Assessing the feasibility of CNN-LSTM Neural Networks versus Decision Trees for climate change-related tropical storm predictions.

Julien Y-Son Kuhn de Chizelle[1,*]

[1] Northwestern University, MSDS Department, Evanston, Illinois

*Address to which correspondence should be addressed:
jkuhndechizelle@gmail.com

Abstract

Understanding the impact of climate change on hurricane frequency and intensity is paramount for governments to provide adequate disaster relief monetary funds and contingency plans for population displacement. Katrina, a 2005 hurricane of category 3, caused just under a thousand accounted casualties and skyrocketing government costs, prompting continuing research for better planning solutions. As no public dataset is available, this report compiles a preliminary template for future use. The test models look for performance differences in parameter inputs as tabular datasets or their equivalent image representations based on Euclidean distances. Tested models output storm intensity and hurricane storm frequency predictions. Results from various Neural Network architectures (including LSTM, CNN, and PINN) and Decision Trees (XGBoost and Random Forest) show that a combination of Neural Networks and Decision Trees ensemble model provides the best compromise for accuracy based on dataset quality and study time available. Finally, IPCC climate change predictions for the year 2100 are extrapolated on an unseen dataset and tested for changes in hurricane frequency and intensity.

Keywords: Hurricane Frequency, Greenhouse Emissions, Convolutional Neural Network (CNN), Long Short-Term Memory Networks (LSTM), XGBoost, Random Forest, Physics-Informed Neural Network (PINN).

Table of Content

I. Introduction

The Intergovernmental Panel on Climate Change (IPCC) consists of a scientific community tasked by the United Nations with assessing and explaining what is known about climate change and the role humans play in it (The Nature Conservancy 2023). Understanding the effect of greenhouse gasses on the likelihood of floods is essential for governments' long-term disaster preparation. The cost of inadequate hurricane preparation goes far beyond the high monetary recovery cost (Richard 2020). Hurricane Katrina in 2005, a category 3, caused 971 fatalities during the storm and 15 additional fatalities during evacuation (CNN 2023). Therefore, preparing ourselves for hurricanes to protect the affected population is paramount.

The IPCC report claims more frequent and severe storms and other extreme weather events as we keep inputting greenhouse gasses into the atmosphere. Researchers such as (Winkler et al. 2022) found, using paleo-reconstruction from sediment data, that "recurrent intervals based on the 170-year instrumental record can severely underestimate the threat hurricanes pose certain localities". In other words, climate change could significantly impact hurricane frequency and intensity. Looking at hurricane locations from the HURDAT dataset for the Atlantic Coast of the United States over the past 170 years, (Vecchi et al. 2021) found "spurious increasing trends in recorded basin-wide hurricanes." Furthermore, the authors found, using Poisson and Binomial regression models, that local climate variabilities and aerosols masked hurricane frequency increases from the mid to late 20th century, skewing the dataset to favor an increase in hurricane frequency in the last decade. Future research needs to account for this flaw in data collection for hurricane frequency.

This report aims to build an artificial intelligence model capable of predicting whether specified meteorological and greenhouse gas concentrations can be used to predict storm occurrence and intensity. Extrapolating this model to IPCC greenhouse and temperature predictions for the year 2100 gives us an understanding of hurricane and storm frequencies as climate changes.

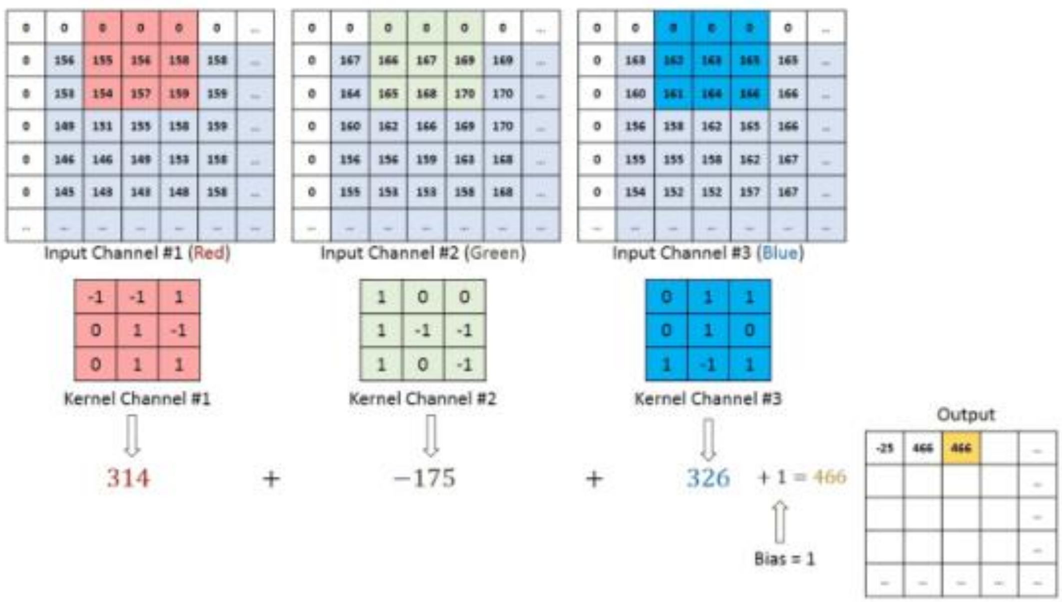
II. Literature Review

Predicting storm frequencies is of great interest to several researchers. (Boussioux et al. 2022) focused on combining a deep-learning encoder/decoder pair alongside gradient-boosted trees to predict the intensity and geographical coordinate forecasting track for North Atlantic and Eastern Pacific hurricanes. The authors used "historical storm data, reanalysis maps, and operational forecast data" from the 1980s onward and input them into a combination of a Convolutional Neural Network (CNN) encoder (LeCun et al. 1989) with a choice of either a Gated Recurrent Units (GRU) (Chung et al. 2014) decoder or a Transformers decoder (Vaswani et al. 2017). Other researchers (Bose et al. 2021) have used an LSTM-RNN model and described them as the best model for short-term storm trajectory forecasting for at least 12 hours.

Other researchers looked at the frequency of hurricanes related to climate change using meteorological equations and predicting risks to coastlines (Marsooli et al. 2019). While the model explained in this paper does not delve into these equations, a future study using a Physics-Informed Neural Network (PINN) architecture could use them (Cai et al. 2021). However, while many researchers targeted hurricane paths (Mudd et al. 2014; Cloud et al. 2019), the author is not aware of research looking at predicting the number of tropical storm landfalls on specified geographical coordinates from the Gulf of Mexico.

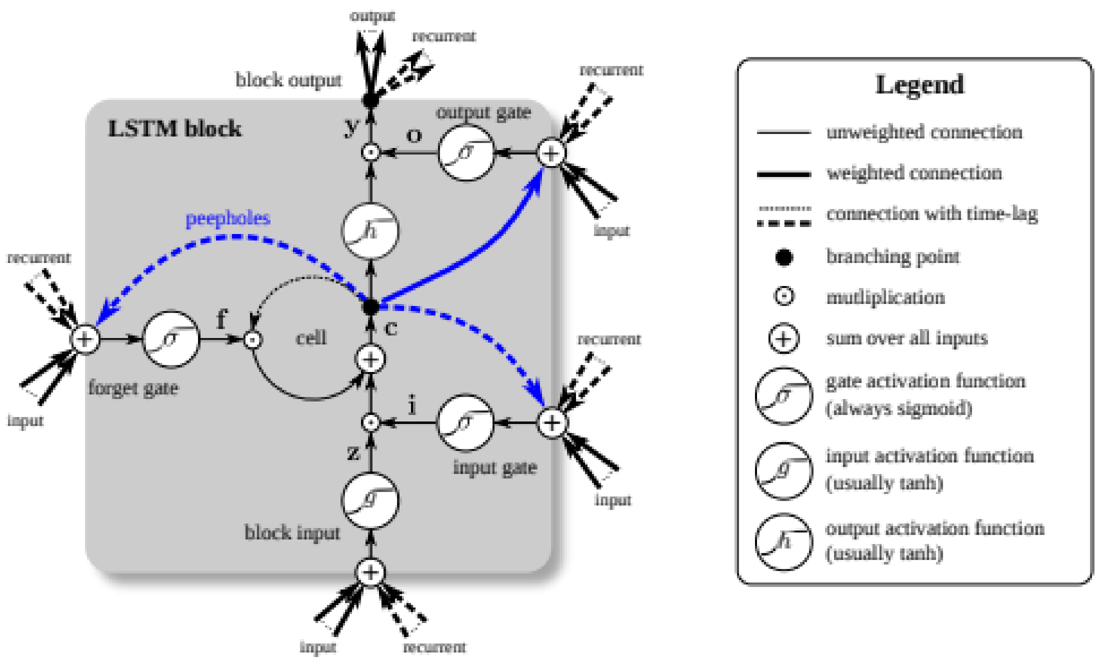
When it comes to Neural Networks for storm predictions, the two most common algorithms are Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN) – such as Long-Short Memory (LSTM). CNN's primary purpose is to break down an array into components defining patterns, with each layer learning a characteristic feature. As commonly thought, the deeper the layer networks, the better the Neural Network can learn (Figure 1).

<u>Figure 1: Visual representation of a CNN using a 3x3 matrix (Saha 2018)</u>



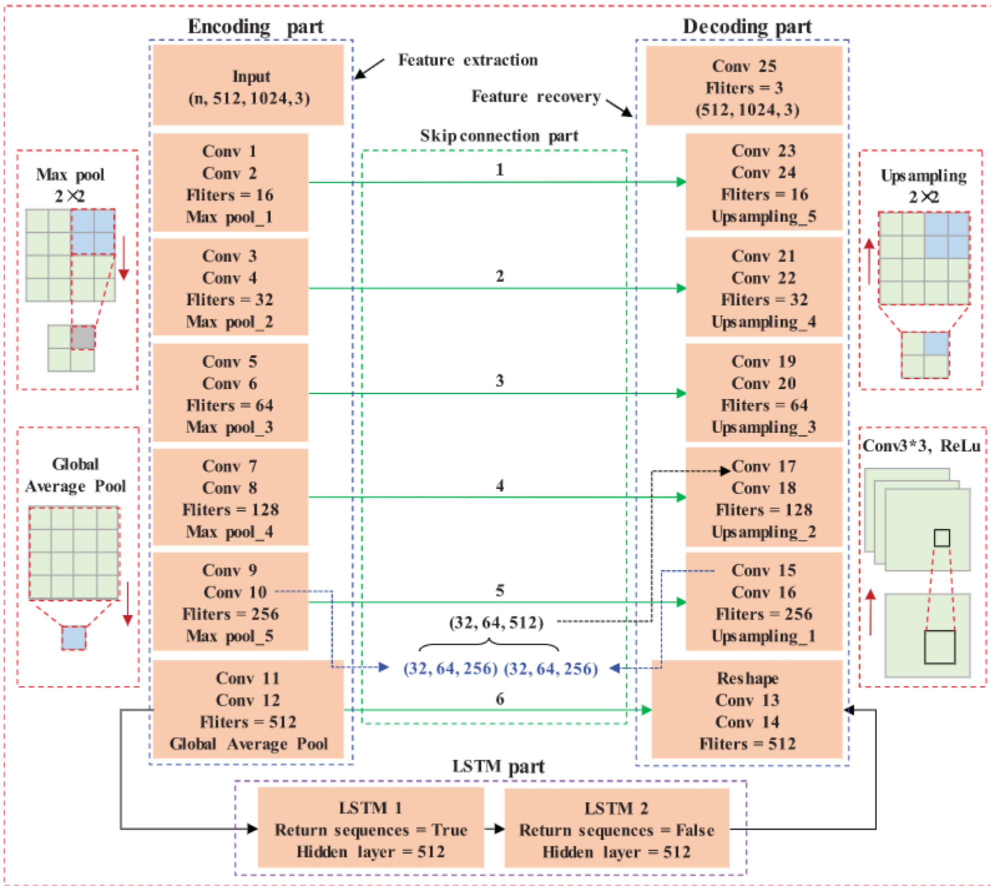
The second architecture, the LSTM, considers temporal trends (for example, speech and phrase recognition). While other algorithms have attempted RNNs, the LSTM is the most popular as it provides "non-linear gating units which regulate the information flow into and out of the cell" (Greff et al. 2015; Figure 2).

