
Flood
89.06%
Rain
0.00%
Ice
0.00%
Snow
1.88%
Blizzard
0.00%
Hurricane
0.00%

THANK YOU

REFERENCES

Akhyar, Akhyar, et al. “Deep Artificial Intelligence Applications for Natural Disaster Management Systems: A Methodological Review.” *Ecological Indicators*, Elsevier, 6 May 2024, www.sciencedirect.com/science/article/pii/S1470160X24005247.

“Predictive Analytics in Disaster Prevention: Machine Learning Models for Early Warning Systems - IEEE Public Safety Technology Initiative.” IEEE Public Safety Technology, IEEE, https://publicsafety.ieee.org/topics/predictive-analytics-in-disaster-prevention-machine-learning-models-for-early-warning-systems#:~:text=Neural%20networks%2C%20particularly%20deep%20learning,interconnected%20nodes%20organized%20in%20layers. Accessed 6 Mar. 2025.

Yu, Rose. “Neural Point Process for Learning Spatiotemporal Event Dynamics.” *Rose Yu*, 2022, https://proceedings.mlr.press/v168/zhou22a/zhou22a.pdf.