Predicting Hurricane Categories Using Machine Learning

Hubert Hwang
Master of Applied Data Science
University of North Carolina at Chapel Hill
Chapel Hill, NC 27599
hwhwang@unc.edu

Sooji Rhodes
Master of Applied Data Science
University of North Carolina at Chapel Hill
Chapel Hill, NC 27599
rhodesso@unc.edu

1 Introduction

Hurricanes are among the most destructive natural disasters, inflicting widespread damage and loss of life. Accurate prediction of hurricane intensity is vital for disaster preparedness and mitigation. Traditional meteorological models rely on physical simulations, which are computationally expensive and often limited by simplifying assumptions. Machine learning (ML) offers a complementary, data-driven approach to identifying patterns in historical data and improving predictive accuracy.

This project aims to classify hurricanes into intensity categories using a range of ML methods. Specifically, we explore Logistic Regression, Random Forest, Gradient Boosting, Support Vector Machines (SVM), Artificial Neural Networks (ANN), and Long Short-Term Memory (LSTM) networks. By comparing these methods, we identify their strengths, limitations, and potential contributions to improving hurricane forecasting.

1.1 Data Collection Instruments

Hurricanes destructiveness and fast moving nature create unique ways for scientists to collect crucial data in order to forecast the storms category. The collection of data has been done through many devices that are oceanographic or aerial some include:

- **Aircraft:** NOAA deploys the use of planes to collect data that use on board instruments, dropsondes are dropped out from the aircraft that measure atomspheric conditions, and other instruments are launched into the storm.
  - **P3-Orion:** flies directly into the storm.
  - **Gulfstream IV-SP:** flies overhead of the storm.
- **Gliders:** Known as AUV's, autonomous underwater vehicles, which are deployed during the hurricane season (June 1 - Nov 30 in Atlantic basin) and gather data that is more representative of the ocean conditions. This tool collects data by using small changes in buoyancy, with wings for propulsion, to convert vertical motion into horizontal motion. These gliders will dive up to 1,000 meters in order to collect salinity and temperature data.
- **Dropsondes:** Oldest hurricane observation instrument that has been used since 1996. These tools are launched from the air crafts to gather data inside the storm. Once dropped, a parachute is deployed to slow its descent in order to capture data through sensors. Information such as temperature, relative humidity, pressure, wind speed, wind direction, and dew points.
- **Saildrones:** This tool is used to represent the intersection of oceanic and atmospheric observations. This tool takes measurements below and above the water. Scientists are then provided with data about the boundary layer where the ocean meets the atmosphere.

Foundation Models for Science Workshop,38th Conference on Neural Information Processing Systems (NeurIPS 2024).

Typically, storms contain water and wind that are too strong to enter and therefore this device is deployed to capture wave action, air temperature, wind speed, temperature and salinity below the ocean's surface.

- **Uncrewed Aircraft System (UAS):** Also known as drones, have assisted scientists with the ability to study the boundary layer where the atmosphere ineracts with the ocean's surface. These devices are used to collect humidity, wind speed, temperature, and atmospheric pressure in the harshest conditions. These devices are especially useful in capturing data in the boundary layer which improves forecasting intensity and behavior.
- **Airborne Expendable BathyThermograph (AXBT):** This instrument is launched from the hurricane P3 aircraft and does not capture anything until it hits the surface of the water. Once this hits the water, the ocean temperature is measured as a function of depth. This gives an idea to scientist on how the storms passing overhead mixes, cools, or heats the ocean. This is very important to determine intensity because the thermal structure of the ocean contributes heavily to a storms development and strength.
- **Argo Float:** This device moves up and down between the surface and mid-water level as it drifts with ocean currents. These are placed year-round globally to measure temperature and salinity in the upper 2,000 meters. The Argo float reaches deeper than gliders and reveal the depth at which the atmosphere and ocean interact.
- **Drifter:** These are drifting surface buoys that are deployed and tracked by satellites as they float in the ocean. The buoys capture a measure of mixed layer currents, atmospheric pressure, waves , winds, sea surface temperature and salinity. These are placed year round and have the ability to be pulled into developing storms which can provide very useful information about the conditions of the ocean inside a hurricane.

