Hurricane Wind Speed Prediction Using Machine Learning and Deep Learning

Arsh Modak, Omkar Waghmare
https://github.com/arshmodak/Hurricane-Wind-Speed-Prediction-using-Deep-Learning-and-Machine-Learning

1. SUMMARY

1.1. Problem Statement and Overview

Tropical cyclones have become more destructive in the last decades due to increase in surface temperature as a result of global warming. Hurricanes/Tropical Cyclones are one of the costliest natural disasters globally because of the wide range of associated hazards. Hurricanes can cause upwards of 1000 deaths in a single event and are responsible for more than 100,000 deaths worldwide. During a tropical cyclone, humanitarian response efforts hinge on accurate risk approximation models that depend on wind speed measurements at different points in time throughout a storm's life cycle.

Direct measurements of the winds within a tropical cyclone are sparse, particularly, over open ocean. Thus, diagnosing the intensity of a tropical cyclone is initially performed using satellite measurements. According to the National Hurricane Center (NHC), an accurate assessment of intensity using satellite data remains a challenge.

For several decades, forecasters have relied on visual pattern recognition of complex cloud features in visible and infrared imagery. However, visual inspection is manual, subjective and often leads to inconsistent estimates. This is the reason why we want to design and develop a system using Deep Learning and Machine Learning which predicts the hurricane's speed using satellite images.

1.2. Data

The data was prepared by the NASA IMPACT team and Radiant Earth Foundation. It consists of single-band satellite images captured by GOES (Geostationary Operational Environmental Satellites) of 600 storms over two oceans (Atlantic Ocean and East Pacific).

There are a total of 114,634, 366x366 single-band images; 70,257 images in the train set and 44,377 in the test set. Each image is associated with the following metadata:

- *Image ID*: unique image identifier.
- *Storm ID*: unique storm identifier.
- *Ocean*: 1-Atlantic, 2-East Pacific.
- *Relative Time*: timestamp in milliseconds relative to the first image.
- *Wind Speed*: speed of the wind in knots.

1.3. Proposed Methods

We will be creating a baseline CNN model and use state-of-the-art (SOTA) CNN architectures for our task. Additionally, we will perform feature extraction techniques to use them for our standard ML models.

We will perform hyperparameter tuning for all our models and evaluate them using metrics such as MSE, MAE and RMSE.

2. PROPOSED PLAN

2.1. Metadata Extraction

Training and testing sets are comprised of two main folders, one for images along with a JSON file for its metadata and another file for its corresponding wind speed labels. We have extracted meta data regarding the images from multiple JSON files and created a dataset for future use.

2.2. Exploratory Data Analysis

We performed EDA on the metadata dataset that we created in the previous step. We also visualized the images of the storms at varying speeds. The results can be viewed in Section 3.

2.3. Image Processing and Feature Extraction

For the baseline model, we will be using the single band images as is. For the SOTA CNN architectures, we will convert our single-band images to RGB (three-band) images by concatenating three images at different consistent timesteps to create one image. Additionally, we will experiment with various filters and different timesteps. We will also resize all our images to 224x224x3 to accommodate for the input layers of the SOTA CNNs.

For the standard ML models, we will extract features using our trained SOTA CNN models as well as prepare a Bag of Visual Words.

We will transform the image using various Data Augmentation techniques such as cropping, resizing, flipping, translating, normalizing etc.

2.4. Model Creation and Training

For our baseline model we will create a simple CNN architecture. We will also experiment with SOTA CNN architectures such as DenseNets, ResNet, VGGs. We will use the extracted features from CNNs for traditional supervised ML models such as Linear Regression, Decision Trees and Random Forest Regressors, etc.

2.5. Hyperparameter Tuning

We will experiment with various model hypermeters and optimization techniques for our models. If needed we will also perform Cross Validation and other regularization techniques.

2.6. Evaluation and Comparison

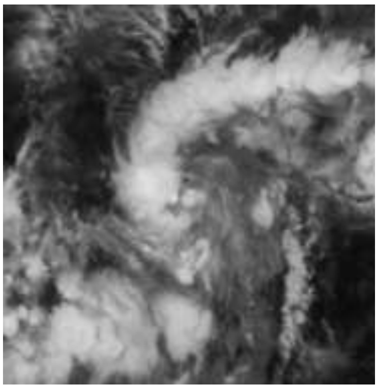
Since our response variable is continuous in nature, we will evaluate our models using metrics such as MSE, MAE and RMSE and compare and reason model performance.

