PHURIE: Hurricane Intensity Estimation from Infrared Satellite Imagery using Machine Learning

Amina Asif[1], Muhammad Dawood[1], Bismillah Jan[1], Javaid Khurshid[1], Mark DeMaria[2] and Fayyaz ul Amir Afsar Minhas[1,*]

1. *Department of Computer and Information Sciences, Pakistan Institute of Engineering and Applied Sciences (PIEAS), PO Nilore, Islamabad, Pakistan.*
2. *National Hurricane Center, National Oceanic and Atmospheric Administration (NOAA), Miami, FL, United States of America*

** corresponding author email: afsar@pieas.edu.pk; fayyazafsar@gmail.com, Phone: +92-51-2207381 to 85 Extension: 3164, Fax: +92-51-9248600*

ABSTRACT: Automated prediction of hurricane intensity from satellite infrared imagery is a challenging problem with implications in weather forecasting and disaster planning. In this work, a novel machine learning based method for estimation of intensity or maximum sustained wind speed of tropical cyclones over their life-cycle is presented. The approach is based on a support vector regression model over novel statistical features of infrared images of a hurricane. Specifically, the features characterize the degree of uniformity in various temperature bands of a hurricane. Performance of several machine learning methods such as Ordinary Least Squares Regression, Backpropagation Neural Networks and XGBoost regression has been compared using these features under different experimental setups for the task. Kernelized support vector regression resulted in the lowest prediction error between true and predicted hurricane intensities (approximately 10 knots or 18.5km/hour), which is better than previously proposed techniques and comparable to SATCON consensus. The performance of the proposed scheme has also been analyzed with respect to errors in annotation of center of the hurricane and aircraft reconnaissance data. The source code and webserver implementation of the proposed method called PHURIE (PIEAS HURricane Intensity Estimator) is available at the URL: http://faculty.pieas.edu.pk/fayyaz/software.html#PHURIE.

Keywords: Hurricane Intensity Prediction; Tropical Cyclones; Machine learning based forecasting; Support Vector Regression

1. Introduction

Hurricanes are among the most destructive natural phenomena on earth. They form over warm tropical and subtropical oceans during summers or early fall. Upon making landfall, hurricanes can cause significant property damage, and loss of life [1]. Timely analyses and forecasts of track, intensity and wind structure can help authorities raise warnings, evacuate high-risk regions, estimate expected losses, and minimize mortalities.

Due to the limited availability of direct measurements, satellite images of hurricanes throughout their lifecycles have been analyzed for the past several decades. One of the earliest methods for tropical cyclone (TC) intensity estimation is the Dvorak technique [2], which is a manual method that characterizes a TC based upon the cloud structure seen in an image. To reduce the reliance on human experts, the Objective Dvorak Technique [3] was proposed in 1989 for automatic intensity estimation based on rules similar to original Dvorak technique. More sophisticated rules were introduced in the Advanced Dvorak's Technique [4], which resulted in an improvement in prediction accuracy. However, human involvement was still required and the method could not be automated completely. Since then, many studies have been carried out to help automate the process for improvement in speed and reduction in need for human intervention. A brief description of several of such studies is presented below.

Piñeros et al. proposed a method based on the variance of the deviation angle of brightness temperature values in infrared (IR) images [5]. Their method was built on the premise that the lower the variance in the histogram of deviation angles, which is inversely proportional to TC organization, the higher would be intensity of the TC. A sigmoid curve was fit to use variance of deviation angles for intensity estimation. In their study, Piñeros et al. used IR images from the GOES-12 satellite for hurricanes in years 2004-2009 in the North Atlantic Basin. Their method

gives a Root Mean Squared Error (RMSE) of 14.7 knots when evaluated over a randomly selected set of hurricanes over the period 2004-2008. The same model, when trained over data from 2004-2008 and tested over TC IR images from year 2009, produced an RMSE of 24.8 kt. An improved version of their technique was presented by Ritchie et al. [6]. That study added some additional constraints to the existing technique and re-trained it after removing low intensity (<34 kt) TC images from data and using data from an additional year (2010). This resulted in an RMSE of 12.9 kt. The Deviation Angle Variation technique was used to estimate the intensities of TCs in the north Pacific ocean in a 2013 study [7] with an RMSE of 14.3 kt. [8] proposed a k-nearest neighbor based algorithm for TC intensity estimation. Their algorithm estimated the intensity based on the intensity of the 10 most similar images to the query image. In a study carried out by Jaiswal et al., brightness temperature histograms in the radial and angular directions were computed and histogram matching was used for intensity predictions [9]. Their study used TC data collected using satellites GOES-8 and -12 from 2000-2005 from the HURSAT database [10]. The method yielded an overall RMSE of 15.5 kt. The study by Zhao et al. presents a multiple regression based method using deviation angle and radial profiles in IR images for intensity estimation [11]. The method was tested on hurricane data from Northwestern Pacific Ocean over the years 2008 and 2009 and an RMSE of 12.1 kt was reported.

