

Deepti: Deep Learning-based Tropical Cyclone Intensity Estimation System

Manil Maskey, *Member, IEEE*, Rahul Ramachandran, *Member, IEEE*, Muthukumaran Ramasubramanian, Iksha Gurung, Brian Freitag, Aaron Kaulfus, Drew Bollinger, Dan Cecil, JJ Miller

Abstract—Tropical cyclones are one of the costliest natural disasters globally because of the wide range of associated hazards. Thus, an accurate diagnostic model for tropical cyclone intensity can save lives and property. There are a number of existing techniques and approaches that diagnose tropical cyclone wind speed using satellite data at a given time with varying success. This paper presents a deep learning-based objective, diagnostic estimate of tropical cyclone intensity from infrared satellite imagery with 13.24 kt Root Mean Squared Error (RMSE). In addition, a visualization portal in a production system is presented that displays deep learning output and contextual information for end users, one of the first of its kind.

Index Terms—Deep learning, wind speed estimation, research to production, machine learning lifecycle, model interpretation.

I. INTRODUCTION

HURRICANES can cause upwards of 1000 deaths in a single event and are responsible for more than 100,000 deaths worldwide [1]. Economical losses from several hurricanes over the last few years is estimated to be more than \$50 billion and accounting for inflation, some historical storms caused more than \$100 billion in damage [1]. The Insurance Institute reports there is more than \$10 trillion in insured coastal property that is vulnerable to landfalling hurricanes in the U.S. alone. There are a number of hazards associated with hurricanes and many factors that can contribute to the final death toll and damage losses. Hurricane damage models approximate risk using an exponential [2] or power [3] of wind speed. Thus, being able to accurately diagnose the intensity of a tropical cyclone (defined as the maximum sustained surface wind speed) is essential for disaster preparedness and response.

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. An accurate assessment of intensity using satellite data remains a challenge as the National Hurricane Center (NHC) estimates a 10-20% uncertainty in its post-analysis intensity estimates when only satellite observations are available. Similarly, the average 24-h forecast intensity error for any given storm is on the order of 10-20%, as reported by the NHC. Thus, improving initial intensity estimates from satellite imagery could mean significant improvements in short-term intensity forecasting; thereby improve our nation's disaster readiness and response.

Operational forecasters primarily rely on adaptations of Dvorak technique, a satellite image-based technique developed in the 1970s [4], [5], [6], [7], [8] for estimating tropical cyclone intensity when direct measurements from aircraft are not available. These techniques utilize brightness temperatures from infrared satellite measurements and employ human visual inspection of features with the cloud field, such as symmetry of the inner core and curvature of the outer rainbands. There are known limitations with Dvorak-based techniques, namely, the subjectivity of humans in classifying storm intensity based on cloud features. Two well-trained analysts using this technique can derive different intensity estimates for the same storm at the same time. Even if they agree with each other to within 0.5 Tropical Number (or T-Number; the intensity classification scale utilized by the Dvorak technique), the precision of Dvorak-based estimates, there can be a 12 knot (hereby abbreviated as kts) or approximately 6 ms^{-1} difference in wind speed estimates for hurricane-forced storms. Furthermore, based on anecdotal evidence, large differences are sometimes seen in cases with complex cloud structures, and when comparing estimates from different agencies. One such example of a complex case is given by the NHC advisory on Tropical Storm Ophelia, October 10, 2017 UTC 1500:

“Dvorak intensity estimates range from T2.3/33 kts from University of Wisconsin - Cooperative Institute for Meteorological Satellite Studies (UW-CIMSS) to T3.0/45 kts from the Tropical Analysis and Forecast Branch (TAFB) to T4.0/65 kts from the National Oceanic and Atmospheric Administration/National Environmental Satellite, Data, and Information Service Satellite Analysis Branch (SAB). For now, the initial intensity will remain at 45 kts, which is an average of the scatterometer winds and all of the other available intensity estimates.”

Two human experts at TAFB and SAB differed by 20 kts in their Dvorak analyses, and the automated version at the University of Wisconsin was 12 kts lower than either of them. Passive microwave imagery from modern satellites can reveal structural details that can mitigate some of the difficulties associated with the Dvorak technique, but such imagery is not available at a temporal frequency required by operational forecasters. This need was noted by the hurricane science community and is prominently highlighted at the 2017 Tropical Cyclone Operations and Research Forum/71st Interdepartmental Hurricane Conference with a focused session especially dedicated to “Techniques to improve the utility of microwave satellite and radar data for tropical cyclone intensity and location analysis (e.g., a ‘Dvorak-like’ technique

M. Maskey is with the National Aeronautics and Space Administration, Marshall Space Flight Center, Huntsville, AL, 35805 USA e-mail: manil.maskey@nasa.gov.

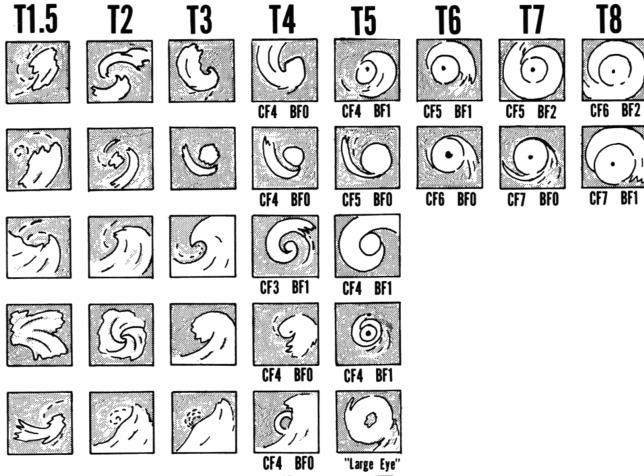


Fig. 1: Dvorak's illustration of common development patterns and corresponding intensities [5]

using microwave imagery)." Thus, there is clearly a vital need to develop automated, objective, accurate tropical cyclone intensity estimation tool from satellite data. To take advantage of such an automated tool, a full system that can be deployed and operate on all storms in real-time is required. Such a system would have a profound impact on disaster preparedness.

The main contributions presented in this paper are: 1) the development of deep learning model for objective estimation of tropical cyclone intensity using a convolutional neural network (CNN) on satellite images; 2) extensive evaluation of the deep learning model and analysis of how the model makes a decision for classification, 3) the systematic process of transitioning the model to a production environment; 4) the design and implementation of a tropical cyclone monitoring and intensity estimation system; 5) interactive portal for situational awareness and evaluation of wind speed estimation; and 6) labeled training images of hurricanes with maximum wind speed.

The rest of paper is organized as follows: Section 2 provides related work. Section 3 outlines our methodology and Section 4 concludes the paper with possible future work.

II. RELATED WORK

Some related work in estimation of wind speed from satellite images are discussed here along with existing portals that provides real-time information of tropical cyclones. We also provide background information on convolutional neural network, which is the basis of our deep learning model.