3. PROJECT MILESTONES

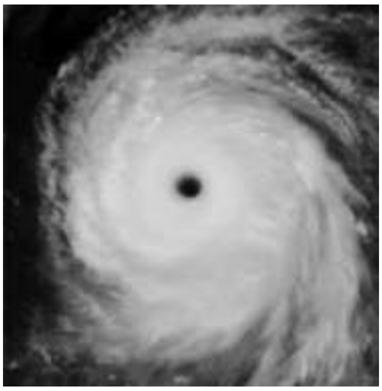
- **October 19:** Completing of preparation of images for baseline model and implementation and initial results of baseline model.
- **November 2:** Hyperparameter Tuning results for baseline model. Completion of preparation of images for pre-trained models (converting single-band images to RGB images based on timestep).
- **November 16:** Completion of pre-trained model training, hyperparameter tuning.
- **November 30:** Feature Extraction using CNNs and Implementation of Machine Learning Models
- **December 14:** Model evaluation and Performance Comparison. Completion of Final Report.

4. PRELIMINARY RESULTS

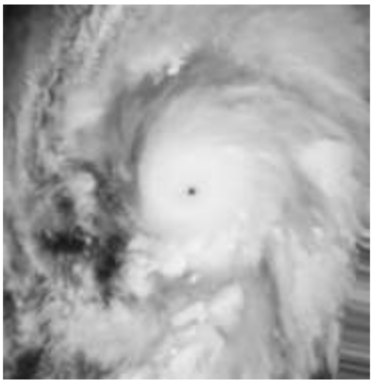
Images of the same storm at different wind speeds