Figure 2: Visual representation of an LSTM algorithm (Greff et al. 2015)



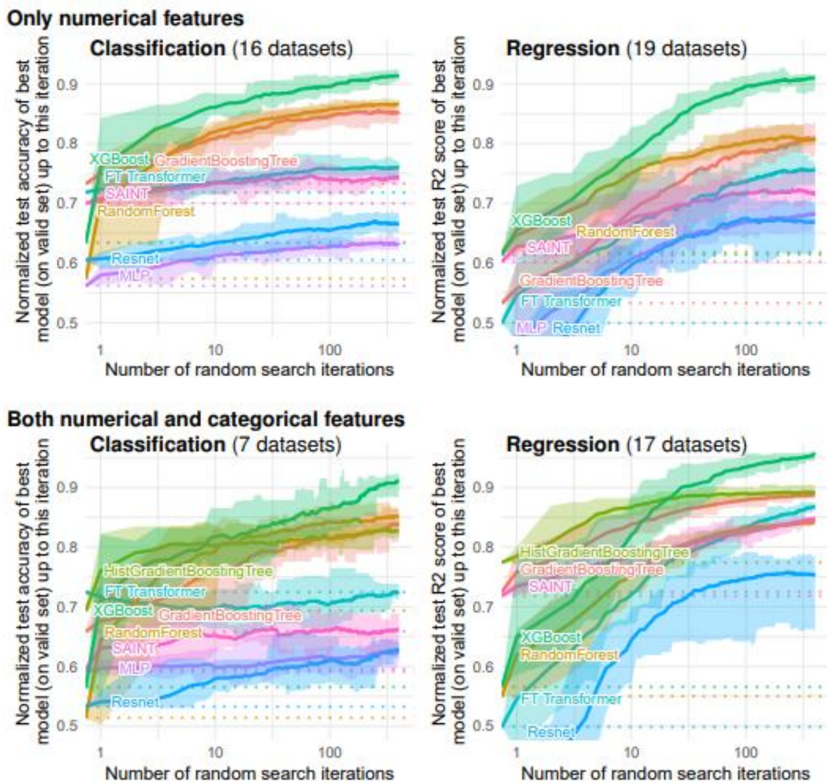
Previous authors found that a combination of LSTM and CNN outperforms each algorithm (Hou et al. 2022; Kim and Cho 2019). By combining LSTM and CNN approaches, (Hou et al. 2022) proposed a new deep U-Net-LSTM architecture combining the strength of both algorithms. Applying the two LSTM layers between the encoder and decoder portion of the U-Net (), the algorithm can increase prediction capabilities compared to other algorithms for hydrodynamics predictions (Figure 3; Ronneberger et al. 2015).

Figure 3: (Hou et al. 2022)'s proposed deep U-Net-LSTM framework



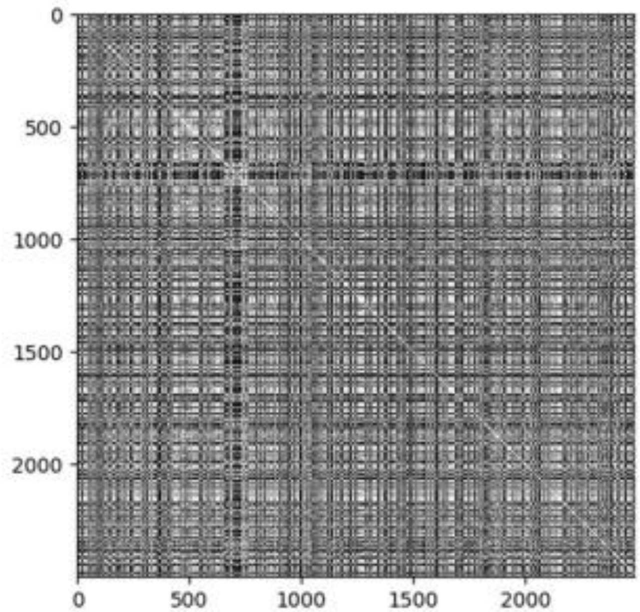
Interestingly, other researchers found that Neural Networks are not always the best solution for tabular datasets, like the one proposed in this report. (Grinszatjn et al. 2022) found that neural network predictions from tabular datasets suffer from three main points. First, “neural networks are biased to overly smooth solutions.” Second, “uninformative features affect more MLP-like neural networks.” Third, “data are non-invariant by rotation.” (Grinsztajn et al. 2022) found a clear advantage towards decision-tree algorithms from an analysis of numerical features using both neural networks and decision-trees (Figure 4).

Figure 4: Neural Network performance versus decision trees (Grinsztajn et al. 2022)



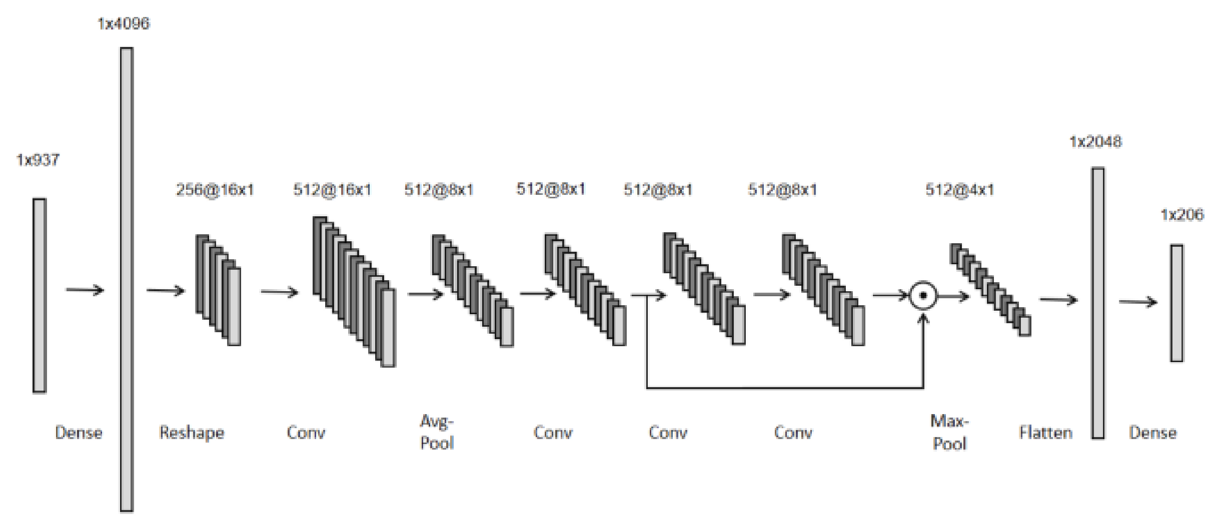
The tabular dataset challenge regarding neural networks has been tackled by (Zhu et al. 2021; Sharma et al. 2019; Villanueva 2021). (Zhu et al. 2021) created an image generator for tabular data (IGTD) capable of assigning features to a pixel "so that similar features are close to each other in the image." The optimization uses iterations to reduce the difference: $\mathrm{err}(\boldsymbol{R}, \boldsymbol{Q}) = \sum_{i=2}^{N} \sum_{j=1}^{i-1} \mathrm{diff}\left(r_{i,j}, q_{i,j}\right)$ between two rankings and assign a pixel intensity for each feature of the sample (Figure 5). Furthermore, the algorithm does not need domain expertise of the dataset for a preliminary understanding of the features represented in each tabular column. The results proved very promising, with an R^2 prediction accuracy reaching 0.86 on a sampled dataset. (Sharma et al. 2019) provided the same type of algorithm for gene clustering on various types of non-image datasets.

Figure 5: IGTD conversion of CCL gene expression data using Euclidean distances.



On the other hand, (Villanueva 2021) described the second-place algorithm for the "Mechanism of Action" Kaggle competition, whereby a competitor transformed the tabular dataset originally given into a feature space through a combination of dense and convolutional layers (Figure 6). While the algorithm in question removes the tabular shape of the dataset, it plots the tabular dataset's equivalent in multiple dimensions through a dense layer and then applies convolutional layers to learn features.

Figure 6: Soft-Ordering 1-dimensional CNN



III. Data

a. Data gathering

The current report relies on datasets collected from a variety of online sources. The paper's author is unaware of a pre-established combination of datasets of interest involving greenhouse gases for such a study. As such, a new dataset is built.

The newly created dataset consists of 12 columns (Table 1) and 28907 entries, each column of originally different file formats that need cleaning, briefly described below.

Table 1: Combined dataset

Column Name	Description	Source
date	Date given in Year/Month/Day format for data collection	All sources below
coordinates	Geographical coordinates of data collection (latitude, longitude)	All sources below
air_pressure(mBar)	Air pressure measured in mBar at location	(GCOOS, 2023)
air_temperature	Air temperature measured in degree Celsius at location	(GCOOS, 2023)
mass_concentration_of_chlorophyll_in_sea_water	Chlorophyll mass measured at location	(GCOOS, 2023)
mass_concentration_of_oxygen_in_sea_water	Oxygen concentration measured at location	(GCOOS, 2023)
relative_humidity	Relative humidity measured at location	(GCOOS, 2023)
sea_water_practical_salinity	Practical salinity measured at location	(GCOOS, 2023)
tropical_storm_occurrence	If a storm occurred at that location in the given date.	(Knapp et al., 2018)
USA_SSHS	Storm category based on NOAA terminologies.	(Knapp et al., 2018)
xco2	Average number of CO2 molecules uptake in the atmosphere above Gulf of Mexico as detected by OCO-2 satellite.	(EarthData, 2023)
Methane_ppbv	Average number of methane molecules uptake in the atmosphere in Alabama as detected by JAXA GoSAT satellite.	(EORC-JAXA, 2023; GoSAT Project, 2023)

While meteorological and methane data are collected as Excel CSV files, CO_2 concentrations consist of data from the OCO-2 satellite as nc4 files. Proper conversion to CSV files is warranted on the satellite dataset using the R software package ncdf4. Each dataset is additionally split into CSV files for each day, requiring a concatenation of the CSV datasets into single files for each attribute column name (achievable through the Windows command prompt). Finally, as the given datasets are too large to run on the author's computer, the data was reduced to no decimal place coordinates, no repeated coordinates per day, study years between 2015 and 2021 (with an unseen dataset for the year 2021), latitudes between 27 and 30.5, and longitudes between -97 and -87. The time frame also corresponds to a period post-aerosol hurricane frequency flaw, as described in the introduction of this paper and (Vecchi et al. 2021).

b. Data preprocessing

After importing the dataset into a new Jupyter notebook, one notices that a large percentage of the columns contain significant missing data with respect to coordinate values (Table 2).

Table 2: Missing values in each dataset

Column Name	**Percentage NaN (%)**
date	0
coordinates	0
air_pressure(mBar)	14
air_temperature	9
mass_concentration_of_chlorophyll_in_sea_water	96
mass_concentration_of_oxygen_in_sea_water	90
relative_humidity	81
sea_water_practical_salinity	71
tropical_storm_occurrence	0
USA_SSHS	99
xco2	91
average_Methane_ppb	99

It is important to notice that while xco2 and average methane have a high number of NAs, these values are averaged over a coordinate space and shift in coordinates through time. Therefore, the null values are filled using appropriate values corresponding to the respective dates. Furthermore, while the author found values charts for methane data from JAXA satellite for the years 2017 onward, this type of data did not appear to exist earlier. However, global methane maps in png format are available from (GOSAT Project 2023) where the author can infer methane concentrations from the missing years. Similarly, the "USA_SSHS" column corresponds to tropical storm occurrence and can therefore be replaced by a number not used in the column previously. Finally, as the rest of the columns with a high percentage of null values are hard to fill, we will discount them from the dataset. The final dataset consists of 8 columns and 26372 entries.

Following preprocessing, correlation plots for model variables (air pressure, air temperature, CO_2, and methane) against target variables (tropical storm strength, USA_SSHS, and tropical storm occurrence) (Figure 7) and the equivalent Pearson Correlation plot (Figure 8) do not show noticeable patterns between the variables.