The objective of our study is to develop a machine learning based automated system that can predict intensity of a hurricane when given its satellite infrared (IR) image. The workflow of the proposed system is illustrated in Figure 1. The proposed system computes statistical and deviation angle-based features for an input IR image. For prediction, the features are passed to a machine learning model that has been trained using existing data comprising of satellite images of previous hurricanes with known intensity. In this paper, we present details of our proposed method. The

dataset, feature extraction and machine learning models are described in section 2, results are presented in section 3 and conclusions are summarized in section 4.

2. Methods

In this section, we present details of the dataset, feature extraction technique, machine learning models and the experimental setup employed in our study. The primary task of the proposed technique is to use machine learning for predicting the maximum sustained windspeed or intensity of a hurricane (in knots or kilometers per hour) from infra-red satellite images of the hurricane. Section 2.1 provides a description of the dataset used for training and evaluation of the machine learning model. In section 2.2, we explain feature extraction methods. Analysis of feature importance is presented in section 2.3. Different machine learning models analyzed in the study are described in section 2.4. Post-processing and experimental setup used for performance evaluation have been explained in sections 2.5 and 2.6, respectively.

2.1 Dataset

Our study used infrared images from the publicly available HURSAT-B1 (version-05) dataset [10] of different hurricanes. The original dataset contained hurricane season data for years 1978 to 2009 and included imagery from multiple satellites including SMS-2, GOES-1 to 13, Meteosat-2 to 9, GMS-1 to 5, MTSAT-1R, MTS-2 and FY2-C/E. HURSAT-B1 contains both visible and IR window channel imagery. Example satellite infra-red images from the dataset are shown in Figure 2 in false coloring. A pixel value corresponds to temperature at a certain location as captured by the satellite with higher temperatures shown in red and lower ones shown in blue. The spatial resolution of the data is about 8 km/pixel (4.32 nautical miles per pixel), i.e., a single pixel represents the average temperature in an 8km × 8km region on the Earth's surface. The dataset contains images from a number of hurricanes taken every 3 hours for each hurricane.

Images in the dataset are centered on the TCs. Information about the intensity of a given image of a hurricane was taken from IBTrACS (International Best Track Archive for Climate Stewardship) [12]. The intensity of a hurricane at a given time is defined as the maximum sustained surface windspeed (in knots) of the hurricane at a height of 10m from the surface of the Earth over a period of 1 minute (60 seconds). Based on the maximum sustained surface windspeed (in knots), a tropical storm can be classified into 5 categories. IBTrACS stores the intensity of the hurricane based on a consensus of automated, semi-automated and aircraft reconnaissance data. In line with previous studies, the best track data was linearly interpolated to match the temporal resolution of the image data. We used the intensity in knots as our target or output value.

We restricted our study to hurricane data collected by GOES-12 satellite in the North Atlantic Basin from years 2004 to 2009. Only infrared (IR) window channel imagery was used in our study. Images taken after a TC made land-fall were removed from the dataset for our experiments. The subset used in the study included a total of 4552 images. Details about the intensity distribution of the sample are presented in Table1.

2.2 Feature Extraction

In satellite IR images, high intensity TCs present themselves as well-organized low-temperature circular cloud structures. For low intensity TCs, the cloud structure becomes less organized. This phenomenon is shown in Figure 2. It can be seen that, as the intensity increases, the cloud structure becomes more symmetric and the organization of the clouds increases. This relationship was also the basic premise of the deviation angle technique described earlier.