A. Dvorak Technique

The Dvorak technique is a satellite image-based tropical cyclone intensity estimation method that has been influential for more than 30 years. With modifications and improvements of this method over the years, including automated versions [9], variations of the Dvorak technique are considered the gold standard for satellite image-based tropical cyclone intensity estimation among tropical meteorologists. The main premise

of the Dvorak technique is that the shape and coverage of the cloud field is related to the intensity of the cyclone. Features such as the length and curvature of the storm's outer rainbands are analyzed to arrive at a particular T-number as shown in Figure 1. Given this visual analysis, certain objective rules are applied based on prior intensity estimates that are used to determine the current intensity. The longevity of the Dvorak technique within the tropical community serves as proof that there is a physical relationship between spatial cloud patterns and tropical cyclone intensity. That said, because the Dvorak technique relies on human interpretation of features in a tropical cyclone cloud field, two well-trained analysts can assign different intensity estimates. Additionally, subtle differences in T-number can result in differences in maximum wind speed by 12 knots or more at hurricane intensities. Furthermore, larger differences are sometimes seen in storms with complex cloud structure, and when comparing estimates from different agencies in different centers.

B. Advanced Dvorak Technique

Olander and Velden [33], [34] built upon the manual Dvorak technique and introduced the Advanced Dvorak Technique which uses satellite imagery as input to an automated algorithm that assigns a T-number. Additional improvements have been incorporated into the ADT such as the inclusion of aircraft measurements and passive microwave data as well as improved tropical cyclone centering; a critical component of the automated algorithm. While the ADT improves upon the manual Dvorak technique, model performance struggles on weaker storms that tend to have a more disorganized cloud distribution and empirical thresholds are retained to constrain the change in cyclone intensity with time.

C. Deviation-Angle Variance Technique (DAVT)

The deviation-angle variance technique (DAVT) was described and applied to the North Atlantic by Pineros et al. [11], [10] and Ritchie et al. [12] and in North Pacific by Ritchie et al. [13]. This technique quantifies the symmetry of tropical cyclones in infrared (IR) satellite imagery. It performs a directional gradient statistical analysis of the brightness of those images. The level of alignment of the gradient vector is the amount of deviation from the perfect radial axis. The variance of this deviation angle is used in the quantification of the cyclone. This technique has two main limitations: (i) it requires images with properly marked tropical cyclone centers and (ii) it uses different models and fitting parameters for tropical cyclones in different regions, reducing broader, global application.

D. Passive Microwave Imagery-based Tropical Cyclone Intensity Estimation

Passive microwave satellite measurements provide more detailed information on the inner structure of the cyclone as microwaves are able to penetrate clouds revealing precipitation within the storm (e.g., [14]). Initial attempts to estimate tropical cyclone intensity directly from passive microwave

imagery [15], [16] were not as successful as initial infrared imagery; however, [34] have incorporated microwave data into the ADT which has reduced estimation errors. Microwave information in the ADT is only applied during periods of storm intensification and when the cyclone has no visible eye. One of the primary limitations of using passive microwave measurements for operational diagnosis of cyclone intensity is the limited temporal frequency of observations compared to infrared geostationary satellite measurements.

E. Convolutional Neural Network

Learning “features” is a core idea behind deep learning, which consists of machine learning, computer vision, and pattern recognition algorithms that have multiple layers, where each layer performs feature detection. The fundamental issue in image classification is bridging the semantic gap of using low-level features (image pixels) to derive high-level abstractions. Deep learning’s hierarchical layered learning approach attempts to bridge that semantic gap. Recent advancements in deep learning techniques have produced state-of-the-art image classification results in many domains [25] mostly using convolutional neural networks (CNN).

Convolutional neural networks (CNNs) have been used for many different computer vision tasks ranging from image classification [17], [18], [19] to object detection [20] and even visual saliency detection [23]. Each task uses a slightly different network layout depending on the objective. However, their basic components are all very similar, consisting of convolutional layers, pooling layers, and fully connected layers.

The primary function of the convolutional layer is to learn feature representations of the inputs, where a feature can be viewed as any type of input pattern (e.g. a cloud band in a satellite image). Convolution is the main operation in this layer where the weights of the convolution filters are learned during each iteration of the network. Feature maps are obtained by convolving the input with a convolution filter and then applying an element-wise non-linear activation function on the results. The activation function, typically the rectified linear unit (ReLU) [22], introduces nonlinearities to the network.

The pooling layer is generally placed in between convolutional layers. It is used to reduce the spatial size of the representation, which in turn reduces the number of parameters and computations in the network. The pooling layer operates independently on every input (feature map) and resizes the input spatially using average pooling or max pooling [24]. By stacking more and more convolutional and pooling layers, the network can potentially generate progressively more abstract features at higher layers of representations [25].

The fully connected layer is typically located after multiple convolutional and pooling layers and is just like a typical neural network with full connections to all activations in the previous layer. The purpose of the fully connected layer is to perform the high-level reasoning and classification, typically using a softmax classifier. Stochastic gradient descent (SGD) [27] is the most common back propagation algorithm used to update the parameters during training. Each parameter update in SGD is computed with respect to a mini batch to

help reduce variance in the parameter update and to lead to a more stable convergence. The learning rate controls the speed at which convergence occurs.

Our previous work [28] has demonstrated that a CNN can be used successfully to estimate tropical cyclone categories. Our results show improvement using CNNs when compared to existing techniques. We used up to 5 kt interval estimation of the wind speed with the data available from the Naval Research Laboratory. We extend this past work by including new real-time Geostationary Operational Environmental Satellite (GOES) imagery, further enhancing the model to estimate the wind speed at 1 kt interval, performing extensive evaluation and building and deploying a production system.

Other research efforts have also used CNNs for hurricane intensity estimation application. Wimmers et al. [36] used CNNs to estimate cyclone intensity using passive microwave imagery (37, 89 GHz bands) as input to the model. The approach uses images from years 1987 to 2012. The wind speed labels are obtained from best track records in the hurricane database (HURDAT2) and joint typhoon warning center in 5 kt intervals. Linear interpolation is then used to get wind speed estimates for training images. The model achieves RMSE of 14.3 in the test set obtained from years 2007 and 2012. Lee et al. [38] used 3-dimensional CNNs to estimate hurricane intensity. The input to their proposed model is satellite imagery from Communication, Ocean and Meteorological Satellite - Meteorological Imager (COMS-MI) and joint typhoon warning center (JTWC) advisories are used to obtain wind speed labels. The research is mainly conducted on Western Northern Pacific (WNP) region. The training data is over sampled through wind speed interpolation and image rotation to have a balanced distribution of images.

Giffard-Roisin et al. [37] uses track data and 3D reanalysis data as input to CNN, along with other features such as location information and maximal sustained windspeed to develop storm track models. They formulate the tracking problem as estimating the displacement between current location and future location of cyclone. Other variations of CNNs (CNN-Long Short Term Memory (ConvLSTM), U-Net) have also been successfully used in tracking and forecasting climate events such as hurricanes [39], [32], [35]. The model presented in this paper does not consider storm track or forecasts and is purely a diagnostic model for estimating tropical cyclone intensity.

III. METHODOLOGY

Development and deployment of the diagnostic tropical cyclone intensity estimation system follows the end-to-end machine learning lifecycle described by Maskey et al. [29]. The machine learning lifecycle is a systematic iterative process of training, testing, and deploying a model to develop an optimized model ready for ingestion into a production system and consumption by the targeted end-users.

