# 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

For phase two, we have converted the data from one channel grayscale images to three channel RGB images by concatenating three images at a fixed interval. To achieve this, we created a new metadata dataframe by grouping the data by "storm_id", sorting it by "relative_time" and creating a new column which contained a list of the paths of the images (code in second image in appendix).

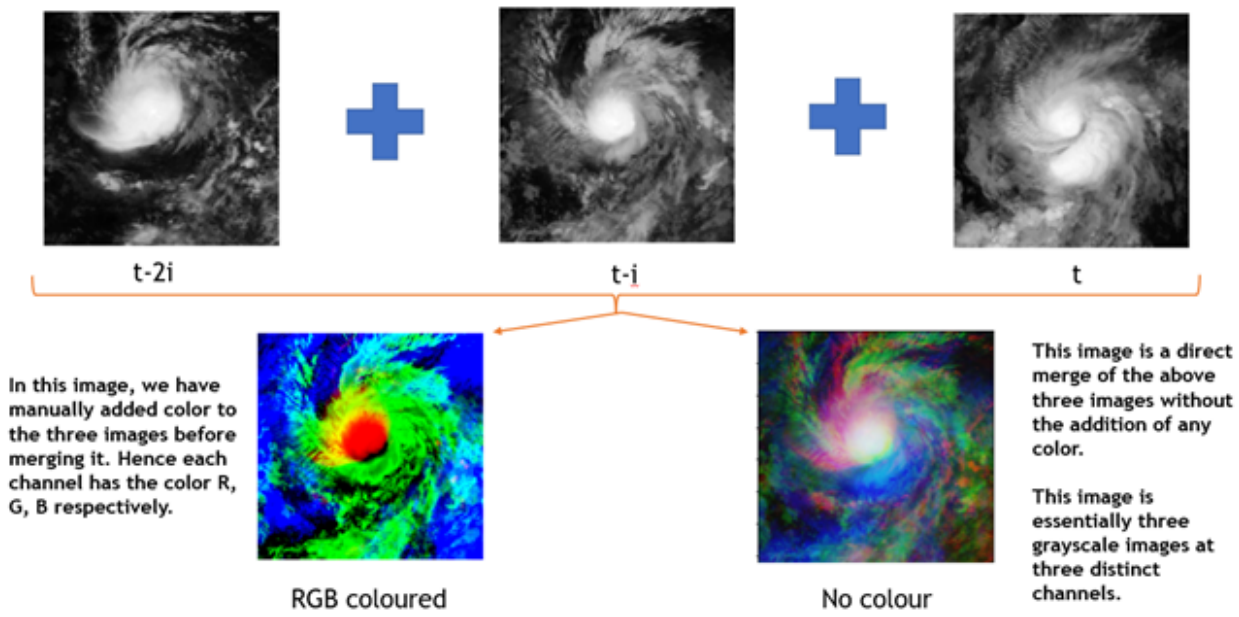



Figure 1: Image Processing results. Original images at different timesteps (top), manually added colors before merging images (bottom left), merging images without adding color (bottom right). For phase two, we will not be using the method on the left since it changes the actual pixel value as we manually add colors.

### 1.3. Proposed Methods

We created a baseline CNN model in phase one that takes simple grayscale images as input and predicts the wind speed. The disadvantage in this case was that the model did not consider historical data/images at previous timestamp of a particular storm and predicted speeds from a single stand alone image.

In phase two, we will attempt to overcome this limitation by using our newly created images discussed in Section 1.2. We are currently training pretrained models and hope to achieve better performance than our baseline CNN model.

Initially, in phase two, we were going to extract features from the best performing CNN and train traditional models for a comparative analysis. However, after careful deliberation, we have decided to drop this idea and pursue a potentially better idea described in detail in Section 2.1 and 2.2. Briefly, we will first use a model to classify an image into a range (bins) of wind speeds and then train another model for each bin which will then predict the precise speed of the hurricane.

## 2. PROPOSED PLAN

### 2.1. New Metadata Creation

We have already created a new metadata for our three channel RGB images. We will segregate our data by wind speed and assign each image to its corresponding range of the bins we create. We will then label encode these bins to make it ready for classification. To balance our dataset, we will make sure each bin has enough images. To achieve this, the bins for higher speeds would be larger than lower speeds, i.e., lower speed: 10 to 20, 20 to 30 etc, and higher speeds: 70 to 90, 90 to 130, 130 to 180 etc. This is required due to there being very few images for extremely high speeds of hurricanes.