2 Related Work

2.1 Literature Review

Research on hurricane prediction has shown the critical impact of environmental factors such as sea surface temperature (SST) and atmospheric pressure:

- **Global warming and hurricane intensity:** Studies like *Recent Increases in Tropical Cyclone Intensification Rates* (1) and *Attribution of 2020 Hurricane Season Extreme Rainfall to Human-Induced Climate Change* (2) highlight how global warming increases the frequency and intensity of hurricanes.
- **Machine learning in hurricane forecasting:** *HurriCast* (3) demonstrated the potential of ML models, particularly ensemble methods and sequential networks, in predicting hurricane behavior.

2.2 Contribution of This Study

This study extends prior research by comparing diverse ML models on a standardized dataset. We emphasize the impact of feature engineering and hyper parameter optimization while addressing challenges like imbalanced classes and large-scale datasets. Our findings provide actionable insights for practitioners and researchers seeking to integrate ML into hurricane forecasting.

3 Methods

3.1 Problem Formulation

Hurricane prediction is framed as a supervised multi class classification problem. The target variable, hurricane category, follows the Saffir-Simpson scale ($y \in \{0, 1, 2, 3, 4, 5\}$). The goal is to learn a function:

$$f(X) \to y$$

where X represents the input features, including atmospheric pressure (*pressure*), sea surface temperature (*sst*), and rolling averages.

3.2 Dataset

We utilized the NOAA HURDAT2 dataset (4) and supplementary data from Kaggle (5). Key features include:

- **Wind speed** (*wind*)
- **Atmospheric pressure** (*pressure*)
- **Sea surface temperature** (*sst*)
- **Rolling averages** (*sst_rolling_mean*, *pressure_rolling_mean*)

3.3 Data Challenges

The vast dataset (1980–2020) posed significant challenges. Missing values in SST and pressure were addressed using KNN imputation, and rolling averages were computed to capture temporal trends. Due to the dataset's size, data processing was conducted in 5-year increments, a compromise necessitated by project time constraints.

3.4 Machine Learning Models

3.4.1 Logistic Regression

$$P(y = k|X) = \frac{e^{\beta_k^T X}}{\sum_{j=1}^{K} e^{\beta_j^T X}}$$

Hyperparameters: Multinomial solver (*lbfgs*), max iterations = 1000.

3.4.2 Random Forest

$$\hat{y} = \text{mode}(\hat{y}_1, \hat{y}_2, \ldots, \hat{y}_B)$$

Hyperparameters: 200 trees, minimum samples split = 5.

3.4.3 Gradient Boosting

$$f_m(x) = f_{m-1}(x) + \eta \cdot g_m(x)$$

Hyperparameters: 100 estimators, learning rate = 0.1.

3.4.4 Support Vector Machine (SVM)

$$\min \frac{1}{2}\|\mathbf{w}\|^2 \quad \text{subject to } y_i(\mathbf{w} \cdot X_i + b) \geq 1$$

Hyperparameters: $C = 10$, kernel = RBF.

3.4.5 Artificial Neural Network (ANN)

$$z^{[l]} = W^{[l]} a^{[l-1]} + b^{[l]}, \quad a^{[l]} = \sigma(z^{[l]})$$

Hyperparameters: 3 layers (128, 64, 32 units), Dropout = 0.4, learning rate = 0.0005.

3.4.6 Long Short-Term Memory (LSTM)

$$h_t = f(W_h \cdot x_t + U_h \cdot h_{t-1} + b_h)$$

Hyperparameters: 50 units, batch size = 16.

4 Experiments and Results

4.1 Performance Comparison of Models

The performance of six machine learning models—Logistic Regression, Random Forest, Gradient Boosting, Support Vector Machines (SVM), Artificial Neural Networks (ANN), and Long Short-Term Memory (LSTM) networks—was evaluated on the hurricane category prediction task. Each model was trained and tested using features such as atmospheric pressure, sea surface temperature (SST), and their rolling averages.