A. Machine Learning Lifecycle

The machine learning lifecycle mainly consists of four phases: problem definition, data collection and analysis, model

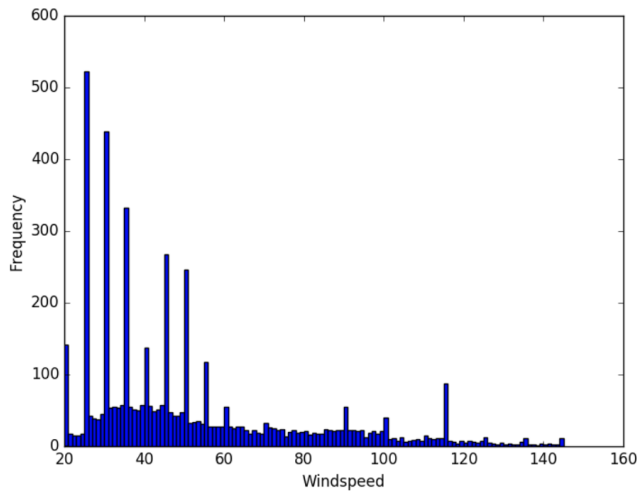


Fig. 2: Distribution of number of images with specific wind speed in knots

development and evaluation, and deployment to production system. Many machine learning projects end after demonstration of model's improvement in accuracy compared to the state of the art. Thus, the steps of the machine learning lifecycle usually omitted are the robust evaluation in understanding of how the model makes decisions and the deployment of the model in a production environment where the model can be evaluated with new real-time data. In this paper, we use the full machine learning lifecycle as a road map to develop our production system.

1) *Problem Definition*: The problem relates to objectively estimating tropical cyclone wind speed using satellite images. During a tropical cyclone, the main factors that contribute to the resulting death toll or damage amount are often attributed to the cyclone's wind speed. Thus, being able to accurately estimate tropical cyclone intensity (wind speed) is essential for disaster preparedness and response. There are inherent issues with current techniques used to estimate tropical cyclone wind speed because of the subjectivity of the Dvorak technique and the empirical thresholds used to constrain the change in cyclone intensity as a function of time. In addition, current techniques are run at hourly temporal frequency or lower and thus, high frequency intensity estimates between official forecast advisories would also provide tropical forecasters with another data point when assessing the maximum wind speed of the cyclone.

2) *Data Collection and Analysis*: The initial training image dataset for this application was constructed using tropical cyclone satellite images from the tropical cyclone repository of the Marine Meteorology Division of U.S. Naval Research Laboratory (NRL) (<http://www.nrlmry.navy.mil>). These satellite infrared (IR) images are captured every 15 minutes and contain additional information such as year, date, time and name of the hurricane.

After several iterations of the beginning phases of the machine learning lifecycle, we realized that NRL image database was not sufficient and more samples were required at a higher temporal frequency. However, real-time image generation from

the NRL data was not possible. Thus, we transitioned to raw GOES data available from NOAA's Comprehensive Large Array-data Stewardship System (CLASS) [30] and wind speed information from HURDAT2, the tropical cyclone best track reanalysis data [40]. HURDAT2 is a storm database that provides various characteristics of any tropical storm. We utilize the location, time and wind speed features from the database. We use storms from year 2000 through 2019, out of which storms from year 2000 - 2016 are used for training and storms from 2017 are used for validation to avoid any intra-storm bias. Storms from 2018 and 2019 were used for model testing. The model was then used in production system to estimate wind speeds 2019 storms.

The training dataset for the intensity estimation model is generated using the following steps:

- Identify HURDAT2 storm intensity, time, and location (latitude, longitude of storm center)
- Create a bounding box around storm using start and end date time of the storm
- Use the bounding box and time to download GOES-8, GOES-10, GOES-11, GOES-12, GOES-13, GOES-15, and GOES-16 IR channel (band 4 GOES-8 through GOES-15 and band 13 for GOES-16) data from NOAA CLASS data catalog
- Create a padding of ± 5 degrees from the center of the storm on both latitude and longitude for every file available through NOAA CLASS
- Match HURDAT2 wind speed to the closest file if there is not an exact match in time
- Interpolate location information and wind speed (1kt interval) between consecutive HURDAT2 observations
- Apply random rotation, random shear, random zoom on training data to create more training samples (only for training set)

Any image with more than 70% missing data were removed from the training set. Note: training set included all wind speed data and images from the East Pacific and Atlantic from year 2000 to 2016. The validation set included all wind speed data and images from East Pacific and Atlantic for year 2017. The test set included all wind speed data and images from East Pacific and Atlantic for year 2018 and 2019. The resulting sets included 97152 training samples, 4840 validation samples, and 6118 test samples. Figure 2 shows the number of tropical cyclone images used for training at 1 kt wind speed interval from 20 kts to 140 kts.

3) *Model Development and Evaluation*: Using this training dataset, we design a deep learning model for the objective estimation of tropical cyclone intensity using a convolutional neural network (CNN) on satellite images. Development of the CNN required several iterations through the machine learning lifecycle to arrive at the final model configuration. Since the work of Pradhan et. 2018 [28], further model iterations revealed that image classification to Saffir-Simpson scale storm intensity and 5 kt wind speed intervals did not perform as well as linear output at 1 kt wind speeds. Thus, this model inputs training samples at 5 kt speed intervals and outputs a maximum wind speed at 1 kt resolution; however, model precision cannot exceed the 5 kt resolution of the input training

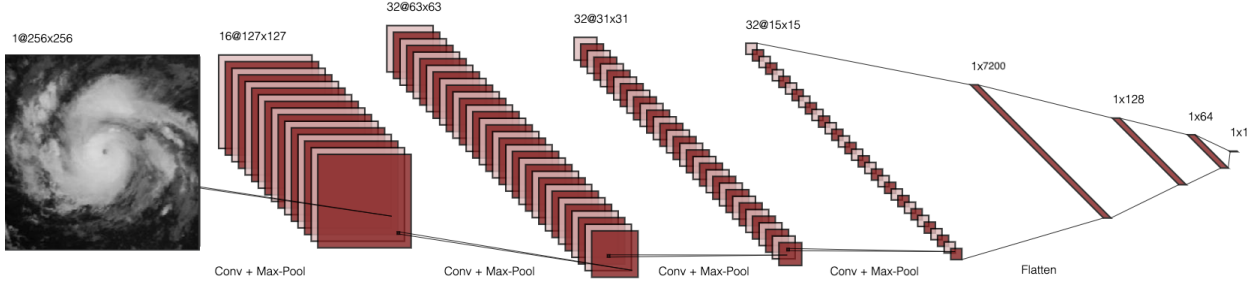


Fig. 3: CNN model architecture used for wind speed estimation

data. Overall, the linear model RMSE was 13.62 kts for all storms occurring in 2018 and 2019 in both the Atlantic Ocean and Eastern Pacific Ocean basins (Table I).