### 2.2. Model Creation and Training

We have started the training of SOTA CNN (densenet161, vgg16, resnet101) architectures to identify the best performing models which will then be used for the architecture described below.

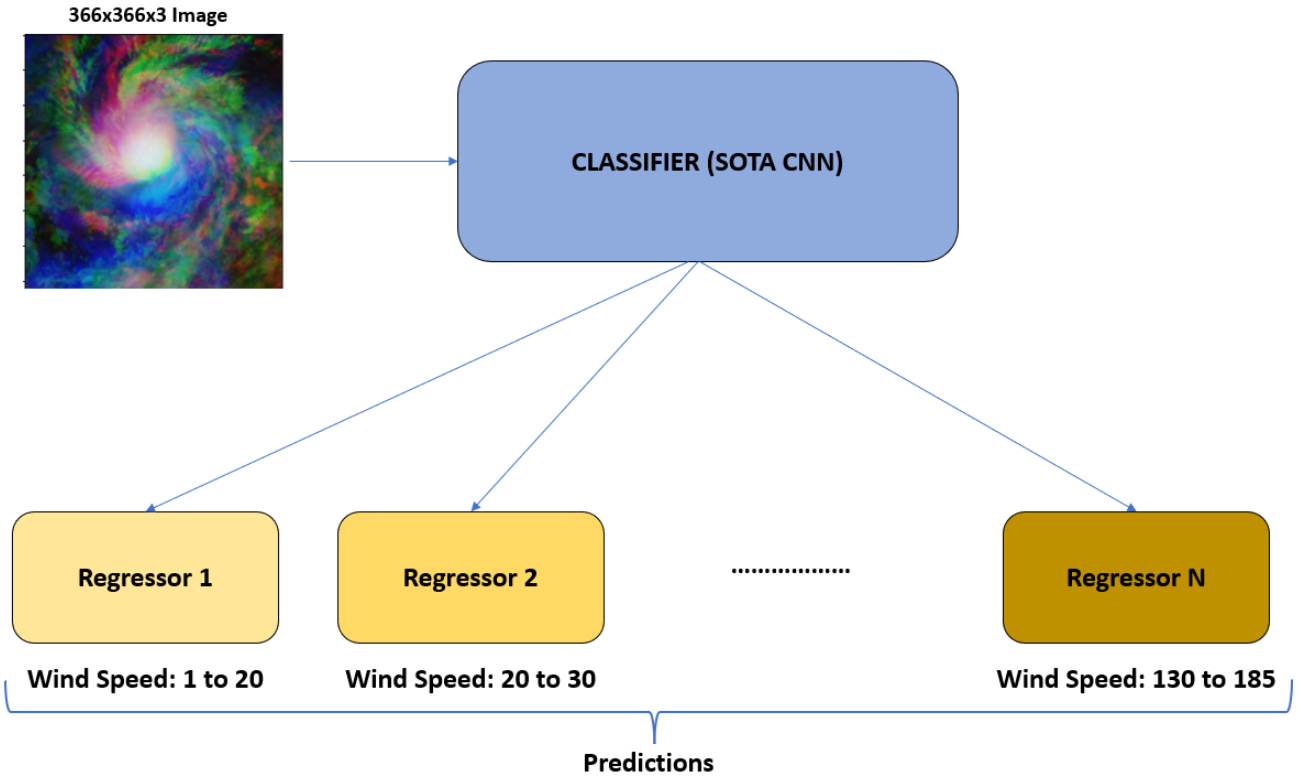



Figure 2: Describes the proposed architecture for phase two. A classifier will be responsible for classifying each RGB image into a specific bin. Each bin will have it's own regressor trained to predict wind speed within its range, thereby giving better predictions for storms with extremely high or extremely low wind speeds.

### 2.3. Hyperparameter Tuning

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

### 2.4. Evaluation and Comparison

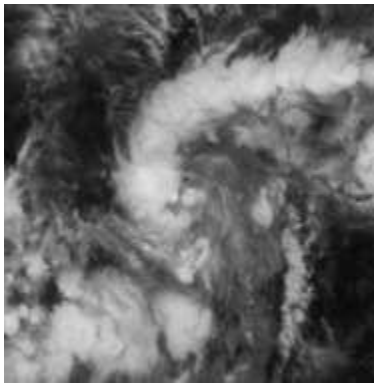
Since our response variable is continuous in nature, we will evaluate our models using metrics such as MSE, MAE and RMSE and compare and reason model performance.