Table 1 summarizes the accuracy of each model, while Figure 2 illustrates the confusion matrix of the Random Forest classifier, which was the best-performing model.

Table 1: Model Performance Comparison

Model	Accuracy (%)
Logistic Regression	89.00
Random Forest	96.12
Gradient Boosting	94.89
SVM	91.00
ANN	90.65
LSTM	87.34

4.2 Key Observations from Model Performance

- **Logistic Regression:** As expected, Logistic Regression provided a strong baseline with an accuracy of 89%. However, its linear nature limited its ability to capture the nonlinear patterns in the data, particularly in underrepresented hurricane categories.
- **Random Forest:** This model achieved the highest accuracy of 96.12%, benefiting from its ability to handle noisy data and imbalances in class distribution. Its feature importance plot (Figure 1) shows that pressure and SST were the most critical features, aligning with domain knowledge.
- **Gradient Boosting:** With an accuracy of 94.89%, Gradient Boosting demonstrated its ability to focus on hard-to-classify instances. Its sequential learning process allowed it to slightly outperform Random Forest in specific scenarios involving rare classes.
- **SVM:** The SVM achieved a solid accuracy of 91%, showing strength in capturing decision boundaries. However, it was computationally expensive and required extensive hyperparameter tuning to optimize performance.
- **ANN:** While ANN achieved a competitive accuracy of 90.65%, it struggled with imbalanced classes and required significant computational resources. Its ability to model nonlinear relationships, however, underscores its potential with larger datasets.
- **LSTM:** Designed for sequential data, LSTM networks achieved 87.34% accuracy. The model effectively captured temporal dependencies but suffered from overfitting due to the limited dataset size.

4.3 Feature Importance Analysis

Feature importance analysis from the Random Forest model, visualized in Figure 1, confirmed that atmospheric pressure and SST were the most influential predictors of hurricane intensity. Rolling averages of these features also contributed valuable insights into temporal trends, but their influence was secondary to the raw pressure and SST values.

4.4 Confusion Matrix Analysis

To further assess model performance, the confusion matrix of the Random Forest classifier is shown in Figure 2. This visualization highlights the model's ability to correctly predict the majority of hurricane categories, with minor misclassifications in higher categories due to data imbalance.