4) *Performance Metrics*: The metrics used in Table I for evaluating the model performance are listed as follows:

- Mean Absolute Error (MAE)

$$\frac{1}{n} * \sum |X_p - X_t| \quad (1)$$

- Root Mean Squared Error (RMSE)

$$\sqrt{\frac{1}{n} * \sum (X_p - X_t)^2} \quad (2)$$

- Bias

$$\frac{1}{n} * \sum (X_p - X_t) \quad (3)$$

- Relative Root Mean Squared Error (RRMSE)

$$\frac{\sqrt{\frac{\sum (X_p - X_t)^2}{n - 1}}}{\bar{X}_p} \quad (4)$$

Where X_p is predicted intensity value and X_t is actual intensity value. n denotes the number of samples.

The final CNN model architecture is shown in Figure 3. A mean squared error (MSE) loss function is used by computing error between the actual wind speed and the estimated wind speed. Thus, the output is treated as a regression output rather than a classification output. The architecture we chose is a variation of VGG-16 model [21]. It differs from VGG-16 model in that VGG-16 model uses 13 convolutional layers and max-pooling layer between each two or three consecutive layers and our CNN model includes 4 convolutional layers where each layer is followed by a max-pooling layer. This was done heuristically to reduce model complexity and avoid model overfitting. Our model also includes 4 dense layers including the output layer which uses a linear activation. Model hyperparameters include: learning rate of $1e-5$, batch size of 60, ReLU activation function, pooling with overlaps, and Adaptive Moment Estimation (Adam) optimization. The model was trained on 12GB Nvidia p100-PCIE Tesla GPU

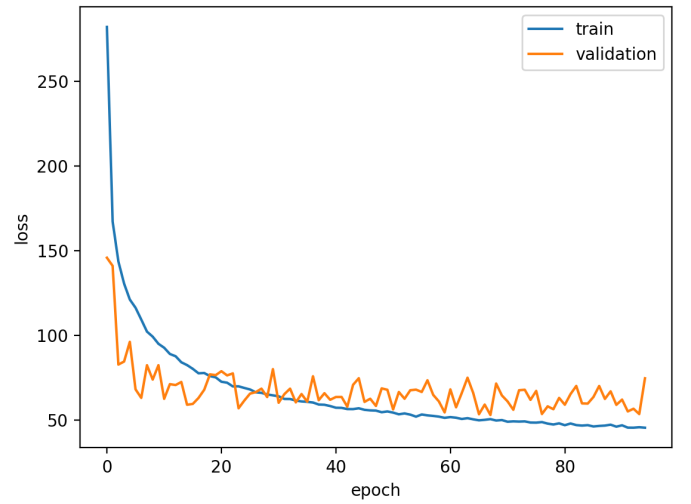


Fig. 4: Training accuracy: model loss mean squared error

with Keras 2.0.8, Tensorflow 1.2.1 and CUDA-8. Eventually, the model reached the RMSE of 10.16 kts for 2017 cyclone data (validation set) as shown in Figure 4.

One of the major criticisms of machine learning techniques from subject matter experts is that it is difficult to determine what the model is actually learning to make classification decisions. Thus, there is a trust factor between the machine learning experts and the physical science community that hinders the transition of these models into a production system. As part of our model evaluation, we use techniques to understand how cyclone intensities are determined within the CNN. Our approach includes tracing the CNN's final intensity back to the original image to discover which pixels contributed most to the classification by using class activation maps (CAMs) [31]. CAMs show which parts of the image the machine learning model uses to determine the maximum wind speed. Encouragingly, the features (and location of pixels) that contribute most to the model output are also present in the cloud patterns used within the Dvorak technique.

Figure 5 shows CAMs for Hurricane Florence in 2018 at various intensity categories. In the Figure 5, the images on the left are the GOES-IR images used as input to the

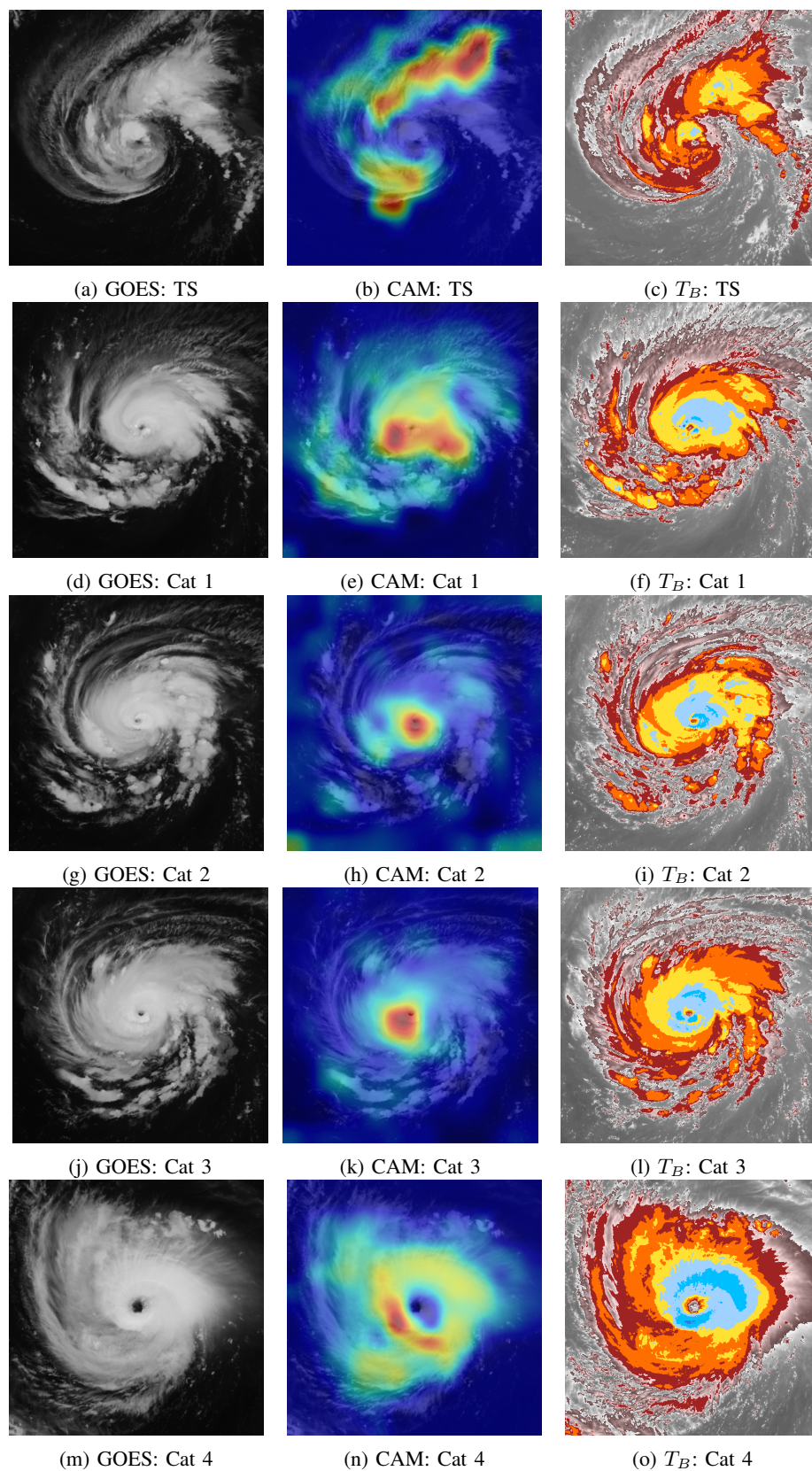


Fig. 5: Hurricane Florence evaluation samples at different categories

TABLE I: Model performance metrics on test dataset (2018 + 2019)