Figure 7: Target variables seaborn pairplot

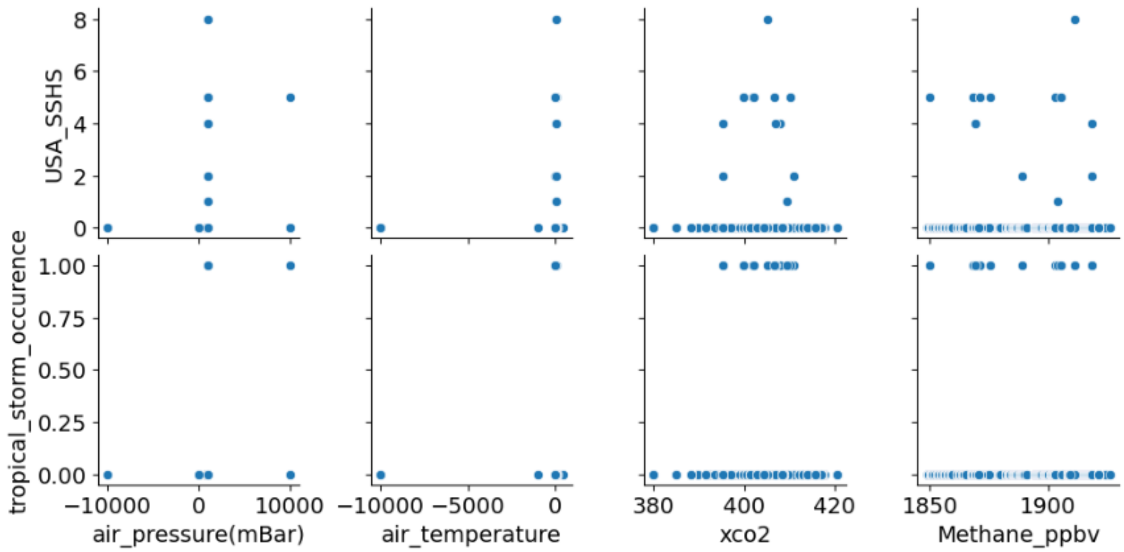
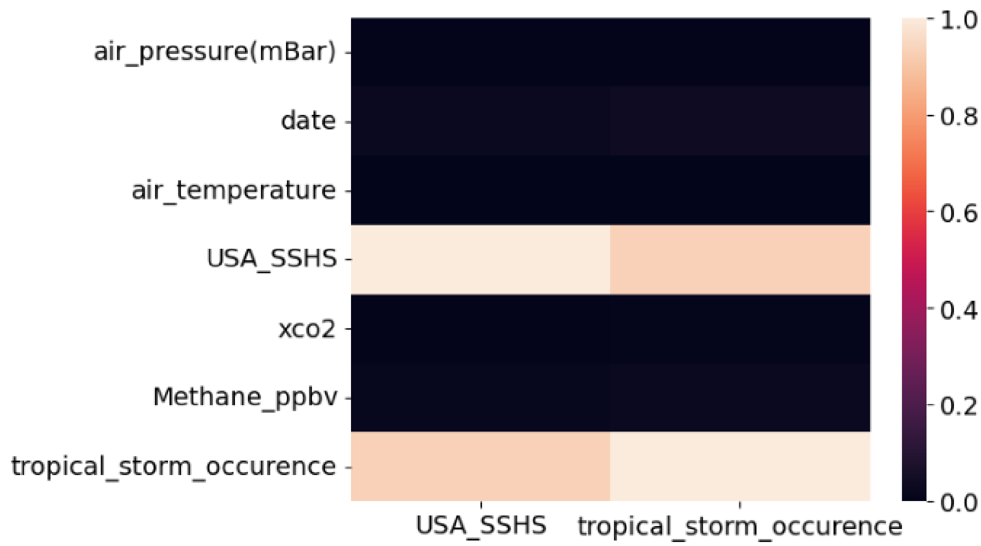


Figure 8: Target variables using Pearson Correlation seaborn plot

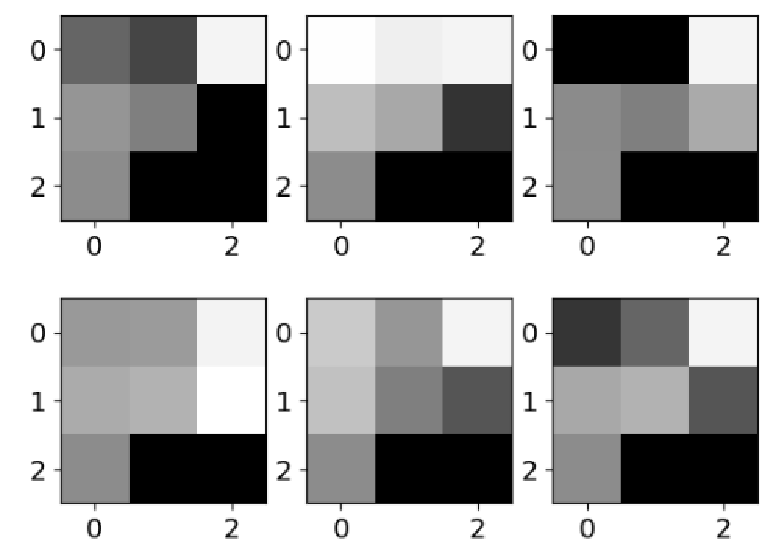


Finally, a limiting factor for an accurate input variable concerns the limited number of entries with a tropical storm hit. Specifically, out of 26372 rows, only 25 entries have a tropical hit. This imbalance of the dataset is resolved using SMOTE oversampling, whereby minority variables repeat in order to have interesting characteristics appear enough times for the model to learn.

IV. Methods

The workflow for algorithm building is broken into two parts: a Neural Network approach and a Decision Tree approach. For the Neural Network approach, the tabular dataset input tests the Soft-Ordering 1-dimensional CNN architecture (Villanueva 2021), and then having that same dataset converted to a pixelized image as described by (Sharma et al., 2019; Figure 9), again tested on the Soft-Ordering 1-dimensional CNN architecture.

Figure 9: Image representation of tabular row 1 to 9 (Sharma et al. 2019)



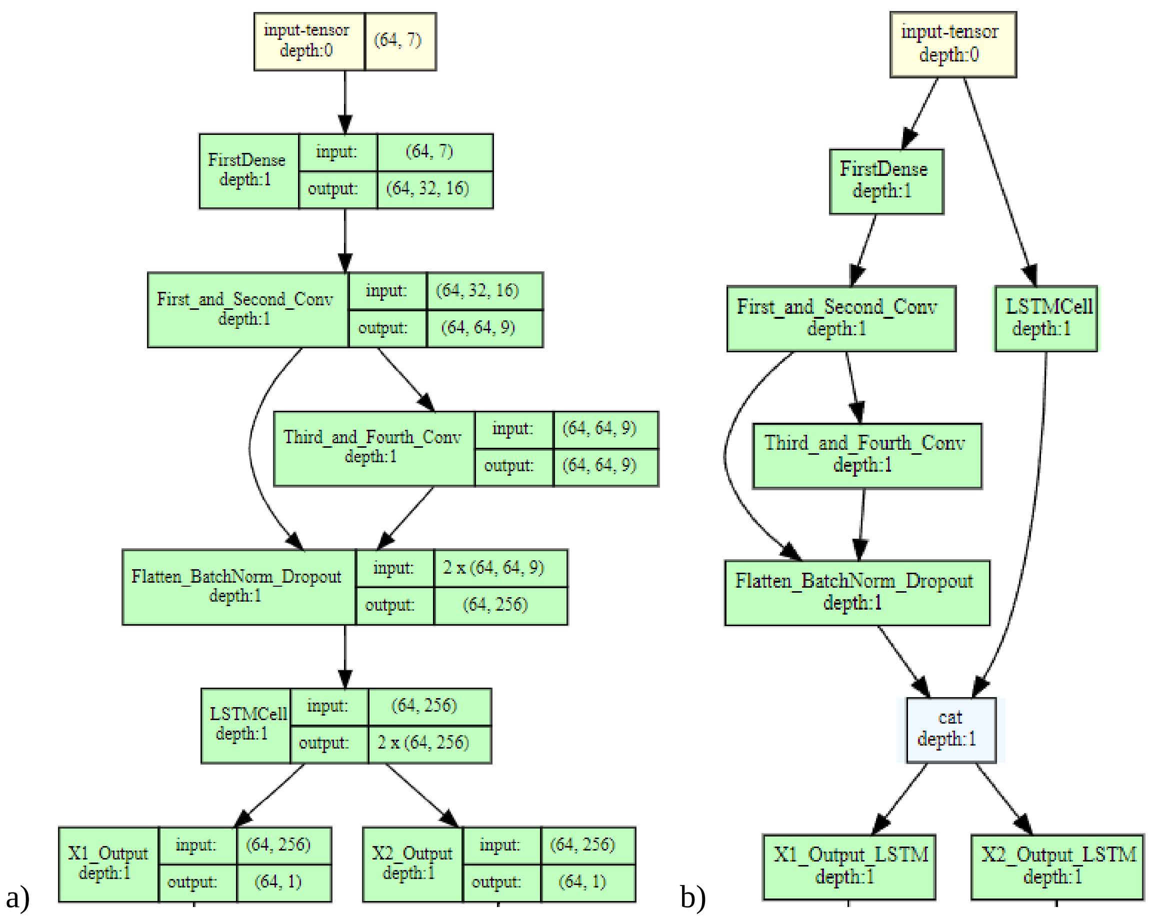
The Decision Tree approach will keep the dataset in its tabular form. The dataset is broken into a training set (80 percent) versus a test set (20 percent). Data from the year 2021 is removed from the dataset in order to act as an unseen dataset for final testing. The model uses seven columns as predictors (‘air_pressure(mBar)’, ‘date’, ‘latitude’, ‘longitude’, ‘air_temperature’, ‘xco2’, and ‘average_Methane_ppb’) and attempts to predict two columns (‘tropical_storm_occurence’, and ‘USA_SSHS’).

Neural Network models are trained using PyTorch. Xgboost package and sklearn for Decision Trees. The code runs on Jupyter Notebook, Windows 11, with AMD Ryzen 7 5700U with Radeon Graphics CPU and 8.00 GB RAM.

a. Neural Network approach

The first Neural Network model tests a Soft-Ordering 1-dimensional CNN with the tabular dataset format and a second network with the same Soft-Ordering 1-dimensional architecture without the first dense layer but using the tabular dataset’s image representation using Euclidean distances. The best model from the two network architectures is used for the Neural Network going forward. The following Neural Network model tests the same architecture, this time applying LSTM layers either at the end of a convolutional block, as described by (Hou et al. 2022; Figure 10), or as a separate branch altogether that is combined with the convolutional output at the end. The two output predictions for storm intensity and occurrence are compiled using a Rectified Linear Unit (ReLU) for multi-label predictions and Softmax activation for the binomial predictions. The ReLU activation is used for multi-label predictions as it retains more hidden layer information during the final output (Nair and Hinton 2010).

Figure 10: a) 1D Neural Network. LSTM and first Dense layer optional depending on model, b) Neural Network with LSTM separate from CNN.



Each of the two models outputs two predictions. The first output aims to predict tropical storm occurrence with accuracy given as:

$$Accuracy\_Precision = \sum_{0}^{length\ of\ training/testing\ set} \frac{y_{pred\_hit} == y_{true\_hit}}{y_{pred\_hit}}$$

The second model output aims to predict storm intensity. Accuracy is given as:

$$Accuracy = \sum_{0}^{length\ of\ training/testing\ set} y_{pred} == y_{true}$$

Three loss functions are provided for each model. One loss analyzes storm intensity ($\sum_0^{length\_training} |y_{pred} - y_{true}|$). Two loss functions look for the storm frequency occurrence: The first compares the standard difference between the two values using Binary Cross Entropy ($\sum_0^{length\_training} |y_{pred} - y_{true}|$). The second loss compares whether each positive hit has been correctly identified ($\sum_0^{y_{true}==1} |y_{pred} - y_{true}|$).

A fourth model is finally creating, incorporating a Physics-Informed Neural Network (PINN) loss (Cai et al. 2021). The PINN regularization term penalizes tropical storm-related intensities if a tropical storm event is not detected.