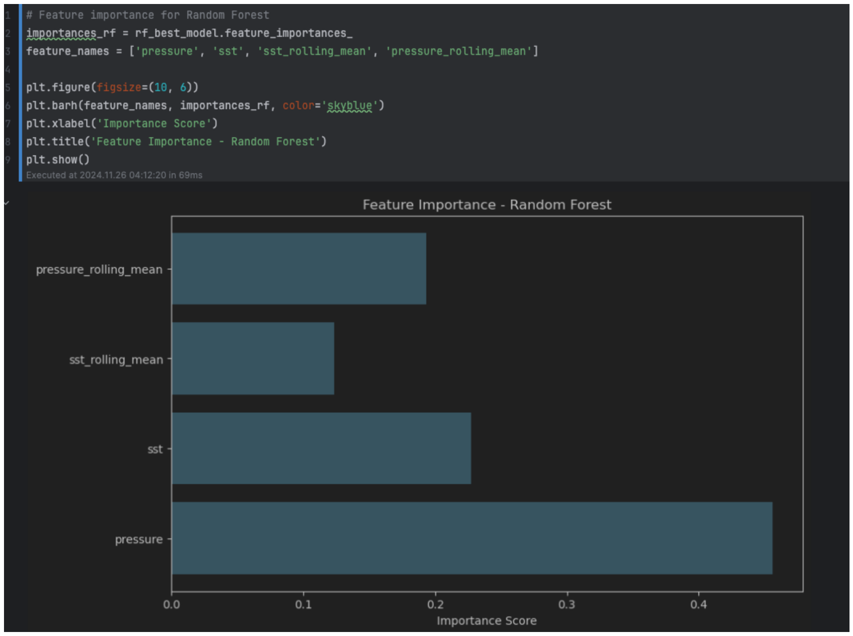

Figure 1: Feature Importance Plot from the Random Forest Model

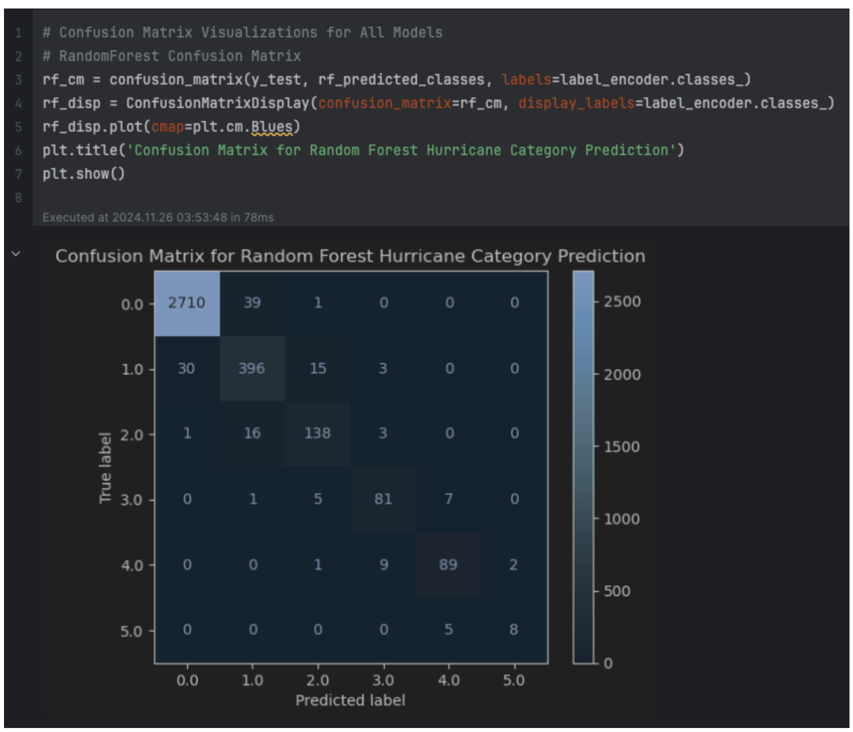

Figure 2: Confusion Matrix for the Random Forest Model

4.5 Insights from Performance Metrics

- **Robustness of Ensemble Models:** The Random Forest and Gradient Boosting models consistently outperformed others due to their ability to aggregate predictions and handle noisy data effectively.

- **Challenges with Neural Networks:** ANN and LSTM models faced challenges due to the relatively small dataset and the imbalanced nature of the target variable. These models would benefit from larger datasets and advanced preprocessing techniques.
- **Imbalanced Classes:** Models struggled to accurately predict higher hurricane categories (e.g., Categories 4 and 5), highlighting the need for techniques like SMOTE to improve class balance.

5 Conclusion

5.1 Accomplishments

This study explored the classification of hurricane categories using six machine learning models: Logistic Regression, Random Forest, Gradient Boosting, Support Vector Machines, Artificial Neural Networks, and Long Short-Term Memory networks. Each model was rigorously evaluated on its ability to handle the nonlinear and temporal complexities inherent in meteorological data. Among these, Random Forest emerged as the best-performing model, achieving an accuracy of 96.12%, closely followed by Gradient Boosting at 94.89%. These results demonstrate the potential of ensemble learning methods in capturing intricate relationships in environmental variables.

Additionally, we demonstrated the effectiveness of feature engineering techniques, such as rolling averages and imputing missing values with KNN, to improve model performance. Our study confirmed the critical role of atmospheric pressure and sea surface temperature (SST) as the most significant predictors of hurricane intensity, with rolling averages providing supplementary insights into temporal trends.