We use the above-mentioned phenomenon to extract features for intensity estimation of TCs. That is, the region around the center tends to exhibit a more uniform low-temperature circular structure in high intensity TCs in comparison to low intensity TCs. Therefore, we compute

statistical features around the center to characterize the TC structure. To compute these features, we first divided each image into 5 circular bands of 8 pixels each (equivalent to 64 km or 34.56 nautical miles) around the center. For each band, mean, standard deviation (SD), entropy, minimum and maximum are computed. Division of images into bands is illustrated in Figure 3. Formulae for computation of statistical features are listed in Table 2. The correlation of these features with hurricane intensity is shown in Figure 4 as discussed in the next section.

In addition to the statistical features, we used variance of the deviation angle histogram as another feature for TC intensity estimation. The idea was motivated from the approach proposed by Piñeros et al. [5]. Deviation angle at a pixel is defined as the angle between the gradient vector and the line joining the hurricane center and that pixel. For well-organized circular structures, most of the deviation angles around the center are zero or near to zero. The concept is illustrated in Figure 5(a-c). Since high intensity TCs exhibit more circular structures, most of the deviation angles in their images would be small and the histogram of these angles will have a low variance. We have used variance of deviation angle histogram for 81x81 pixel window (equivalent to 648×648 km or 350×350 nautical miles) centered at the center of an image as another feature.

2.3 Analysis of importance of features

To assess the effectiveness of statistical features around the center for intensity estimation, we plotted the features against intensity values for hurricane Rita (2005). The scatter plots are shown in Figure 4. It can be seen that a high negative correlation exists for most of the features. For example, the mean temperature of bands 2-4 show negative correlations with magnitude greater than 0.75 with TC intensity. Similarly, the standard-deviation of IR values also show a high inverse correlation. Thus, the mean IR intensities within 24-48 km (12.96-25.92 nautical miles) of the center of the TC and their uniformity are highly predictive of intensity. The entropy

and maximum values of temperatures in various bands are also inversely correlated with intensity. These plots clearly show the efficacy of using these statistical features in our technique.

The effectiveness of the Deviation Angle Variance feature in terms of correlation with true intensity values has also been measured for hurricane Rita (2005). The plot for deviation angle variance versus true intensity values has been shown in Figure 5(d). It is worth mentioning here that simple statistical features such as mean, standard deviation, minimum and maximum temperatures for the third band produce comparable correlation values as the complex deviation angle variance-based feature. Hence, we deduce that, the statistical features despite being simpler, are as informative as deviation angle variance feature and hence, may help improve hurricane intensity predictions.

2.4 Machine Learning Models

In this study, our goal is to develop a system that, given a TC image and a center position, can predict its intensity. We have modeled the problem of predicting the intensity of a hurricane at a given time as a regression problem. For this purpose, we consider a dataset of N example training images represented by their d-dimensional feature vectors $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_N$ corresponding to different infra-red satellite images of hurricanes and their associated intensity values $y_1, y_2, \ldots, y_N$ in knots. The objective of hurricane intensity prediction is to develop a machine learning prediction function $f(\boldsymbol{x})$ that can predict the intensity of the hurricane at a given time using a feature vector $\boldsymbol{x}$ corresponding to an image of the hurricane at that time. To choose the best-suited machine learning model for this problem, we carried out detailed performance analysis and comparison over different regression techniques: Ordinary Least Square (OLS) [13] and Support Vector Regression (SVR) [14] with Radial Basis Function (RBF) kernel, feed-forward backpropagation neural networks (BPNNs) [19] and gradient boosted tree (XGBoost) regression [20]. To establish

if these models are significantly effective in comparison to a naïve prediction, we compared their results to a zero-order baseline that uses the average intensity of the hurricanes in our data set as a constant prediction. Multiple machine learning techniques were compared to identify the best suited one for this task and to analyze the effectiveness of features used in this work by studying the difference in prediction errors of these techniques. Low variation in performance across the techniques implies that the features are significantly informative and a difference in choice of machine learning model would not have a considerable impact on the accuracy of the system and that the deployed model will generalize well to unseen cases. Further details of performance comparison are given in Results section. In the following sections, we present description of the various techniques used in this study.

2.4.1 Baseline Method

To establish a baseline, we used the average intensity of TCs in the whole dataset as a zero-order intensity estimator for any given image.