Wind Speed: 30 knots

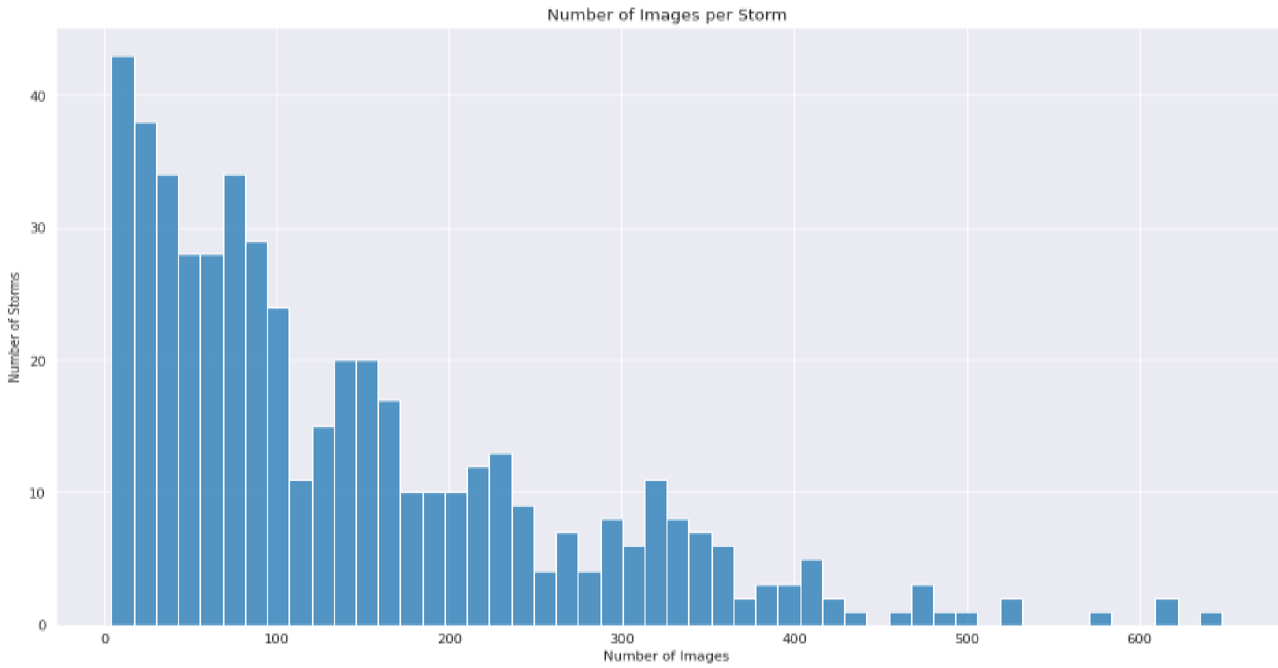


Wind Speed: 150 knots

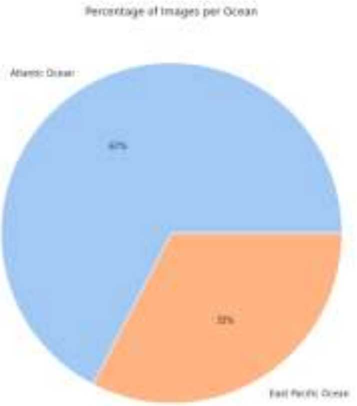


Wind Speed: 185 knots

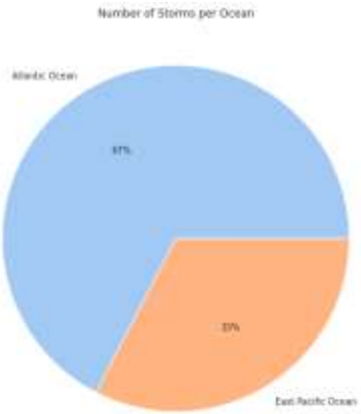
Number of Images for each Storm



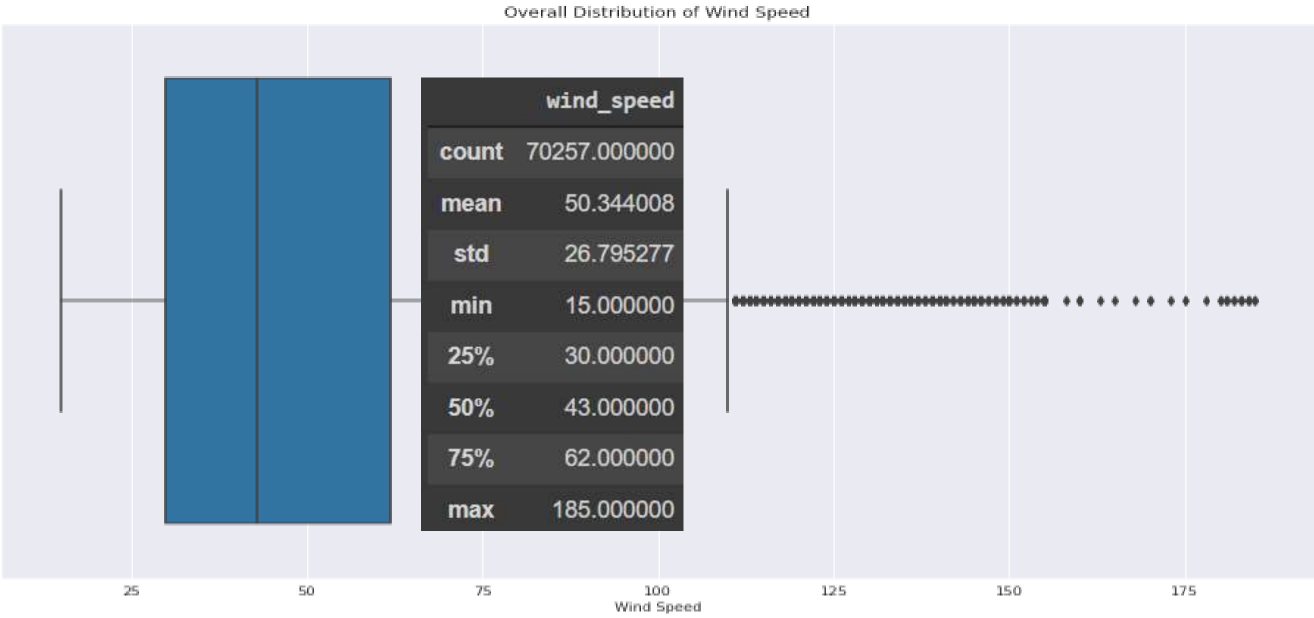
Percent of Images per Ocean



Percent of Storms per Ocean

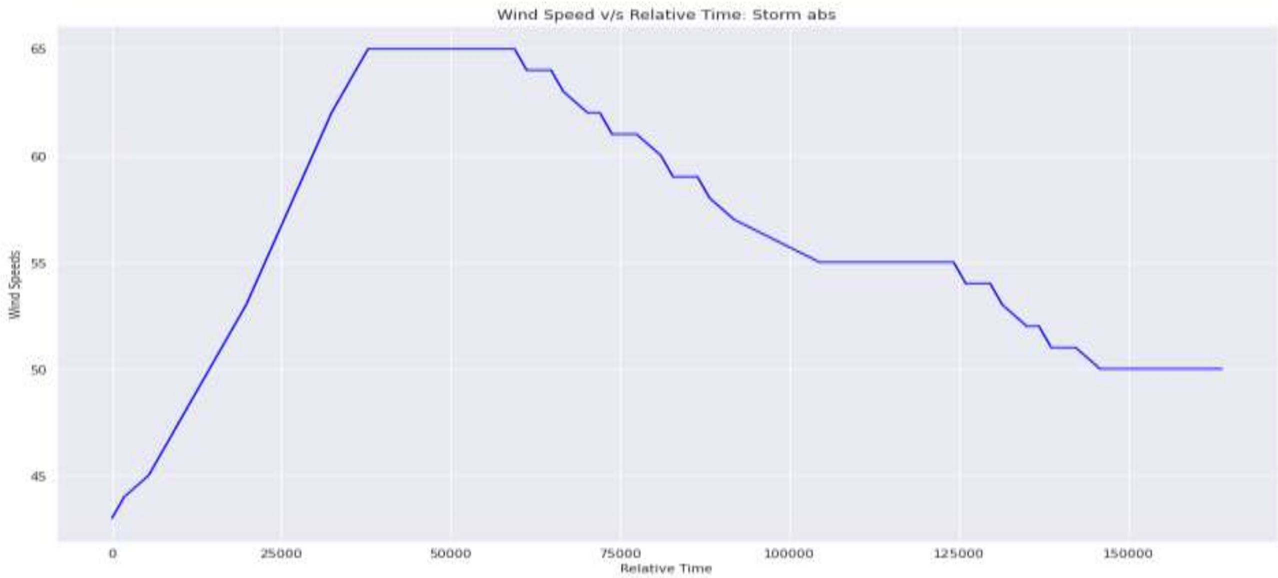