Region	MAE	RMSE	Bias	RRMSE	Support
Atlantic	9.88	13.22	-4.87	0.24	3070
East Pacific	9.95	14.01	-2.70	0.29	3048
Atlantic + East Pacific	9.92	13.62	-3.79	0.27	6118

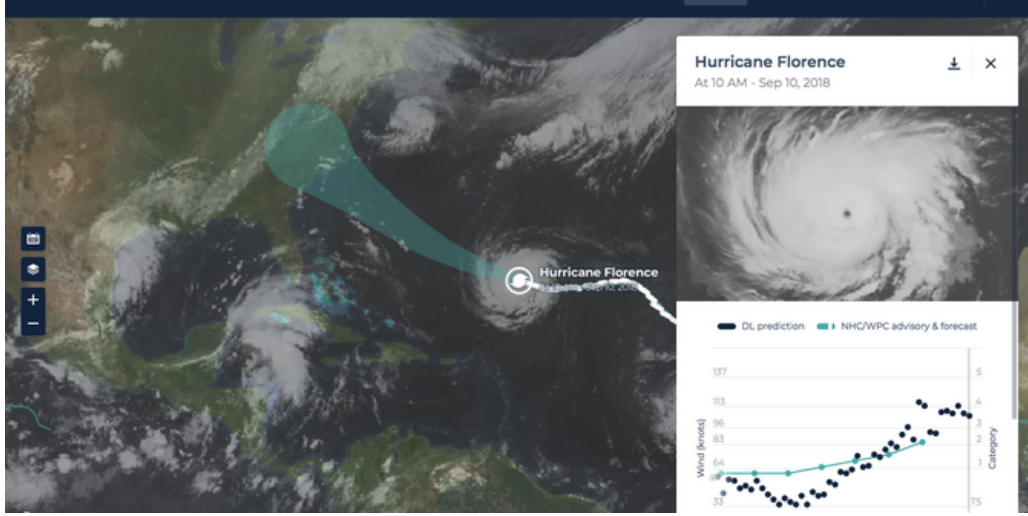


Fig. 6: Hurricane intensity estimation portal

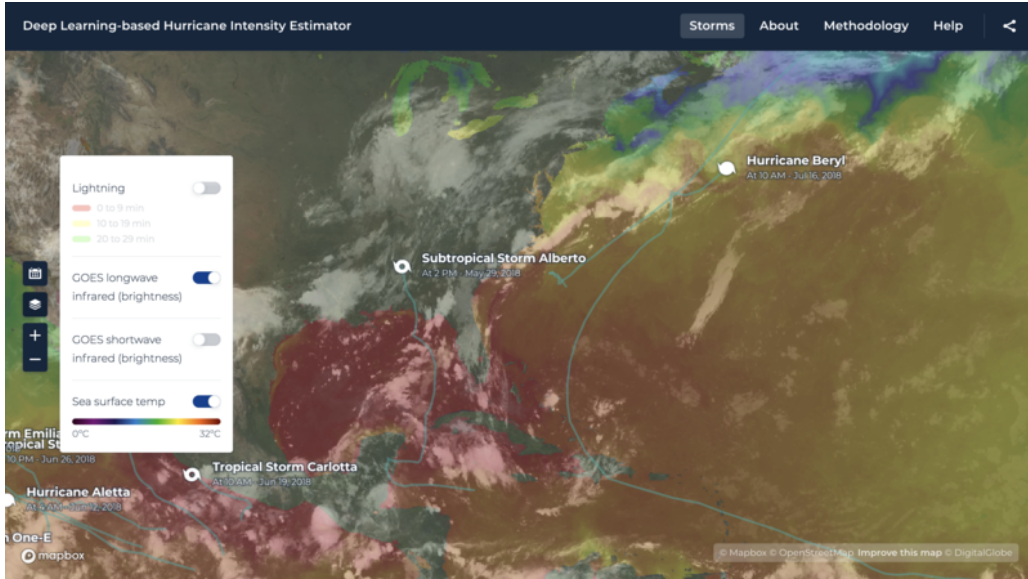


Fig. 7: Sea surface temperature layer

model; the images in the center column are the corresponding CAM images; and the images on the right are the corresponding Dvorak enhanced brightness temperature. The Dvorak enhanced brightness temperature is a color enhancement applied to GOES IR imagery that enhances the contrast of the brightness temperature measurements so the manual Dvorak technique can be more easily applied. Ideally, a perfect model would produce CAMs that match cloud patterns used for classification in the manual Dvorak technique and the Dvorak enhanced brightness temperatures at a given storm intensity. Looking from the top of the image down to the bottom, the storm increases in intensity from a Tropical Storm up to a

Category 4 Hurricane. A comparison between the CAMs, the Dvorak enhanced brightness temperatures and Figure 1, shows that as the storm intensifies, the important features for classification transition from the outer bands to the inner core and the central dense overcast region. For lower T-numbers, the cloud structure is disorganized, which makes it difficult for a ML to learn meaningful information at these intensities. Thus, at Tropical Storm strength as in Figure 5b, CAM images typically show a wide range of features as is shown in Figure 1 for T-numbers ≤ 3 . Looking more closely at the Category 4 images (Figure 5n), it can be seen that the CAM matches quite well with the T-number 6 CF6 BF0 image in Figure 1. This

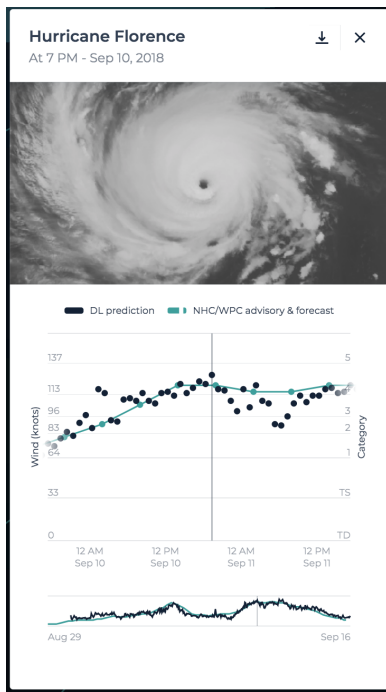


Fig. 8: Information card for a particular estimation. Here the black points represents the DL prediction (deep learning model intensity estimation) and the green line represents the NHC advised wind speeds

provides confidence that the CNN is learning the desired cloud features for classification, particularly at higher intensities.