Iterations are equal to 700, batch size of 64, and epoch number equal to {iterations / ((size of training set) / (batch size))}, as inspired by (kanncaa1 2020). Loss functions are weighted according to their impact on model prediction, requiring a ratio (weight*loss) of the same magnitude across loss functions to penalize each function equally.

b. Decision Tree approach

(Grinsztajn et al. 2022) described XGBoost and Random Forests as performing superior to Neural Networks for tabular datasets. Therefore, this paper runs two model architectures, one for each algorithm. The hyperparameter search reveals that the best XGBoost model contains 1000 estimators, max depth of 10, 2 minimum sample split, 2 minimum samples leaf, and a learning rate of 0.01. A second XGBoost model searches for the importance of each input as it relates to the overall Decision Tree model using a barplot. The Random Forest algorithm additionally uses a hyperparameter search to reveal the best model parameters: 100 estimators, max depth of 100, max features as ‘log2’, minimum sample split of 2, and a squared error criterion. Data preprocessing did not find correlations between variables, so no features were combined to form a new input column.

V. Results

The results of the models given by the Neural Network models and the Decision Tree models are compared with respect to accuracy metrics for hurricane frequency and intensity. The models are run five times and averaged to give the results in (Table 3). Accuracy metrics for each of the five runs, as well as the loss and accuracy training loss are provided in the appendix section. The model runtimes are also provided.

Table 3: Prediction Metrics (average of 5 runs) on test dataset

Model Name	Storm Intensity Accuracy (%)	Frequency Precision (%)	Runtime
Tabular CNN	39.5	54.6	~7 min
Image CNN	37.1	55.9	~ 4 min
LSTM	49.5	20.2	~ 9 min 30 sec
CNN + LSTM separate	45.4	48.1	~ 10 min 30 sec
CNN + LSTM decoder	49.5	52.6	~ 10 min 30 sec
CNN + PINN	39.4	54.3	8 and 16 min
CNN + PINN Image	38.8	40.6	~ 5 min 30 sec
CNN + LSTM + PINN	43.6	49.0	~11 min 30 sec
XGBoost	65.1	99.8	~ 6 min 20 sec
Random Forest	4.5	100	~ 4 min 30 sec
Ensemble (CNN/LSTM decoder, CNN/LSTM separate/PINN, XGBoost)	51.1	99.8	N/A

The same results are collected for predictions on the unseen 2022 dataset (Table 4).

Table 4: Prediction Metrics on unseen dataset

Model Name	Storm Intensity Accuracy (%)	Frequency Precision (%)
Tabular CNN	80.5	0.2
Image CNN	77.5	0.0
LSTM	99.8	0.1
CNN + LSTM separate	91.9	0.1
CNN + LSTM decoder	99.8	0.2
CNN + PINN	77.7	0.2
CNN + PINN Image	78.1	0.1
CNN + LSTM + PINN	86.2	0.3
XGBoost	14.7	0.0
Random Forest	4.5	0.0
Ensemble (CNN/LSTM decoder, CNN/LSTM separate/PINN, XGBoost)	96.6	0.0

VI. Analysis and Interpretation

The first important flaw that needs to be mentioned is that the Neural Network models have a hard time finding a gradient descent route. From the iteration versus accuracy or loss plots, as seen in the appendix, it is clear that the learning is stagnant, altering between 30 and 70 percent accuracy with each iteration. Despite hyperparameter tuning, such as learning rates, the model does not appear to improve. Interestingly, with five model runs, there does not appear to be visible variations in prediction accuracies in both the test dataset and the unseen dataset.

Model results indicate that using a combination of Convolutional Neural Networks (CNN), Long Short-Term Memory (LSTM), and Physics-Informed Neural Network (PINN) provides the best accuracy for intensity values out of five runs. Furthermore, XGBoost Decision Trees provide the best precision for frequency predictions out of five models. Therefore, the best Neural Network model and the best Decision Tree model are combined in an ensemble model, increasing the overall storm intensity and frequency accuracies.

When one compares the accuracy and precision metrics for storm frequency and intensity (Figure 11; Figure 12), one sees a clear overfit on the test set compared to the unseen dataset. This is evident as the model learns from a limited range of dates between 2015 and 2021, leaving the unseen dataset of 2022 with unseen data – that is, we should mention, also being influenced by climate change scenarios. Similarly, based on the average metrics for each run, there is an increase in storm intensity accuracy on the unseen dataset as compared to the testing dataset, while at the same time a significant decrease in storm frequency precision on the unseen dataset as compared to the testing dataset. This leads to the conclusion that the model is able to find better patterns in storm intensity predictions, and is not able to understand how storm intensity is related to storm frequency occurrences (despite manually forcing this distinction in the PINN implementation). After manually looking for storm frequencies based on intensity measurements, the author can conclude that there is not enough storm occurrence data as the test set is not able to find any storm occurrences as predictions despite having 2978 storm occurrences after SMOTE increase.

Figure 11: Confusion Matrix for Storm Frequency predictions

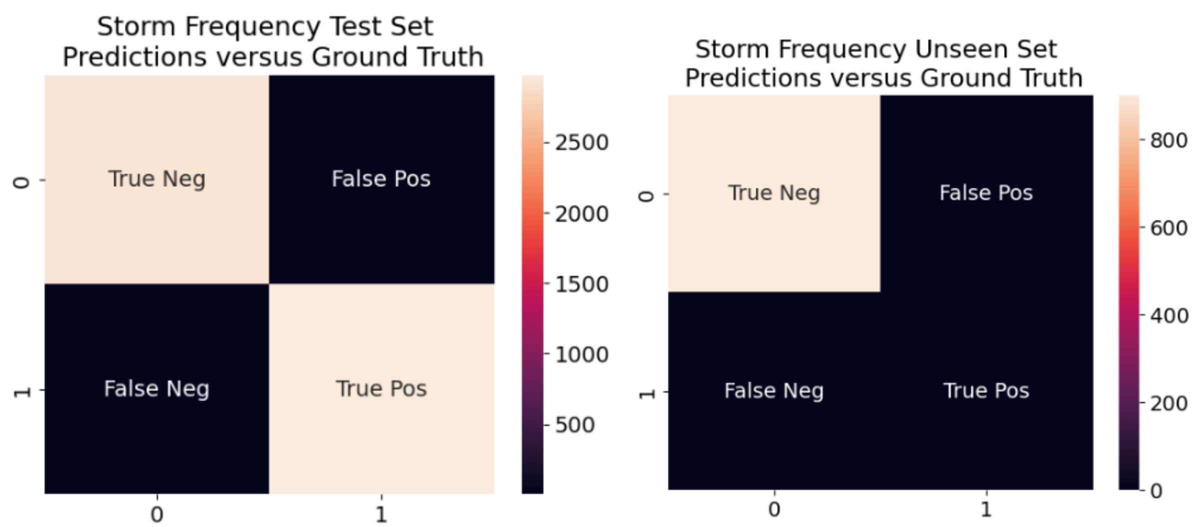
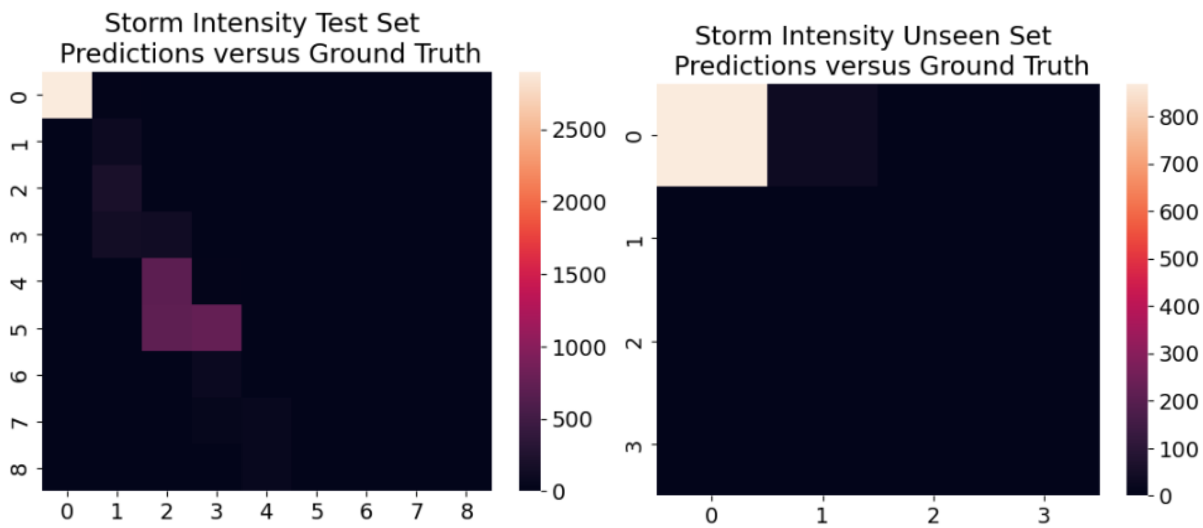
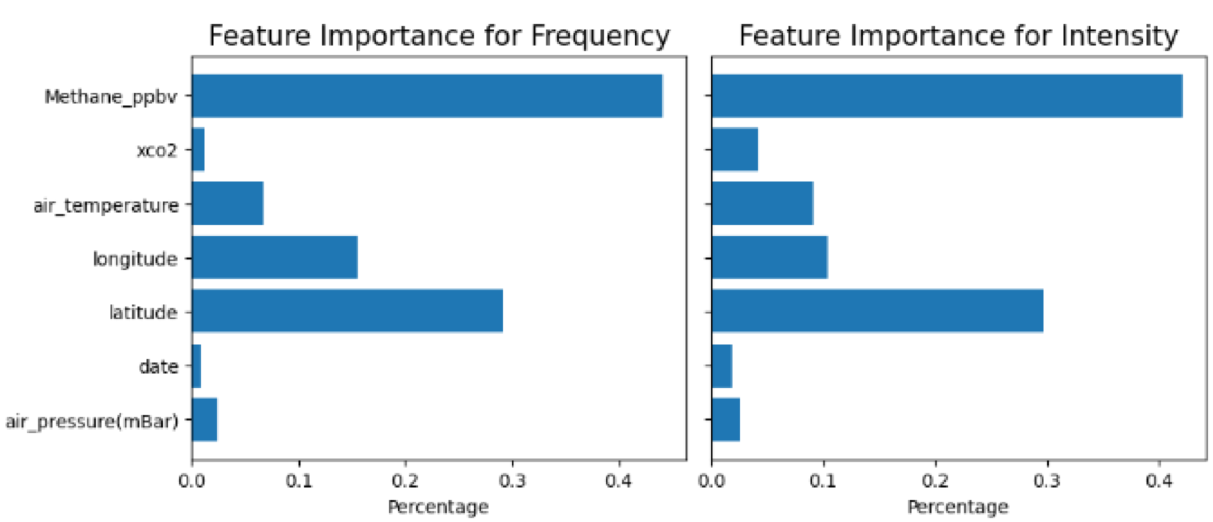


Figure 12: Confusion Matrix for Storm Intensity predictions



Interestingly, going beyond model performance, feature importance based on XGBoost analysis shows that latitude, air temperature, and methane data contribute the most to the final model prediction (Figure 13). This finding further reinforces the notion that greenhouse gas and air temperature change, relevant for climate change studies, could be a factor in determining storm frequency occurrences.

Figure 13: Feature importance horizontal barplot based on XGBoost model analysis



Based on IPCC predictions for the year 2100, the ensemble model predicted the storm frequency occurrences based on a changing climate condition extracted from the testing dataset. The carbon dioxide values were doubled (IPCC 2018), the air temperature increased by 3*C (Herring et al. 2012), and methane concentrations doubled (IPCC 2021). The predictions, based on a model with low confidence, did not show an increase in storm frequency.

VII. Conclusions

This report describes a first attempt at predicting storm frequencies in the Gulf of Mexico related to greenhouse gases and climate change. From a tabular dataset built on datasets collected from various governmental and academic sources, neural networks and decision trees algorithms were built to test which architecture provides the best model predictions. The dataset was cleaned and preprocessed to account for the oversampling of days without hurricanes. The dataset was input in its tabular form and an equivalent image representations. To increase model prediction accuracy, the neural network outputs included losses from storm intensity, storm frequency occurrence, and physical PINN constraints penalizing storm intensities that did not register as a frequency occurrence. The result, despite needing more work, shows a decent accuracy compromise combining an XGBoost decision tree and a neural network combining LSTM, PINN, and CNN in tabular data form. Extrapolating the model further to include IPCC predictions in air temperature, carbon dioxide, and methane shows, with small confidence, no increase in the frequency of hurricanes.