## 3. PROJECT MILESTONES

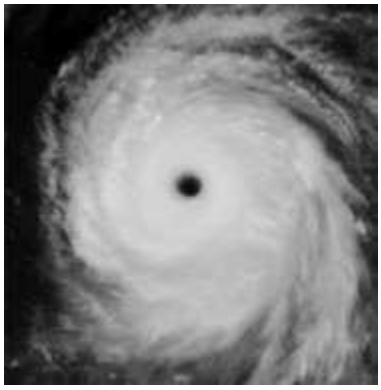
- **October 19:** Completing preparation of images for baseline model and implementation and initial results of baseline model. **(Complete)**
- **November 2:** Hyperparameter Tuning results for baseline model. Completion of preparation of images for pre-trained models (converting single-band images to RGB images based on timestep). **(Complete)**
- **November 16:** Creation of new metadata and data loader for phase two **(Complete)**, training and evaluation of pre-trained models, hyperparameter tuning. **(In-progress)**
- **November 30:** Training classifier and constructing a pipeline for the proposed architecture.
- **December 14:** Model evaluation and Performance Comparison. Completion of Final Report.

## 4. PRELIMINARY RESULTS

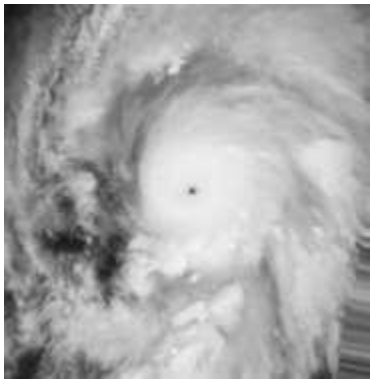
Images of the same storm at different wind speeds



Wind Speed: 30 knots



Wind Speed: 150 knots



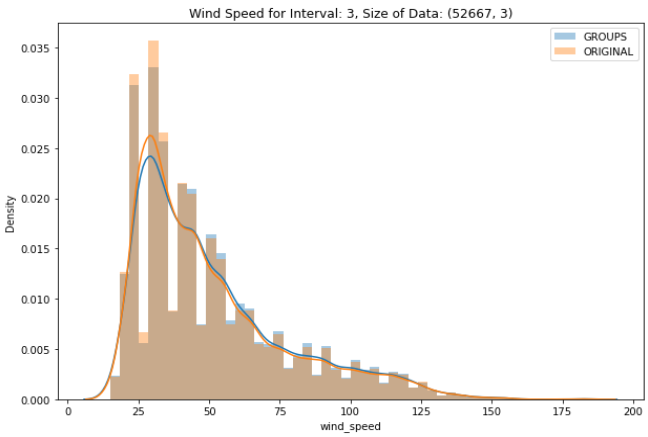
Wind Speed: 185 knots

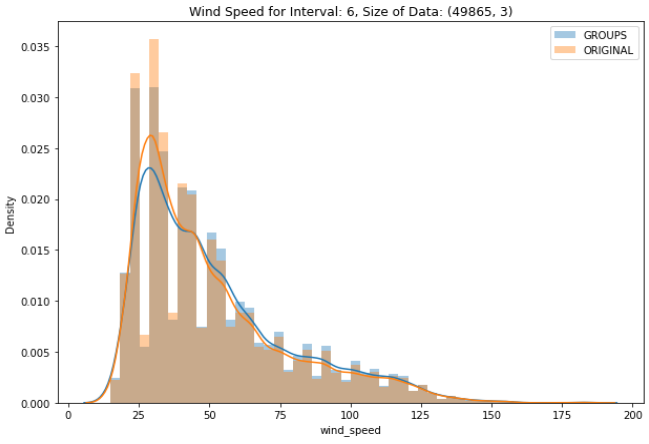
From the data, we could see that we have fewer images for storms with extremely lower speeds (0 -20 Knots) and extremely higher speeds (150 - 185 Knots), and hence the models were not able to accurately predict speeds for such storms.