5) *Deployment to Production System*: Machine learning researchers closely collaborated with domain experts, an end user engagement team, user interface (UI) experts, software engineers, and system architects to design an interactive real-time production system: a tropical cyclone intensity estimation portal (<http://hurricane.dsig.net/>). The portal consists of an interactive user interface where users can visually explore the estimated wind speed along with contextual information. The contextual information includes a map that shows storm location, the satellite image used for estimation, temporal information such as past observations and estimates, and several additional environmental layers (lightning, sea surface temperature, and full-disk IR satellite imagery). The event-based workflow utilized by the production system is shown in Figure 9. The design consideration of the portal focuses around a set of features:

- Continuous monitoring of NHC outlook for “invest” areas or potential areas of tropical cyclone development for triggering the wind speed estimation workflow.
- Displaying estimated wind speed and complementary information over a map.
- Allowing comparison of estimated wind speed against operational forecasts.
- Allowing download of archived estimation and satellite images for historical storms for deep dive of specific storms.

The integral part of the portal is the UI. The primary goal of the UI is to provide an intuitive interpretation of the model

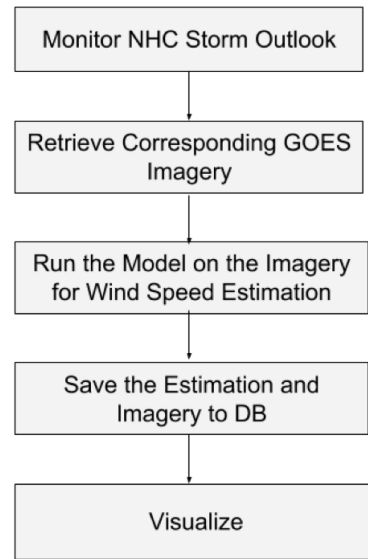


Fig. 9: Event-based workflow for production system

outputs to the general science community. By default, an overview of current storms is presented in the portal. Detailed information about a particular storm, including current historical estimates and observed wind speeds, is presented on demand as shown in Figure 6. Other data layers can be added to the default GOES-IR layer to provide additional contextual information as shown in Figure 7. A separate panel, “information card”, is used to provide a detailed snapshot of the storm as shown in Figure 8. Within the information card, the raw infrared image used for image classification at a given time is shown at the top. Below the image, a 36-hour history of intensity estimates and forecasted wind speeds from the NHC is shown by black dots and a green line respectively. Users can navigate through time with the vertical line to display the GOES-IR image, the estimated wind speed, and the official NHC wind speed for that time. At the bottom, a time series of forecast and estimation is presented for the life of the cyclone. The portal also allows users to download estimated wind speeds along with corresponding times and images for more detailed analysis of a given storm.

The system was successfully deployed prior to the 2018 hurricane season. The deployment allowed end users to explore the model output visually and in real-time. Initial response from the end user community has been very positive. The resulting feedback is being addressed as part of our continuous improvement of the estimation model and the production system.

B. System Architecture

Due to the event-based production system requirement and iterative aspects of model transition, a cloud computing native solution, specifically, Amazon Web Services (AWS) was used. Training of the model is done using on-premise GPU servers to avoid high cost of processing requirements during training. After the model is trained, the model is uploaded to an AWS S3 bucket for use within the production system. Future

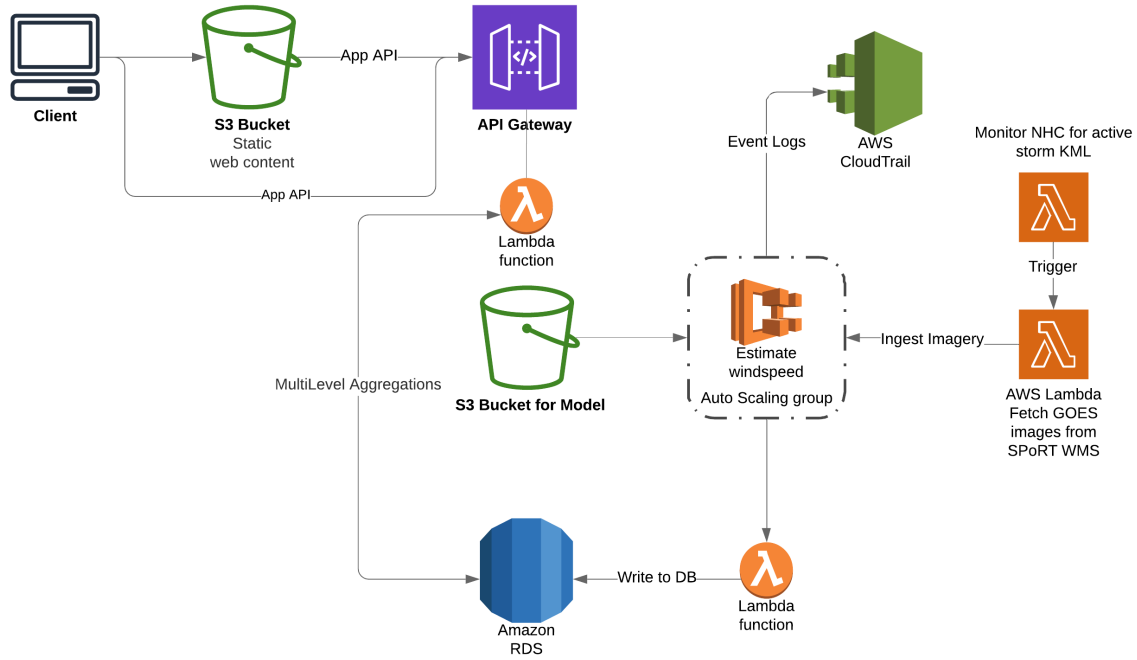


Fig. 10: Architecture of the production system

inferencing automatically uses the newest model in the AWS cloud environment. The architecture of the production system is shown in Figure 10. The service stack was deployed via AWS CloudFormation and leverages the existing open source NASA project called Cumulus [41] for many out-of-the-box features for workflow orchestration. Cumulus is a cloud-based framework for data ingest, archiving, distribution, and management of the Earth observation data. In this deployment, Amazon CloudWatch rules launch a new estimation process at the desired temporal frequency. The machine learning model executes on an Amazon Elastic Container Service (Amazon ECS) cluster, and the results are stored in Amazon Relational Database Service (Amazon RDS). They are then served to the front-end application via the application programming interface (API) gateway. Each process is run as a step function for easier debugging and logging. The portal UI is a statically generated site that invokes separate APIs to retrieve data. The site was developed using React [42] and Redux [43] as the view and state management frameworks. The display of geospatial data is handled by Mapbox GL [44], a mapping technology to display and style vector tiles. This technology provides great flexibility to style the data in the browser, enable complex user interactions, and display any necessary raster layers. The use of AWS provides a stable production system that is easy to deploy and maintain with scalable solution during periods of higher user demand (e.g. hurricane season).

IV. CONCLUSION AND FUTURE WORK

In this paper, we present an end to end deep learning-based wind speed estimation system of tropical cyclones that is triggered in real time. The system include the development of a new convolutional neural network model used to objectively estimate tropical cyclone wind speed using just satellite

images. The model is extensively evaluated and systematically transitioned to production by comparing features identified in CAMs to Dvorak T-number images. We also present a novel way to monitor for new storms and launch the workflow to provide wind speed estimates in real time using a situational awareness portal.