2.4.2 Ordinary Least Square (OLS) Regression

OLS is one of the simplest regression techniques. The principle of OLS is to find a linear function that minimizes the sum of squared errors between target and estimated values for a given dataset. The objective in OLS is to find parameters **w** and b of a linear function $f(\boldsymbol{x}) = \boldsymbol{w^T x} + b$ such that that the difference between the target value y_i and $f(\boldsymbol{x}_i)$ is minimized for all training examples $i = 1 \dots N$. The OLS learning problem can be written as: $\boldsymbol{w}, b = \underset{\boldsymbol{w},\boldsymbol{b}}{argmin} \sum_{i=1}^{N} \big(y_i - f(\boldsymbol{x}_i)\big)^2$. The parameters estimated from training data are then used for estimation of values for independent cases.

There are two shortcomings of using OLS for our problem. First, OLS is prone to over/under-estimation due to the presence of outliers, as its sole aim is to minimize the sum of squared errors [15]. Second, we were not sure if a linear function would successfully be able to model the relationship between the features we extracted and the corresponding intensity values. Therefore, we needed a method that was less sensitive to outliers, offered better generalization and could model non-linear relationships. As a consequence, we used Kernelized Support Vector Regression [14].

2.4.3 KERNELIZED SUPPORT VECTOR REGRESSION

Kernelized SVR is a variant of Support Vector Regression which, originally, is a linear regression technique, i.e., its prediction function can also be written as: $f(\boldsymbol{x}) = \boldsymbol{w}^{\boldsymbol{T}}\boldsymbol{x} + b$. However, it can work for non-linear estimation using kernel functions. For a given dataset, SVR finds a weight vector $\boldsymbol{w}$ such that the norm of $\boldsymbol{w}$ is minimized and the absolute difference between the actual and predicted values for all examples does not exceed a threshold $\varepsilon > 0$. The optimization problem in this case can be given as: $\underset{\boldsymbol{w},b}{min}\ \|\boldsymbol{w}\|^2$ such that $|f(\boldsymbol{x}_i) - y_i| < \varepsilon$ for all $i \in \{1, 2, \dots, N\}$. Minimization of the norm of the weight vector ensures that the weight values do not become large and small changes in the inputs do not cause a large variation in the output. This regularization helps improve prediction performance in high dimensional and noisy feature spaces. To allow some violations, a non-negative slack variable ξ_i is introduced for each example $\boldsymbol{x}_i$and the optimization problem can therefore be modified to $\underset{\boldsymbol{w},b,\boldsymbol{\xi}\geqslant\boldsymbol{0}}{min}\ \|\boldsymbol{w}\|^2 + C\sum_{i=1}^{N}\xi_i$ such that $|f(\boldsymbol{x}_i) - y_i| < \varepsilon + \xi_i$ for $i \in \{1, 2, \dots, N\}$. This problem formulation ensures that the prediction errors are minimal, and the predictor is regularized. The hyper-parameter C controls the amount of penalty imposed for each constraint violation. It is important to note that SVR minimizes the

absolute error and not the square-error function. This reduces the impact of outliers in comparison to OLS. An alternative representation of the SVR [14], allows non-linear regression by using RBF kernel functions $k(\boldsymbol{a}, \boldsymbol{b}) = exp(-\gamma \|\boldsymbol{a} - \boldsymbol{b}\|^2)$ and changing the prediction function to $f(\boldsymbol{x}) = \sum_{i=1}^{N} \alpha_i k(\boldsymbol{x}, \boldsymbol{x}_i)$ [16], [17]. This kernelized formulation of the SVR learns parameters α_i while enforcing regularization and error minimization over training data. The kernel function $k(\boldsymbol{a}, \boldsymbol{b})$ is a symmetric positive definite function that essentially measures the degree of similarity between examples. We have used SVR with RBF kernel for our experiments as RBF has the ability to model spaces of very high dimensionality effectively [18]. The hyper-parameters γ and C are set using nested cross-validation.

2.4.4 Back Propagation Neural Networks

Neural Networks are function approximators inspired from the structure of human brain. They are composed of layers of small computational units called neurons. The output of neurons in a layer is fed to the neurons in the next layer. Each neuron computes its output by applying an activation function over the dot product of its weights and inputs. The final output is computed in the last layer. During training, the objective is to minimize the error between output of the neural network and the target values. To fit a model using a BPNN, an example or a batch of examples from the training data is passed to the network and output is computed. The error is calculated and weights of the network are updated in a direction opposite to the gradient of error [19]. The process is repeated iteratively to minimize training loss. Since the error surface is not always convex, backpropagation may yield sub-optimal solutions. For comparison with our methods we have used a BPNN with two hidden layers, 64 neurons per layer, and rectified-linear unit (ReLu) activation functions with a single output layer neuron. The neural network has been implemented using Keras [21].