VIII. Directions for Future Work

The first main limitation surrounding the dataset concerns the CO2 and methane dataset. These samples are averaged over months and days over the Gulf of Mexico and, as such, do not correspond to CO2 and methane concentration at the specified geographical coordinate and time. As these variables are key to this study, the author wishes to find more accurate CO2 and methane datasets as more satellite measurements are made.

Furthermore, as reported by (Vecchi et al. 2021), aerosols have been found to mask changes in hurricane frequencies. A better understanding of aerosol behavior as it relates to hurricane intensity and frequency is needed. Finally, as this study is limited by computational availabilities, looking more in-depth at data from more detailed geographical coordinates, and expanding the search out of the Gulf of Mexico, would provide greater insights into the relationships at play.

IX. Acknowledgments

The author would like to thank the National Oceanic and Atmospheric Administration (NOAA), the National Centers for Environmental Information (NCEI), and the National Aeronautics and Space Administration (NASA) for providing open-source datasets for this research. Further mention to Northwestern University, Dr. Alianna J. Maren, and Robert Guenther for reviewing this paper and guiding the author through the research process.

Finally, this work would not have been possible without the constant support of the author’s family (Yan, Anh, Nicolas, and Olivier), his girlfriend (Morgane), special mention to Scottey the cat, and to his friends who supported him throughout the Master’s and this Capstone project. The author is forever grateful.

X. References (explain how your work fits into the world, a book)

Bose, Rikhi, Adam L. Pintar, and Emil Simiu. "Forecasting the Evolution of North Atlantic Hurricanes: A Deep Learning Approach." NIST. Rikhi Bose, Adam L. Pintar, Emil Simiu, June 2021. https://www.nist.gov/publications/forecasting-evolution-north-atlantic-hurricanes-deep-learning-approach.

Boussioux, Léonard, Cynthia Zeng, Théo Guénais, and Dimitris Bertsimas. "Hurricane Forecasting: A Novel Multimodal Machine Learning Framework." *Weather and Forecasting* 37, no. 6 (September 2022): 817–31. https://doi.org/10.1175/waf-d-21-0091.1.

Cai, Shengze, Zhiping Mao, Zhicheng Wang, Minglang Yin, and George Em Karniadakis. "Physics-Informed Neural Networks (Pinns) for Fluid Mechanics: A Review." *Acta Mechanica Sinica* 37, no. 12 (2021): 1727–38. https://doi.org/10.1007/s10409-021-01148-1.

Chung, Junyoung, Caglar Gulcehre, KyungHyun Cho, and Yoshua Bengio. "Empirical evaluation of gated recurrent neural networks on sequence modeling." *CoRR* (2014).

Cloud, Kirkwood A., Brian J. Reich, Christopher M. Rozoff, Stefano Alessandrini, William E. Lewis, and Luca Delle Monache. "A Feed Forward Neural Network Based on Model Output Statistics for Short-Term Hurricane Intensity Prediction." *Weather and Forecasting* 34, no. 4 (August 2019): 985–97. https://doi.org/10.1175/waf-d-18-0173.1.

"GCOOS Web Accessible Folder." Gulf of Mexico Coastal Ocean Observing System. Accessed January 16, 2023. https://data.gcoos.org/waf.php.

"Global Concentrations of CO2 and CH4: Greenhouse Gases Observing Satellite." GoSAT Project. Accessed February 14, 2023. https://www.gosat.nies.go.jp/en/global_ghg_simulation.html.

Greff, Klaus, Rupesh K. Srivastava, Jan Koutnik, Bas R. Steunebrink, and Jurgen Schmidhuber. "LSTM: A Search Space Odyssey." *IEEE Transactions on Neural Networks and Learning Systems* 28, no. 10 (2015): 2222–32.

Grinsztajn, Leo, Edouard Oyallon, and Gael Varoquaux. "Why do tree-based models outperform deep learning on typical tabular data?" *NeurIPS 2022* (2022).

Herring, David. "Climate Change: Global Temperature Projections." NOAA Climate.gov. Accessed February 28, 2023. https://www.climate.gov/news-features/understanding-climate/climate-change-global-temperature-projections#:~:text=Results%20from%20a%20wide%20range,gases%20that%20human%20activities%20produce.

"Home." JAXA. Accessed February 14, 2023. https://www.eorc.jaxa.jp/GOSAT/index.html.

Hou, Yuqing, Hui Li, Hong Chen, Wei Wei, Jiayue Wang, and Yicang Huang. “A Novel Deep U-Net-LSTM Framework for Time-Sequenced Hydrodynamics Prediction of the SUBOFF AFF-8.” *Engineering Applications of Computational Fluid Mechanics* 16, no. 1 (2022): 630–45. https://doi.org/10.1080/19942060.2022.2030802.

“Hurricane Katrina Statistics Fast Facts.” CNN Editorial Research, January 16, 2023. https://edition.cnn.com/2013/08/23/us/hurricane-katrina-statistics-fast-facts/index.html.

Intergovernmental Panel on Climate Change (IPCC) “IPCC — Intergovernmental Panel on Climate Change.” (2018) Accessed February 28, 2023. https://www.ipcc.ch/site/assets/uploads/2018/02/ar4-wg1-chapter10-1.pdf.

Intergovernmental Panel on Climate Change (IPCC), “Climate change 2021 the physical science basis. Contribution of Working Group I to the Sixth Assessment Report of the intergovernmental panel on climate change” (2021).

kanncaa1. “Long-Short Term Memory with Pytorch.” Kaggle. Kaggle, April 8, 2020. https://www.kaggle.com/code/kanncaa1/long-short-term-memory-with-pytorch.

Kim, Tae-Young, and Sung-Bae Cho. “Predicting Residential Energy Consumption Using CNN-LSTM Neural Networks.” *Energy* 182 (2019): 72–81. https://doi.org/10.1016/j.energy.2019.05.230.

Knapp, Kenneth R., Howard J. Diamond, James P. Kossin, Michael C. Kruk, and Carl J. Schreck. “International Best Track Archive for Climate Stewardship (IBTrACS) Project.” *National Centers for Environmental Information (NCEI)* 4, Ibtracs NA (January 18, 2023). https://doi:10.25921/82ty-9e16.

LeCun, Yann, Bernhard Boser, John S. Denker, Donnie Henderson, R. E. Howard, Wayne Hubbard, and Lawrence D. Jackel. “Backpropagation Applied to Handwritten Zip Code Recognition.” *Neural Computation* 1, no. 4 (1989): 541–51. https://doi.org/10.1162/neco.1989.1.4.541.

Marsooli, Reza, Ning Lin, Kerry Emanuel, and Kairui Feng. “Climate Change Exacerbates Hurricane Flood Hazards along US Atlantic and Gulf Coasts in Spatially Varying Patterns.” *Nature Communications* 10, no. 1 (August 22, 2019). https://doi.org/10.1038/s41467-019-11755-z.

Mudd, Lauren, Yue Wang, Chris Letchford, and David Rosowsky. “Assessing Climate Change Impact on the U.S. East Coast Hurricane Hazard: Temperature, Frequency, and Track.” *Natural Hazards Review* 15, no. 3 (2014). https://doi.org/10.1061/(asce)nh.1527-6996.0000128.

“OCO-2 Level 2 Bias-Corrected XCO2 and Other Select Fields from the Full-Physics Retrieval Aggregated as Daily Files, Retrospective Processing V10R (oco2_l2_lite_fp) at Ges Disc.”

Catalog. Publisher NASA/GSFC/SED/ESD/GCDC/GESDISC, December 1, 2022. https://disc.gsfc.nasa.gov/datasets/OCO2_L2_Lite_FP_10r/summary?keywords=OCO-2.

Richard, Olivier. "Natural Disasters: Economic Effects of Hurricanes Katrina, Sandy, Harvey, and Irma." Natural Disasters: Economic Effects of Hurricanes Katrina, Sandy, Harvey, and Irma | U.S. GAO. U.S. Government Accountability, September 10, 2020. https://www.gao.gov/products/gao-20-633r.

Ronneberger, Olaf, Philipp Fischer, and Thomas Brox. "U-Net: Convolutional Networks for Biomedical Image Segmentation." *arXiv:1505.04597*, cs.CV (May 18, 2015).

Saha, Sumit. "A Comprehensive Guide to Convolutional Neural Networks-the eli5 Way." Medium. Towards Data Science, December 15, 2018. https://towardsdatascience.com/a-comprehensive-guide-to-convolutional-neural-networks-the-eli5-way-3bd2b1164a53.

Sharma, Alok, Edwin Vans, Daichi Shigemizu, Keith A. Boroevich, and Tatsuhiko Tsunoda. "DeepInsight: A Methodology to Transform a Non-Image Data to an Image for Convolution Neural Network Architecture." *Scientific Reports* 9, no. 11399 (2019). https://doi.org/10.1038/s41598-019-47765-6.

"The Latest IPCC Report: What Is It and Why Does It Matter?" The Nature Conservancy, April 4, 2022. https://www.nature.org/en-us/what-we-do/our-insights/perspectives/ipcc-report-climate-change/.

Vaswani, Ashish, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. "Attention is all you need." *Advances in Neural Information Processing Systems* 30 (2017): 5998-6008.

Vecchi, Gabriel A., Christopher Landsea, Wei Zhang, Gabriele Villarini, and Thomas Knutson. "Changes in Atlantic Major Hurricane Frequency since the Late-19th Century." *Nature Communications* 12, no. 4054 (July 13, 2021).

Villanueva, Martín. "Convolutional Neural Networks on Tabular Datasets (Part 1)." Medium, April 29, 2021. https://medium.com/spikelab/convolutional-neural-networks-on-tabular-datasets-part-1-4abdd67795b6.

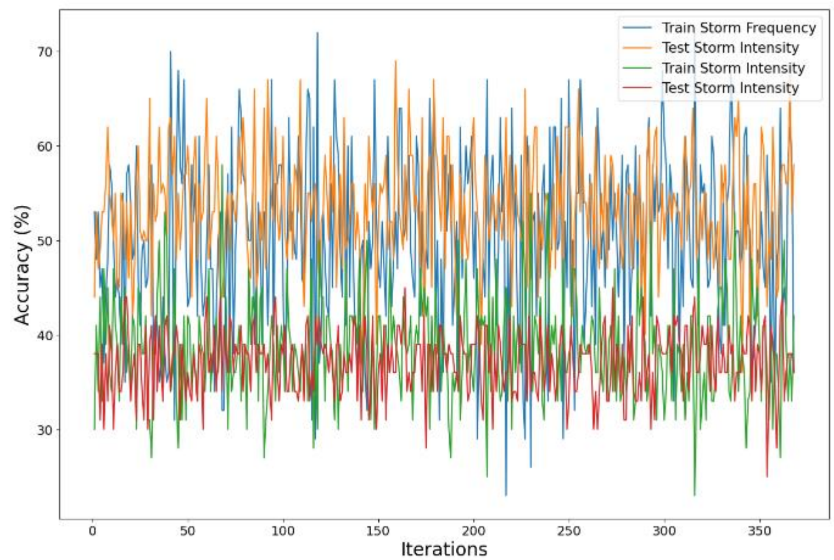
Vinod Nair and Geoffrey E. Hinton. "Rectified Linear Units Improve Restricted Boltzmann Machines." *Proceedings of the 27th International Conference on Machine Learning* (2010).