5.2 Lessons Learned

- **Feature Importance:** Atmospheric pressure and SST were consistently identified as the most influential features across all models. This finding aligns with meteorological studies, validating the reliability of our feature engineering process.
- **Model Strengths:**
  - **Ensemble Methods:** Random Forest and Gradient Boosting showed remarkable robustness to noisy data and imbalances in class distribution. Their ability to focus on difficult-to-classify instances contributed to their superior performance.
  - **Neural Networks:** While ANNs and LSTMs struggled with smaller datasets, they excelled in modeling complex, nonlinear relationships. LSTMs, in particular, highlighted the potential of temporal modeling in hurricane prediction.
  - **Baseline Models:** Logistic Regression provided a valuable benchmark, illustrating the limitations of linear assumptions in a highly nonlinear problem space.
- **Data Challenges:** The vastness of the dataset presented significant challenges. While we focused on data from 1980 to 2020, it was impossible to import and manipulate all the data efficiently within the constraints of our project. The sheer size of the dataset required considerable preprocessing, which limited our ability to leverage even larger datasets or more granular temporal analyses. For instance, breaking the data into 5-year increments could have allowed deeper exploration, but such an approach would have required at least three months to execute. This limitation underscores the importance of scalable data processing techniques and extended project timelines for handling massive datasets.

5.3 Key Insights for the Reader

This study highlights the complementary nature of different machine learning models for hurricane intensity prediction. Readers should recognize the importance of:

- Using ensemble methods for robust performance in noisy and imbalanced datasets.
- Incorporating sequential models like LSTMs to capture temporal dependencies in weather data.
- Employing comprehensive feature engineering to enhance model interoperability and performance.

These insights contribute to the broader field of hurricane prediction by illustrating how machine learning can complement traditional meteorological models, offering more flexible and data-driven approaches.

5.4 How This Work Extends the Field

Our study bridges the gap between physical models and machine learning techniques by providing a comparative analysis of diverse methods on a real-world dataset. By demonstrating the strengths and limitations of each approach, we offer a roadmap for practitioners to select and optimize models based on their specific requirements. Furthermore, this research underscores the potential of combining feature engineering, temporal modeling, and ensemble learning to improve hurricane prediction accuracy.

5.5 Future Directions

- **Addressing Data Imbalances:** Future work should explore techniques like Synthetic Minority Oversampling Technique (SMOTE) or data augmentation to mitigate class imbalances, especially for higher hurricane categories.
- **Incorporating Additional Features:** Features like ocean heat content, wind shear, and satellite-derived cloud patterns could enhance model predictions.
  - **Ocean Heat Content(OHC):** The OHC measure thermal energy in the upper ocean layers which influence hurricane intensity. Surface temperature alone might not tell the full story, therefore OHC accounts for deeper heat reservoirs that could sustain or intensify storms. High OHC provides a greater source of energy, which can lead to rapid intensification.
  - **Mid-Level Moisture:** This is moisture in the mid-troposphere (approximately 3-7 km altitude). Dry air can disrupt convection as it infiltrates storms and weakens their system. Tracking this can give better prediction of how storms develop and decay.
  - **Upper-Level Divergence:** This measure the spread of air outward at high altitudes. This divergence enhances upward motion in the storm which advances further organization of the storm and allows it to strengthen. This is particularly paramount to the prediction of rapid intensification in a sheared environment.
  - **Ocean currents:** Ocean current can influence the movement of warm water masses and redistribute heat. Warm water under a storm can be quickly replenished through strong currents thus sustaining its intensity. The weak currents could possibly lead to faster cooling of surface water. Incorporating current data can enhance forecasting by showing heat availability evolving along a storms path which assist in predicting when a storm is weakening or intensifying over its course.
  - **Sea Surface Salinity (SSS):** Sea surface salinity will affect the oceans circulation and density, this will influence vertical mixing and heat distribution. Upper ocean regions that contain low salinity can stabilize and preserve heat in the surface layers that hurricanes draw upon. Energy availability is reduced in high salinity areas as it may allow cooler water to mix upward. Including sea surface salinity into models will assist in forecast areas that a storm might intensify. For example, low salinity regions may support intensifying of storms as the surface waters remain warm.
- **International Feature Consolidation:** The International Best Track Archive for Climate Stewardship (IBTrACS) is an outstanding initiative that consolidates tropical storm data from multiple international agencies, which creates a comprehensive global dataset. The collaboration among agencies ensures that consistent features are collected and facilitates a unified approach to storm analysis. Most of the features captured internationally are similar to the ones captured by NOAA.
  - **International:**
    - Sea Surface Temperatures (SST)
    - Humidity Levels (Upper and Mid-Level)
    - Ocean Heat Content (OHC)
    - Wind Shear
    - Atmospheric Pressure Patterns