2.4.5 XGBOOST

XGBoost [20] is a random-forest based method that uses gradient boosted decision trees. A decision function that performs minimization of average regression loss is learned using gradient boosting on a set of decision trees trained in an iterative manner. The training in each increment is performed using residual error of the preceding step. Further details of the technique can be found in [20]. In our experiments, we used Python xgboost v. 0.7 API for XGBoost regression.

2.5 Post-Processing

Our model generates predictions using a single image. To reduce noise, a time-smoothing operation is performed after generating predictions for different images of a TC. For this purpose, we used a simple linear exponentially weighted averaging filter that, at a time step t, produces a weighted average of predicted intensities for current and previous time steps as follows: $g(\boldsymbol{x}_t) = 0.41f(\boldsymbol{x}_t) + 0.25f(\boldsymbol{x}_{t-1}) + 0.15f(\boldsymbol{x}_{t-2}) + 0.1f(\boldsymbol{x}_{t-3}) + 0.06f(\boldsymbol{x}_{t-4}) + 0.03f(\boldsymbol{x}_{t-5})$. It is important to note that the coefficients of the filter sum to 1.0 and decrease exponentially with time.

2.6 Experimental Setup

We performed multiple experiments over features and regression models discussed earlier for the 2004-2009 sample. We have used Root Mean Squared Error (RMSE) [22] as the performance metric to evaluate and compare the efficacy of our methods with previously published works. Results for the experiments are presented and discussed in Section 3.

2.6.1 LEAVE ONE TC OUT CROSS VALIDATION

For all TCs over the period 2004-2009, we left one hurricane out for testing and trained over the rest. RMSE scores for each of the test hurricanes were computed and then averaged. The experiment was performed for all of the regression techniques described in section 2.4: OLS, SVR, Feed-forward BPNN and XGBoost.

2.6.2 Stratified error analysis

We have performed stratified error analysis of our method for different stages of TC development to get an idea of prediction accuracy for low vs. high intensity hurricanes using leave one TC out cross-validation.

2.6.3 Comparison with Deviation Angle Variance Technique

To compare our method with the deviation angle variance based method, we replicated the experiments carried out in [5]. Two experiments were performed in the study. The first experiment uses data from 2004-2008. The following hurricanes were left out for testing: Bonnie (2004), Earl (2004), Jeanne (2004), Matthew (2004), Nicole (2004), Dennis (2005), Irene (2005), Katrina (2005), Nate (2005), Rita (2005), Tammy (2005), Delta (2005), Debby (2006), Isaac (2006), Arthur (2008), Cristobal (2008), Fay (2008), Hanna (2008), Kyle (2008) and Paloma (2008). The rest of the TCs over the period 2004-2008 were used for training.

In the second experiment, all TCs from 2004-2008 were used for training and data from 2009 were used for testing. We report the RMSE results for both OLS and SVR.

2.6.4 Leave One Year Out Cross Validation

In this experiment, we used the data for all years from 2004-2009. TCs from one year are left out for testing and training is performed over the rest. This experiment was performed to compare our method with the improved version of the DAV technique [5] proposed by [6]. Their experiment used long range IR images from GOES-E satellite and used data of one additional year (2010). We report RMSE results for our data using the same Leave One Year Out Cross validation method.

2.6.5 Comparison with Aircraft Reconnaissance Data

Aircraft reconnaissance data is available for several hurricanes and it gives very reliable estimates of hurricane intensity at certain times. We compared the predictions of the proposed model with aircraft measurements by performing leave one TC out cross-validation and restricting our error evaluation to only those times that were within 3 hours of an aircraft pass.