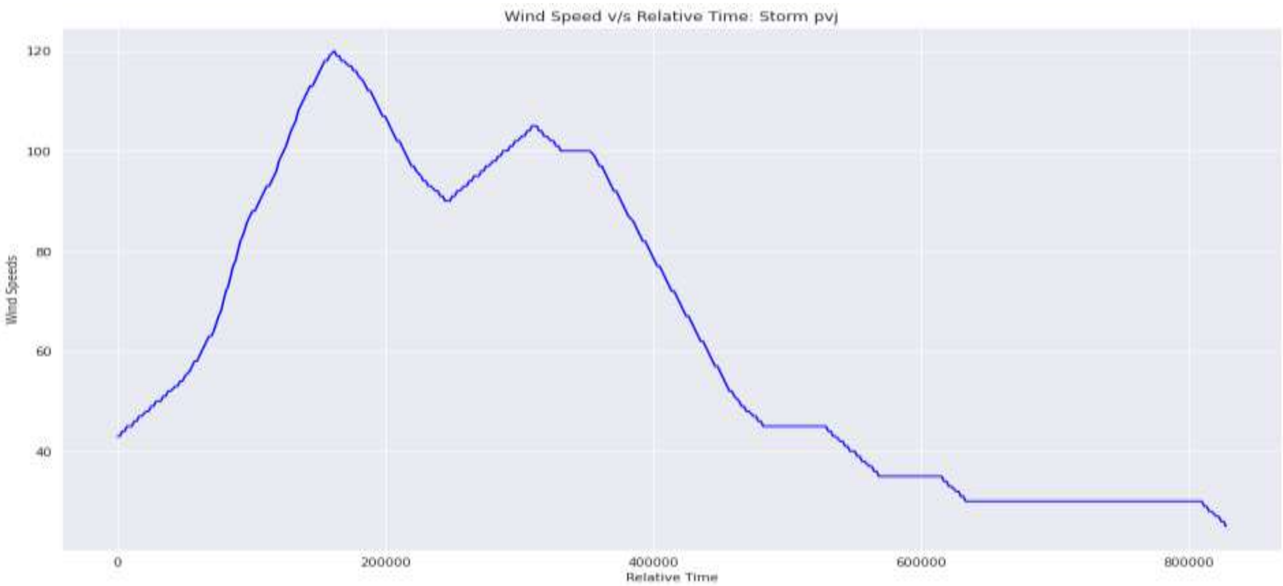
Distribution of Wind Speeds



	wind_speed
count	70257.000000
mean	50.344008
std	26.795277
min	15.000000
25%	30.000000
50%	43.000000
75%	62.000000
max	185.000000

Wind Speed v/s Relative Time per Storm





5. REFERENCES

https://ieeexplore.ieee.org/document/9149719

https://mlhub.earth/10.34911/rdnt.xs53up

https://www.drivendata.org/competitions/72/predict-wind-speeds/