  - **Unique from NOAA:**
    * Ocean Currents
    * Sea Surface Salinity (SSS)
    * Upper-Level Divergence
- **Model Enhancements:**
  - Training LSTMs on larger datasets could reduce over-fitting and improve temporal modeling.
  - Combining Random Forest, Gradient Boosting, and neural networks could leverage their complementary strengths for improved accuracy and generalizability.
- **Overcoming Data Constraints:** To better analyze the massive datasets, future efforts should focus on breaking down the data into smaller temporal chunks or leveraging distributed computing frameworks. This approach would allow a more thorough analysis without overwhelming computational resources.
- **Real-Time Applications:** Integrating these models into real-time forecasting systems could provide rapid predictions, aiding disaster management and preparedness.

5.6 What We Learned

From this project, we learned that machine learning can offer significant improvements in hurricane prediction, but only when accompanied by rigorous preprocessing, effective feature engineering, and careful model selection. The study also highlighted the challenges of working with massive datasets and the need for scalable solutions in such scenarios. Most importantly, this research underscores the value of ensemble methods and sequential modeling in capturing complex patterns in environmental data.

Readers should leave with an understanding of the potential and challenges of applying machine learning to hurricane prediction. This includes recognizing the importance of feature selection, the strengths of various algorithms, and the computational hurdles posed by large-scale data. By providing a framework for future exploration, this study contributes meaningfully to the field of hurricane forecasting.

In conclusion, this study demonstrates the trans-formative potential of machine learning in hurricane intensity prediction. By leveraging diverse methodologies and carefully engineered features, we pave the way for more accurate and reliable hurricane forecasting tools. Future work should build on these findings to address current limitations and unlock further advancements in this critical area.

A Appendix

Refer to the supplementary document: **Final_Notebook_Updated.pdf**.

References

[1] Bhatia, K. T., et al. (2019). Recent increases in tropical cyclone intensification rates. *Nature Communications*.

[2] Attribution of 2020 hurricane season extreme rainfall to human-induced climate change. *Nature Communications*.

[3] HurriCast: An Automatic Framework Using Machine Learning for Hurricane Forecasting. *arXiv*.

[4] NOAA HURDAT2 Dataset. `https://www.nhc.noaa.gov/data/hurdat/hurdat2-1851-2023-051124.txt`.

[5] Kaggle Atlantic Hurricane Database. `https://www.kaggle.com/datasets/utkarshx27/noaa-atlantic-hurricane-database`.

[6] National Oceanic and Atmospheric Administration (NOAA). Observational Instruments. `https://www.aoml.noaa.gov/observational-instruments/`.

[7] Water and energy budgets of hurricanes: Case studies of Ivan and Katrina. Journal of Geophysical Research: Atmospheres, 112(D23). https://agupubs.onlinelibrary.wiley.com/doi/full/10.1029/2006JD008303

[8] Wang, Y., Wu, C.-C. (2004). Current understanding of tropical cyclone structure and intensity changes–A review. Meteorology and Atmospheric Physics, 87(4), 257-278. https://link.springer.com/article/10.1007/s00703-003-0055-6

[9] Hanley, D., Molinari, J., Keyser, D. (2001). A composite study of the interactions between tropical cyclones and upper-tropospheric troughs. Monthly Weather Review, 129(10), 2570-2584. https://journals.ametsoc.org/view/journals/mwre/129/10/1520-0493_2001_129_2570_acsoti_2.0.co_2.xml

[10] Balaguru, K., Taraphdar, S., Leung, L. R., Foltz, G. R. (2012). Increase in the intensity of postmonsoon Bay of Bengal tropical cyclones. Geophysical Research Letters, 39(14). https://agupubs.onlinelibrary.wiley.com/doi/full/10.1029/2012GL052371