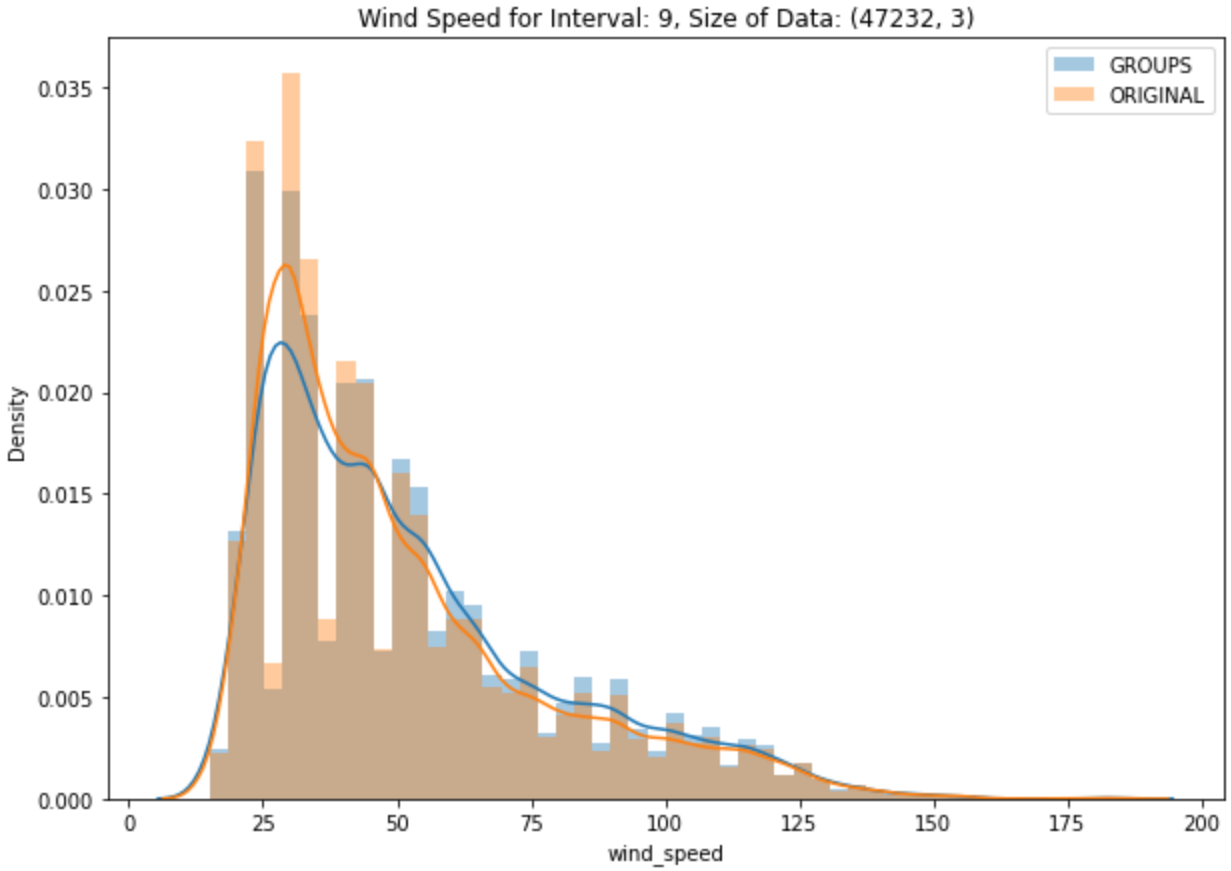
The code below takes a dataframe and interval as inputs and returns a new dataframe that contains the list of the three image paths at the specified intervals which will be used to create the new RGB images for training the SOTA CNN Architectures.

```
def get_data_in_intervals(metadata, interval):

  data_dict = dict()
  counter = 0
  for name, group in metadata.groupby(by="storm_id"):
    group = group.sort_values(by="relative_time").reset_index()
    # print(name, group.index)
    for i in range(0, group.shape[0]):
      i2 = i-interval
      i3 = i2-interval
      if i3 > 0 and i2 > 0:
        # print(i, i2, i3)
        lis = [group.image_path[i3], group.image_path[i2], group.image_path[i]]
        # storm_id = metadata.storm_id[i3]
        wind_speed = group.wind_speed[i]

        if counter not in data_dict.keys():
          data_dict[counter] = {"storm_id": name, "image_list": lis, "wind_speed": wind_speed}
        counter += 1

  df = pd.DataFrame.from_dict(data_dict, orient="index").reset_index().drop(columns="index")

  return df
```

The image below compares the distribution of the original data with the new data created using the above code snippet for 3,6,9 intervals respectively. As we can see, the new data distribution is not varying much from the original data, thereby maintaining the relative number of images for each wind speed.






## 5. REFERENCES


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

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

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