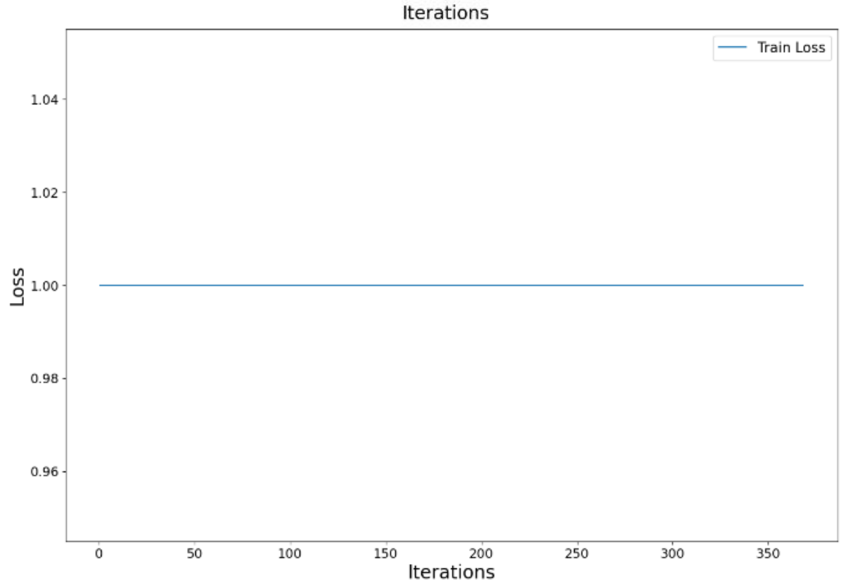
Winkler, Tyler S., Peter J. van Hengstum, Jeffrey P. Donnelly, Elizabeth J. Wallace, Nicole D'Entremont, Andrea D. Hawkes, Christopher V. Maio, Richard M. Sullivan, and Jonathan D. Woodruff. "Oceanic Passage of Hurricanes across Cay Sal Bank in the Bahamas over the Last 530 Years." *Marine Geology* 443 (January 2022): 106653. https://doi.org/10.1016/j.margeo.2021.106653.

Zhu, Yitan, Thomas Brettin, Fangfang Xia, Alexander Partin, Maulik Shukla, Hyunseung Yoo, Yvonne A. Evrard, James H. Doroshow, and Rick L. Stevens. "Converting Tabular Data into Images for Deep Learning with Convolutional Neural Networks." *Scientific Reports*, 11325, 11, no. 1 (2021). https://doi.org/10.1038/s41598-021-90923-y.

XI. Appendices

a. Tabular CNN

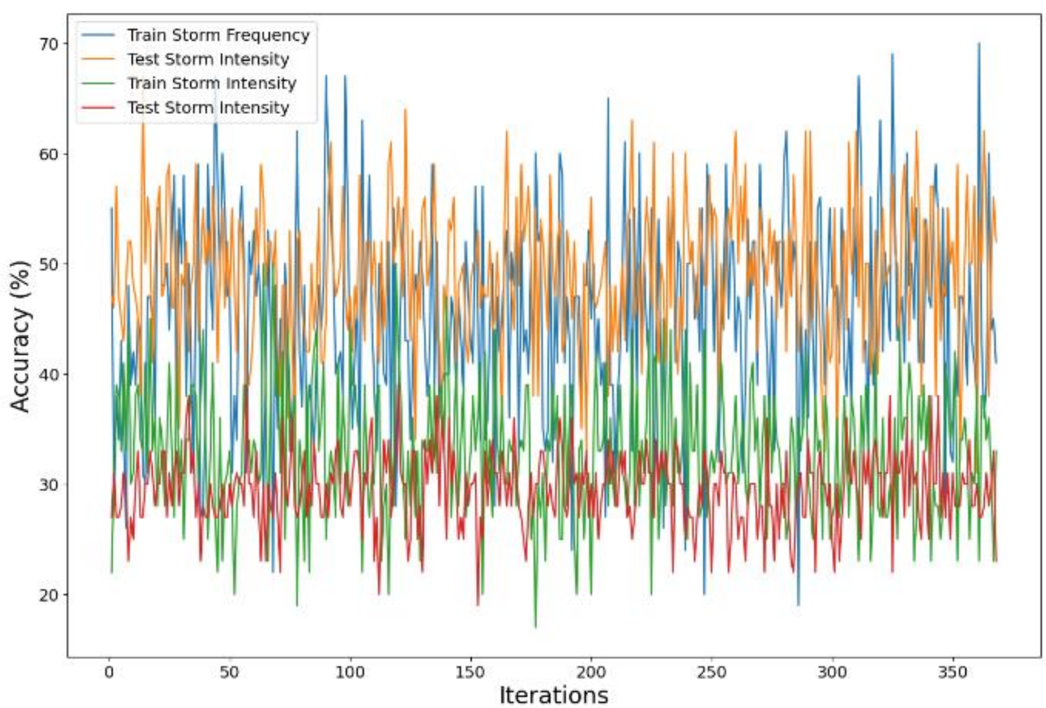


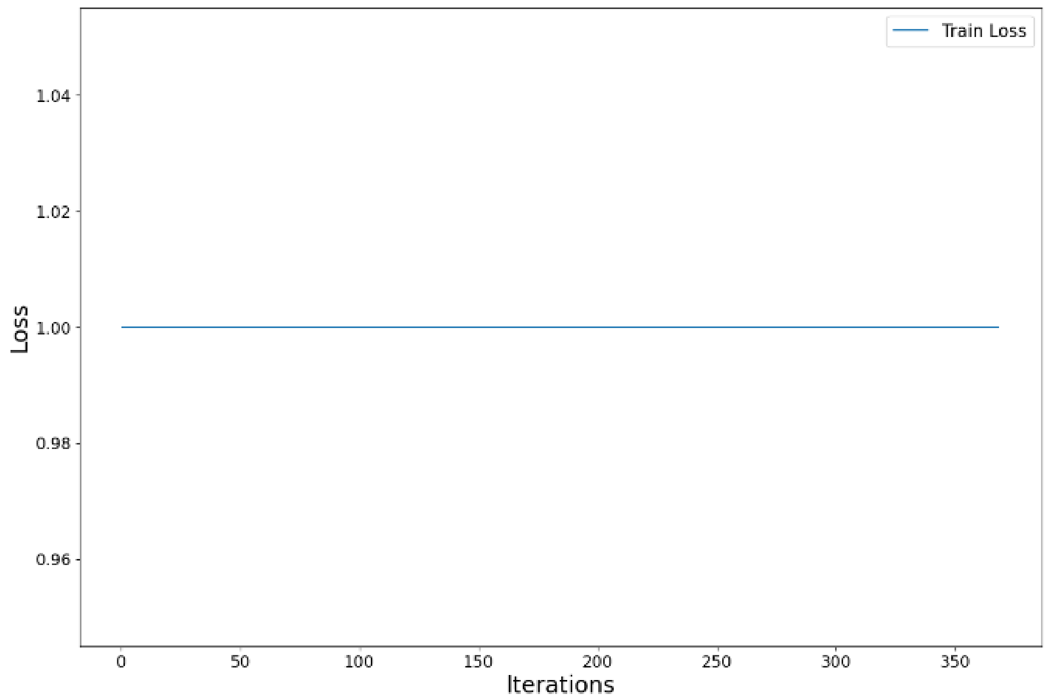


Results from five runs:

Model Name	Storm Intensity Accuracy (%)	Frequency Precision (%)	Model Runtime
Tabular CNN Test	37.7	50.2	7 min 5 sec
	38.2	50.5	6 min 23 sec
	41.7	61.6	7 min 10 sec
	41.6	55.7	7 min 20 sec
	38.1	55.0	7 min 19 sec
Tabular CNN Unseen	75.2	0.5	
	80.5	0.2	
	82.07	0.0	
	81.5	0.5	
	83.1	0.0	

b. <u>Image CNN</u>

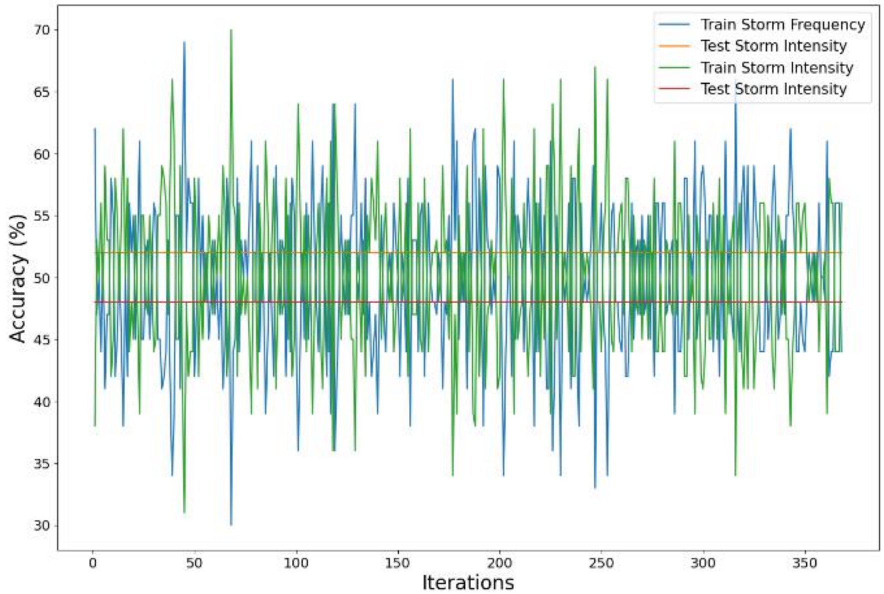


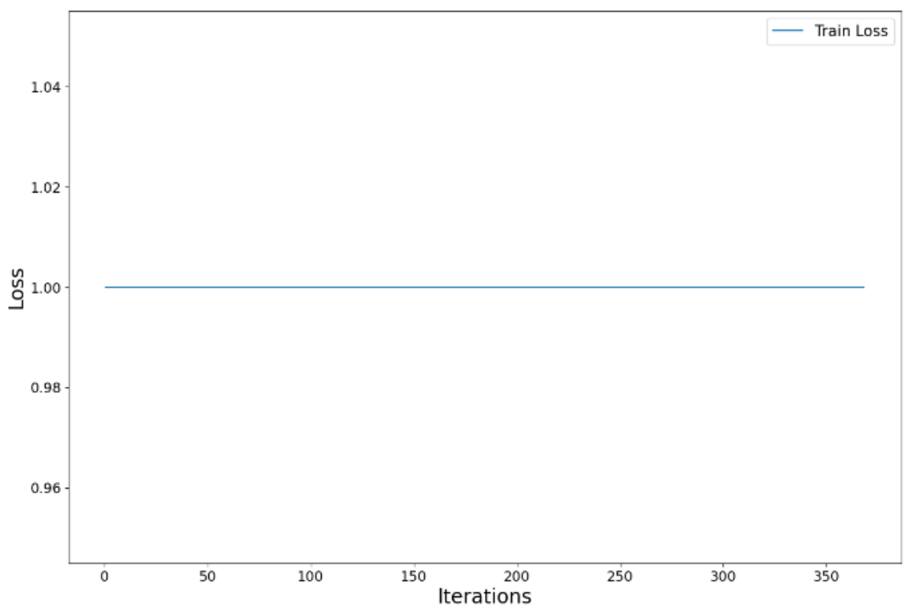


Results from five runs:

Model Name	Storm Intensity Accuracy (%)	Frequency Precision (%)	Runtime
Image CNN Test	32.4	44.9	4 min 24 sec
	42.6	63.5	3 min 54 sec
	31.9	59.9	4 min 4 sec
	37.5	58.5	3 min 58 sec
	40.9	52.7	4 min 1 sec
Image CNN Unseen	76.1	0.0	
	77.9	0.0	
	76.6	0.2	
	78.8	0.0	
	78.0	0.0	

c. <u>LSTM</u>



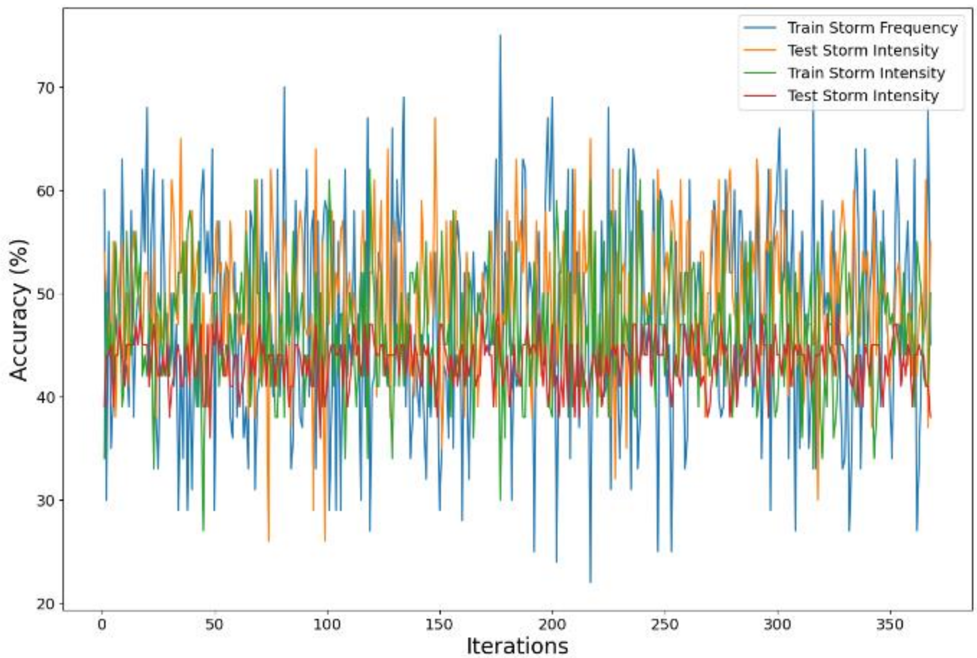


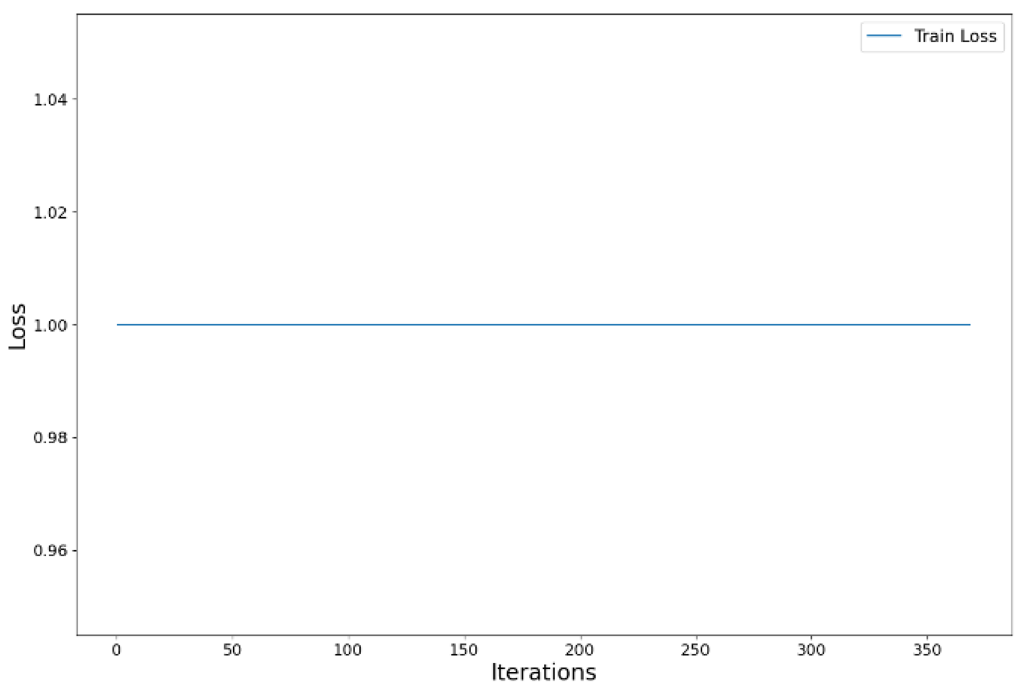
Results from five runs:

Model Name	Storm Intensity Accuracy (%)	Frequency Precision (%)	Runtime
LSTM Test	49.5	50.5	9 min 35 sec
	49.5	0.0	9 min 44 sec
	49.5	50.5	9 min 17 sec
	49.5	0.0	9 min 17 sec

	49.5	0.0	9 min 23 sec
LSTM Unseen	99.8	0.2	
	99.8	0.0	
	99.8	0.2	
	99.8	0.0	
	99.8	0.0	

d. <u>CNN + LSTM separate</u>

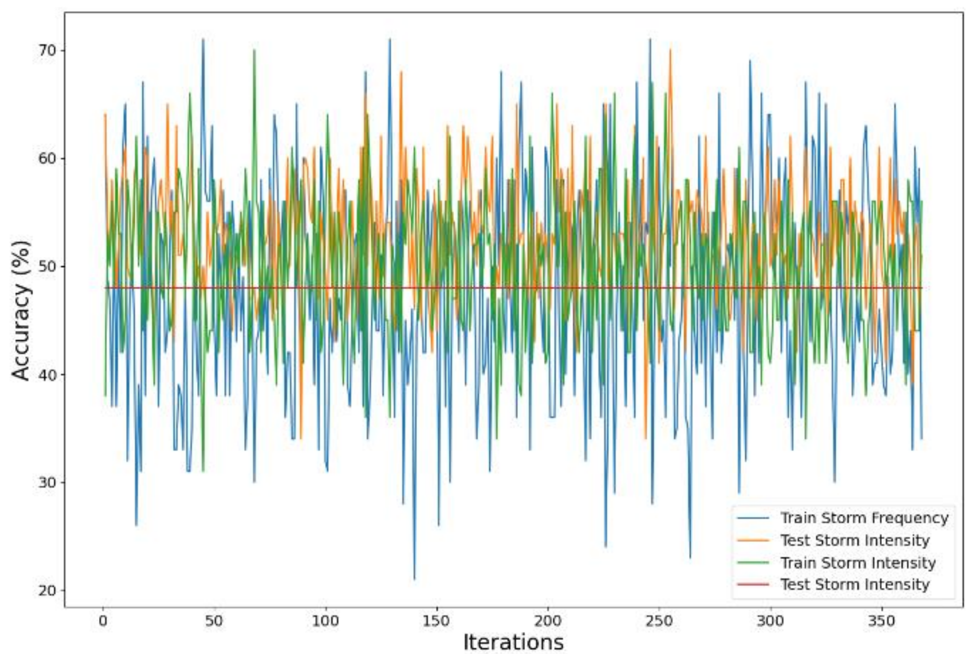


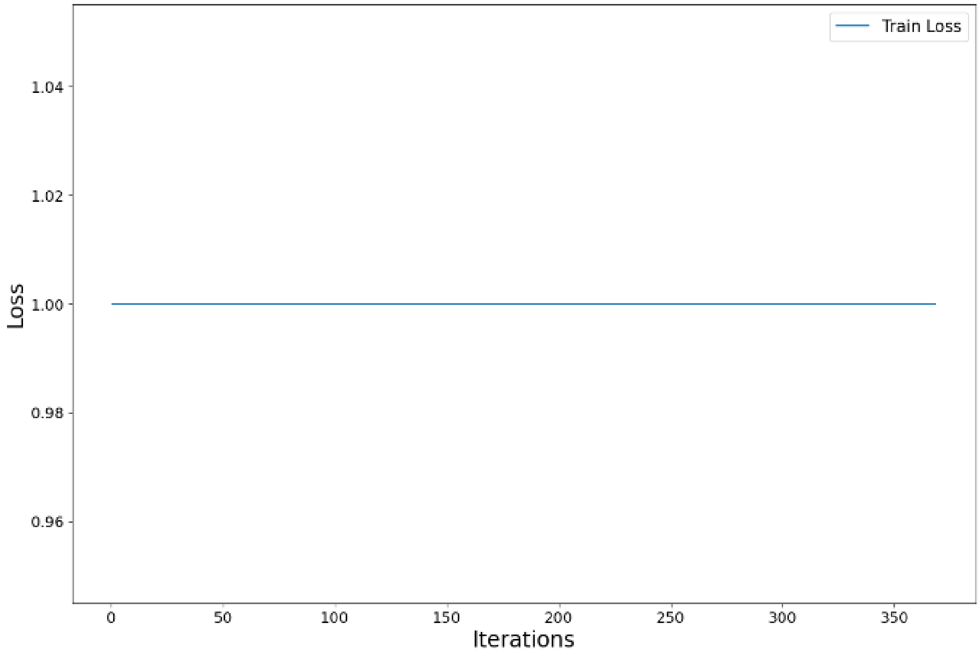


Results from five runs:

Model Name	Storm Intensity Accuracy (%)	Frequency Precision (%)	Runtime
CNN + LSTM Test	45.4	49.5	10 min 36 sec
	45.3	50.4	10 min 16 sec
	48.4	40.6	10 min 32 sec
	46.5	51.0	10 min 31 sec
	41.3	48.9	10 min 34 sec
CNN + LSTM Unseen	89.1	0.2	
	93.5	0.0	
	97.2	0.0	
	92.6	0.2	
	87.1	0.3	

e. <u>CNN + LSTM decoder</u>

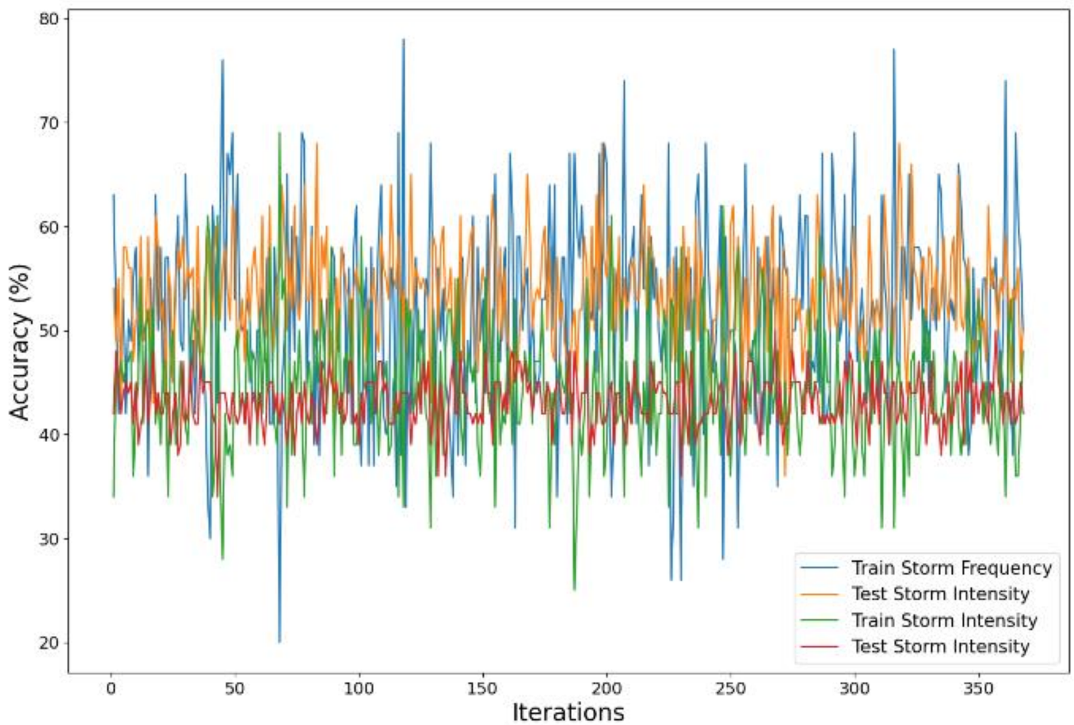




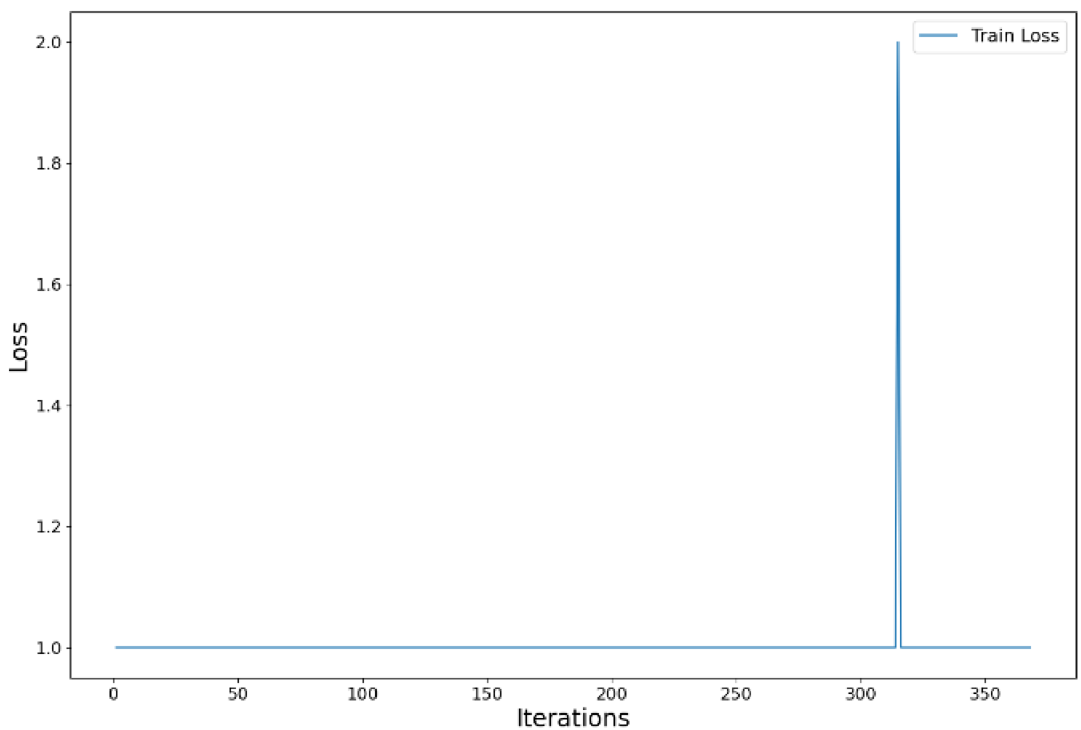
Results from five runs:

Model Name	Storm Intensity Accuracy (%)	Frequency Precision (%)	Runtime
CNN + LSTM Test	49.5	47.9	10 min 32 sec
	49.5	56.3	11 min 4 sec
	49.5	48.0	10 min 23 sec
	49.5	60.1	10 min 20 sec
	49.5	50.5	10 min 58 sec
CNN + LSTM Unseen	99.8	0.4	
	99.8	0.2	
	99.8	0.2	
	99.8	0.2	
	99.8	0.0	

f. CNN + PINN



Train Storm Frequency
Test Storm Intensity
Train Storm Intensity
Test Storm Intensity
Accuracy (%)
Iterations

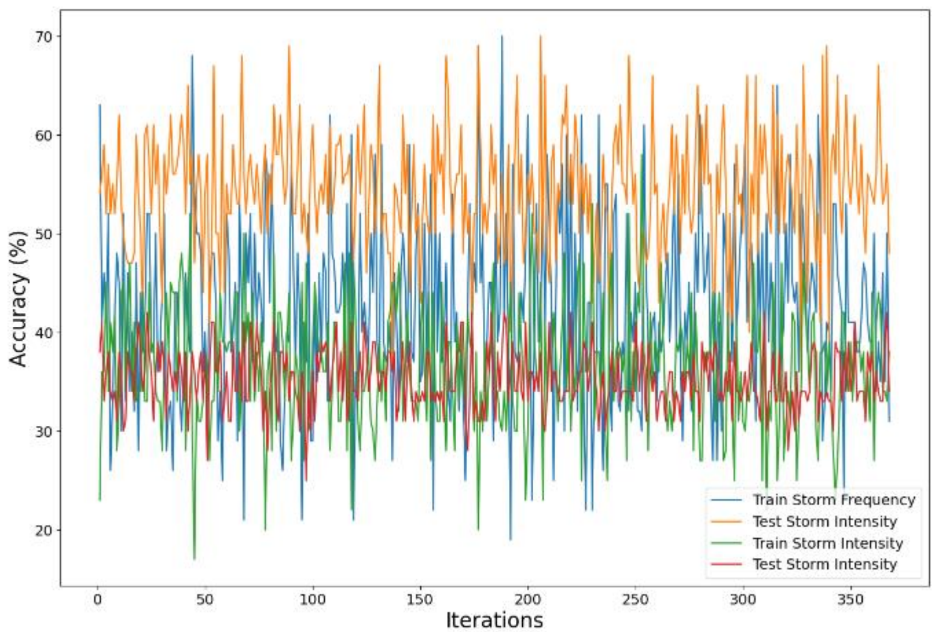


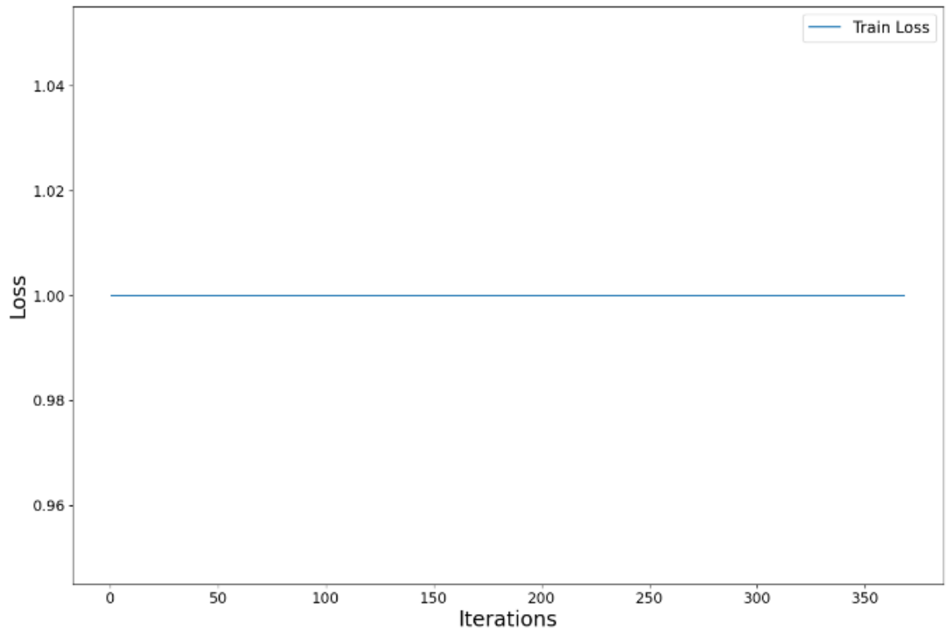
Train Loss
Loss
Iterations

Results from five runs:

Model Name	Storm Intensity Accuracy (%)	Frequency Precision (%)	Runtime
CNN + PINN Test	44.2	51.9	8 min 26 sec
	32.1	57.1	8 min 27 sec
	42.4	54.8	8 min 13 sec
	38.2	55.3	15 min 24 sec
	40.2	52.4	16 min 26 sec
CNN + PINN Unseen	81.9	0.2	
	66.0	0.2	
	75.9	0.2	
	80.6	0.2	
	84.1	0.2	

g. <u>CNN + PINN Image</u>

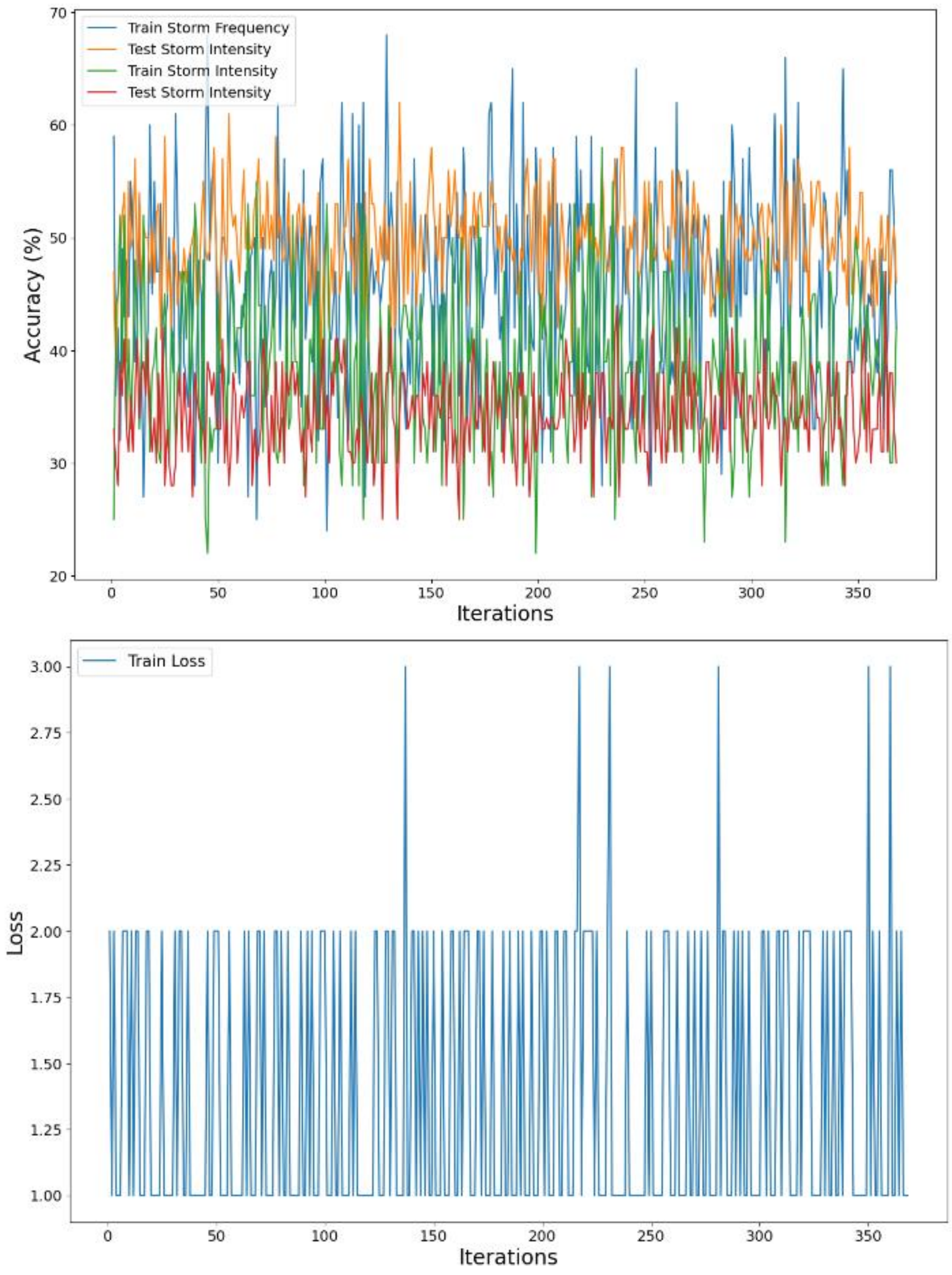




Results from five runs:

Model Name	Storm Intensity Accuracy (%)	Frequency Precision (%)	Runtime
CNN + PINN Test	35.7	43.2	5 min 16 sec
	39.4	41.2	5 min 36 sec
	37.3	52.4	5 min 39 sec
	41.7	29.2	5 min 40 sec
	39.7	37.1	5 min 1 sec
CNN + PINN Unseen	78.4	0.0	
	77.9	0.0	
	74.1	0.5	
	77.9	0.0	
	82.4	0.2	

h. CNN + LSTM + PINN



Results from five runs:

Model Name	Storm Intensity Accuracy (%)	Frequency Precision (%)	Runtime
CNN + LSTM + PINN Test	42.3	47.7	12 min 7 sec
	47.9	36.0	11 min 21 sec
	44.9	51.0	11 min 29 sec

	37.5	45.9	11 min 15 sec
	45.4	64.5	11 min 26 sec
CNN + LSTM + PINN Unseen	88.2	0.4	
	92.1	0.2	
	89.9	0.4	
	72.2	0.3	
	88.4	0.2	

i. <u>XGBoost</u>

Results from five runs:

Model Name	Storm Intensity Accuracy (%)	Frequency Precision (%)	Runtime
XGBoost Test	65.1	99.8	6 min 40 sec
	65.1	99.8	6 min 27 sec
	65.1	99.9	6 min 17 sec
	65.1	99.8	6 min 15 sec
	65.1	99.9	6 min 16 sec
XGBoost Unseen	14.7	0.0	
	14.7	0.0	
	14.7	0.0	
	14.7	0.0	
	14.7	0.0	

j. <u>Random Forest</u>

Results from five runs:

Model Name	Storm Intensity Accuracy (%)	Frequency Precision (%)	Runtime
Random Forest Test	4.6	100.0	4.19 sec
	5.8	100.0	4.26 sec
	5.7	100.0	4.39 sec
	4.7	100.0	4.19 sec
	1.6	100.0	4.29 sec
Random Forest Unseen	4.6	0.0	
	5.8	0.0	
	5.7	0.0	
	4.7	0.0	
	1.6	0.0	