In addition, we find that the amount of time spent on development of algorithm is considerably lower than creating a large-scale reliable training dataset of images and corresponding wind speeds. Deployment of the machine learning model is a non-trivial task and requires several iterations with updated training data and model configurations. From the software engineering perspective, the model was treated like source code and versioned appropriately. Finally, for the successful execution of an end to end machine learning project, a diverse team of machine learning experts, domain experts, end users, software engineers, and user interface designers is needed.

Various future work could be considered including use of passive microwave data to estimate wind speed for tropical cyclones at lower intensity as in [34]. In addition, a detailed analysis of a particular storm to understand model performance with storm structural changes during rapid intensification is another future work that could be studied.

ACKNOWLEDGMENT

The authors would like to thank the NASA Earth Science Data Systems program for the support in developing this system.

REFERENCES

- [1] E. Blake, E. Rappaport, and C. Landsea, "The deadliest, costliest, and most intense United States tropical cyclones from 1851 to 2006 (and other frequently requested hurricane facts)," NOAA Technical Memo. NWS-TPC-5, 2007.

- [2] Z. Huang, D. Rosowsky, and P. Sparks, "Long-term hurricane risk assessment and expected damage to residential structures," *Reliability Engineering & System Safety*, vol. 74, pp. 239-249, 2001.
- [3] K. Emanuel, "Global Warming Effects on U.S. Hurricane Damage," *Weather, Climate, and Society*, vol. 3, pp. 261-268, 2011.
- [4] V. Dvorak, "A technique for the analysis and forecasting of tropical cyclone intensities from satellite pictures," *NOAA Technical Memo*, NESS 36, 15 pp. 14, 1972.
- [5] V. Dvorak, "A technique for the analysis and forecasting of tropical cyclone intensities from satellite pictures," *NOAA Technical Memo*, NESS 45, 19 pp. 14, 1973.
- [6] V. Dvorak, "Tropical cyclone intensity analysis and forecasting from satellite imagery," *Monthly Weather Review*, vol. 103(5), pp. 420-430, 1975.
- [7] V. Dvorak, "Tropical cyclone intensity analysis using satellite data," *NOAA Technical Report*, 11, 1984.
- [8] V. Dvorak, "Tropical clouds and cloud systems observed in satellite imagery: Tropical cyclones," *Workbook*, vol. 2, 1995.
- [9] T. Olander and C. Velden, "The current status of the UW-CIMSS Advanced Dvorak Technique (ADT)," 30th Conference on Hurricanes and Tropical Meteorology, American Meteorological Society, 2012.
- [10] M. Pineros, E. Ritchie, and J. Tyo, "Estimating tropical cyclone intensity from infrared image data," *Weather and Forecasting*, vol. 26, 2011.
- [11] M. Pineros, E. Ritchie, and J. Tyo, "Objective measures of tropical cyclone structure and intensity change from remotely sensed infrared image data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 46(11), 2008.
- [12] E. Ritchie, G. Valliere-Kelley, M. Piñeros, and J. Tyo, "Tropical cyclone intensity estimation in the North Atlantic basin using an improved deviation angle variance technique," *Weather Forecasting*, vol. 27, pp. 1264-1277, 2012.
- [13] E. Ritchie, K. Wood, O. Rodríguez-Herrera, M. Piñeros, and J. Tyo, "Satellite-derived tropical cyclone intensity in the North Pacific Ocean using the deviation-angle variance technique," *Weather Forecasting*, vol. 29, pp. 505-516, 2014.
- [14] J. Hawkins, T. Lee, J. Turk, C. Sampson, J. Kent, and K. Richardson, "Real-time inter-net distribution of satellite products for tropical reconnaissance," *Bulletin of American Meteorological Society*, vol. 82, pp. 567-578, 2001.
- [15] R. Bankert and P. Tag, "An automated method to estimate tropical cyclone intensity using SSM/I imagery," *Journal of Applied Meteorology*, vol. 41, pp. 461-472, 2001.
- [16] S. Hoshino and T. Nakazawa, "Estimation of tropical cyclone's intensity using TRMM/TMI brightness temperature data," *Journal of Meteorological Society Japan*, vol. 85, pp. 437-454, 2007.
- [17] A. Krizhevsky, I. Sutskever, and G. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," *Advances in Neural Information Processing Systems*, pp. 1097-1105, 2012.
- [18] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Boston, MA, pp. 1-9, 2015.
- [19] C. Szegedy, S. Reed, D. Erhan, and D. Anguelov, "Scalable, high-quality object detection," *arXiv*, 2015.
- [20] C. Szegedy, A. Toshev, and D. Erhan, "Deep Neural Networks for Object Detection," *Advances in Neural Information Processing Systems*, pp. 2553-2561, 2013.
- [21] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 3rd International Conference on Learning Representations (ICLR), pp. 1-14, 2015.
- [22] V. Nair and G. Hinton, "Rectified Linear Units Improve Restricted Boltzmann Machines," *Proceedings of the 27th International Conference on Machine Learning (ICML)*, pp. 807-814, 2010.
- [23] L. Wang, H. Lu, X. Ruan, and M. Yang, "Deep networks for saliency detection via local estimation and global search," *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3183-3192, 2015.
- [24] Y. Boureau, J. Ponce, and Y. LeCun, "A Theoretical Analysis of Feature Pooling in Visual Recognition," *Proceedings of the 27th International Conference on Machine Learning (ICML)*, pp. 111-118, 2010.
- [25] Y. Bengio and A. Courville, "Deep Learning of Representations," *Handbook on Neural Information Processing*, vol. 49, pp. 1-28, 2013.
- [26] Y. Bengio, A. Courville, and P. Vincent, "Representation Learning: A Review and New Perspectives," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, pp. 1798-1828, 2013.
- [27] L. Bottou, "Large-Scale Machine Learning with Stochastic Gradient Descent," *Proceedings of COMPSTAT*, pp. 177-186, 2010.
- [28] R. Pradhan, R. Aygun, M. Maskey, R. Ramachandran and D. Cecil, "Tropical Cyclone Intensity Estimation Using a Deep Convolutional Neural Network," *IEEE Transactions on Image Processing*, vol. 27(2), pp. 692-702, 2018.
- [29] M. Maskey, R. Ramachandra, I. Gurung, B. Freitag, J.J. Miller, M. Ramasubramanian, D. Bollinger, R. Mestre, D. Cecil, A. Molthan, C. Hain, "Machine Learning Lifecycle for Earth Science Application: A Practical Insight into Production Deployment," *IEEE International Geoscience and Remote Sensing Symposium*, Yokohama, Japan, pp. 10043-10046, 2019.
- [30] NOAA Comprehensive Large Array-Data Stewardship System. Accessed on Dec. 15, 2019.
- [31] B. Zhou, A. Khosla, A. Lapedriza, A. Oliva, and A. Torralba, "Learning Deep Features for Discriminative Localization," *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, pp. 2921-2929, 2016.
- [32] M. Mudigonda, S. Kim, A. Mahesh, S. Kahou, K. Kashinath, D. Williams, V. Michalski, T. O'Brien, and Mr. Prabhat, "Segmenting and Tracking Extreme Climate Events using Neural Networks," 31st Conference on Neural Information Processing System, pp. 1-5, 2017.
- [33] T. Olander and C. Velden, "The Advanced Dvorak Technique: Continued Development of an Objective Scheme to Estimate Tropical Cyclone Intensity Using Geostationary Infrared Satellite Imagery," *Weather and Forecasting*, vol. 22, pp. 287-298, 2007.
- [34] T. Olander and C. Velden, "The advanced dvorak technique (ADT) for estimating tropical cyclone intensity: Update and new capabilities," *Weather and Forecasting*, vol. 34, pp. 905-922, 2019.
- [35] S. Gao, P. Zhao, B. Pan, Y. Li, M. Zhou, J. Xu, S. Zhong, and Z. Shi, "A nowcasting model for the prediction of typhoon tracks based on a long short term memory neural network," *Acta Oceanologica Sinica*, Vol. 37, pp. 8-12, 2018.
- [36] A. Wimmers, C. Velden, and J. Cossuth, "Using deep learning to estimate tropical cyclone intensity from satellite passive microwave imagery," *Monthly Weather Review*, Vol. 147, pp. 2261-2282, 2019.
- [37] S. Giffard-Roisin, M. Yang, G. Charpiat, C. Kumler Bonfanti, B. Kégl, and C. Monteleoni, "Tropical Cyclone Track Forecasting Using Fused Deep Learning From Aligned Reanalysis Data," *Frontiers in Big Data*, Vol. 3, pp. 1-13, 2020.
- [38] J. Lee, J. Im, D. Cha, H. Park, and S. Sim, "Tropical Cyclone Intensity Estimation Using Multi-Dimensional Convolutional Neural Networks from Geostationary Satellite Data," *Remote Sensing*, vol. 12, pp. 108, 2019.
- [39] S. Kim, H. Kim, J. Lee, S. Yoon, S. Kahou, K. Kashinath, and Mr. Prabhat, "Deep-hurricane-tracker: Tracking and forecasting extreme climate events," In *Proceedings 2019 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pp. 1761-1769, 2019.
- [40] C. Landsea and J. Franklin, "Atlantic Hurricane Database Uncertainty and Presentation of a New Database Format," *Monthly Weather Review*, Vol. 141, pp. 3576-3592.
- [41] Cumulus Framework, GitHub Repository, <https://github.com/nasa/cumulus>
- [42] React, GitHub Repository, <https://github.com/facebook/react>
- [43] Redux, GitHub Repository, <https://github.com/reduxjs/redux>
- [44] mapbox-gl-js, GitHub Repository, <https://github.com/mapbox/mapbox-gl-js>



Dr. Manil Maskey received Ph.D. degree in Computer Science from the University of Alabama in Huntsville in 2019. He is a Research Scientist with the National Aeronautics and Space Administration (NASA) Marshall Space Flight Center (MSFC), Huntsville, AL, USA. He also leads the Advanced Concepts team, within the Inter Agency Implementation and Advanced Concepts (IMPACT). Currently, he chairs the IEEE Geoscience and Remote Sensing Society, Earth Science Informatics Technical Committee and leads the machine learning activities for the NASA Earth Science Data Systems program. His research interests include computer vision, visualization, knowledge discovery, cloud computing, and data analytics.



Dr. Rahul Ramachandran is a senior Research Scientist for the National Aeronautics and Space Administration (NASA) at Marshall Space Flight Center (MSFC). His research centers on Earth Science Informatics and Data Science with a focus on the application of novel computational methods and information technology to the acquisition, storage, processing, discovery, interchange, analysis and visualization of Earth Science data and information. In addition, Dr. Ramachandran manages the Inter Agency Implementation and Advanced Concepts (IMPACT) team at NASA MSFC. The IMPACT team seeks to infuse NASA's Earth Science data into other agencies and organizations application workflows. The IMPACT team monitors trends across the informatics, data science, and information technology fields to inform strategy and develop effective new solutions for Earth Science data management and dissemination. Dr. Ramachandran received the Presidential Early Career Award for Scientists and Engineers (PECASE) award in 2009 and the NASA Exceptional Achievement Medal in 2018. He is also a Senior IEEE member.



Aaron Kaulfus received his B.S. (2012) in Meteorology from the University of Oklahoma and M.S. in Atmospheric Science (2015) from the University of Alabama in Huntsville. He is currently working on his Ph.D. in Atmospheric Science at the University of Alabama in Huntsville. He joined the Inter Agency Implementation and Advanced Concepts Team (IMPACT) in 2018 as a Research Associate where he has been involved in machine learning applications for event detection and estimation using high volume Earth science datasets. His research interests include satellite remote sensing, air quality, data analytics, and developing innovative data management and stewardship solutions for data and knowledge discovery. Currently, he is a member of IEEE and the Earth Science Informatics Technical Committee.



Muthukumaran Ramasubramanian is a PhD candidate in Computer Science at the University of Alabama in Huntsville. He is a Machine Learning researcher working with NASA Inter Agency Implementation and Advanced Concepts Team (IMPACT). His research areas include deep learning, computer vision, natural language processing, radio frequency domain machine learning and large scale machine learning.



Drew Bollinger leads the Labs Team at Development Seed. Drew is a data analyst, software developer, and machine learning engineer, with experience building geo-interfaces and running advanced statistical and spatial analysis on open data sets. He has delivered several impactful workshops at Satsummit, IGARSS, and most recently at the Africa Geospatial Data and Internet Conference in Accra, Ghana.



Iksha Gurung holds a MS in Computer Science from the University of Alabama in Huntsville. He is a Research Scientist at the NASA Inter-Agency Implementation and Advanced Concepts Team (IMPACT), where he leads the development of machine learning models for variety of Earth science applications. His research interests include machine learning for remote sensing, cloud computing, and web framework.



Dr. Daniel J. Cecil received Ph.D. degree in Atmospheric Science from Texas AM University in 2000. He is a Research Scientist with the National Aeronautics and Space Administration (NASA) Marshall Space Flight Center (MSFC), Huntsville, AL, USA. His research has generally focused on thunderstorms and tropical cyclones, particularly using satellite-based remote sensing to understand and diagnose properties of those phenomena.



Dr. Brian Freitag holds a PhD in Atmospheric Science from the University of Alabama in Huntsville. He currently serves as a Research Scientist at the University of Alabama in Huntsville publishing high-value Earth science data sets and providing subject matter expertise on the deployment of machine learning models relevant to Earth science. Brian also serves on the American Meteorological Society Board on Urban Environment. His research interests include urbanization, land/use land cover change, mountain meteorology, and extreme weather.



Jeffrey Miller received a M.S. in Atmospheric Science from The University of Alabama in Huntsville in 2018. In 2017 Jeffrey began working with the Inter Agency Implementation and Advanced Concepts Team (IMPACT) through The University of Alabama in Huntsville as a research associate. Jeffrey now works as a meteorologist for Climate Forecast Applications Network in Atlanta, Georgia. His interests include applying machine learning to atmospheric science problems, tropical cyclone forecasting, and working with remote sensing data.