2.6.6 Center Annotation Error Analysis

As the proposed scheme relies on center annotations for feature extraction, we also analyzed the effect of error in annotating the center of the hurricane on intensity estimation. For this purpose, we selected a single hurricane from every year at random for evaluation through leave one TC out cross-validation. The annotated center in IR images of a hurricane was shifted along both axes by a random amount within the interval $[-r, +r]$ prior to feature extraction and intensity prediction. The magnitude of the shift, r, was varied from 0 to 10 pixels (corresponding to a maximum center position error of 80 km or 43.2 nautical miles) to model the effect of center annotation errors of existing center prediction methods [1]. This process is repeated 5 times for each hurricane to get reliable estimates. The prediction error of the proposed technique is then plotted against the magnitude of the shift in the annotated center for analyzing the effect of center annotation error on intensity prediction error.

2.6.7 Analysis of Images from Other Channels

The focus of this study has been to predict TC intensity from IR images. However, in order to assess the effectiveness of the features proposed in this work over data from other channels, we have also evaluated leave one year out cross-validation analysis over other available channels including: Visible channel observations (VSCHN), water vapor observations (IRWVP), Near-infrared channel observations (IRNIR) [10].

2.7 PHURIE Webserver

We have developed a freely available webserver called PHURIE (Python HURricane Intensity Estimator) for the proposed method which accepts an IR image in netcdf file format, extracts features, and generates a prediction from machine learning model. The center of the image input should correspond to the center of the hurricane. PHURIE uses a kernelized SVR model, since the SVR based models had shown to generally outperform others in different experiments. Details of the performance comparison of different regression techniques is presented in the Results section. It is important to note that the webserver generates predictions using a single input image without any post-processing. The websever has been developed using Python and scikit-learn and is available at the URL: http://faculty.pieas.edu.pk/fayyaz/software.html#PHURIE.

3. Results and Discussion

In this section results for all the experiments performed under the setup discussed in the previous sections are presented and discussed.

3.1 Leave One TC out Cross Validation

Using the mean intensity (zero-order predictor) as the predicted intensity for a given image, gives an RMSE of 24.3 kt as shown in Table 3. This is the expected maximum error of any technique and is used as a baseline for comparison.

For evaluation of our method, we used Leave One Hurricane Out model as described in section 2.4.1. Mean RMSE values obtained using SVR, OLS, BPNN and XGBoost regression models are presented in Table 3. Using SVR, we obtained a mean RMSE of 11.2 kt. As expected, the proposed method performs much better than the zero-order predictor. The error of SVR is much lower than other machine learning models. Furthermore, the post-processing smoothing step

reduces these errors even further to 9.5 kt which is comparable to CIMSS satellite consensus (SATCON) intensity prediction error (9.1kt) [23].

We have also performed leave-one-TC-out cross-validation with a feed-forward back-propagation neural network. The average RMSE for the neural network is 12.0 knots which is marginally higher than RMSE of 11.2 knots obtained with support vector regression. We tuned different parameters of the neural network but no significant reduction in error was noted. We have also used XGBoost regression for this problem which gives an RMSE of 11.3 knots after optimization of various hyper-parameters such as the number of estimators, subsampling, etc.

3.2 Stratified Error Analysis

Figure 6 shows a scatter plot of the true and SVR-predicted intensities for all images of all TCs. The overall Pearson correlation between true and predicted intensities is 0.91, whereas, the overall RMSE is 10.6 kt, which indicates the effectiveness of our approach. Figure 6 also shows RMSE errors for different TC stages. We hypothesize that the increased error at higher intensities is a consequence of the presence of relatively fewer training images at these intensities (see table 1) and the nature of the error function (RMSE) being used. For a deeper evaluation of the performance of our method, we present plots of SVR-predicted vs. actual intensities for hurricanes Katrina and Rita (2005) in figures 7 and 8 respectively. A high correlation can be observed for both the cases. It can be seen that, in contrast to most of the existing techniques, our method performs well even for low-intensity images.

3.3 Comparison with Deviation Angle Variance Technique

Results of the two experiments replicated from [5] are given in Tables 4 and 5. The comparison of our approach with their results using the same experimental conditions show that

all machine learning models used in this work outperform their approach in both the experiments. A major improvement has been seen in the second experiment (Table 5), where the TCs from years 2004-2008 were used for training and testing was performed on hurricanes in 2009. We obtained a mean RMSE of 13.4 kt compared to the previously reported 24.8 kt [5]. Post-processing using temporal smoothing filter improves the results even further to 11.5 knots. It is important to note that the proposed scheme offers better accuracy than the recently published method by Zhao et al [11] which gives an RMSE of 12.1knots over typhoons in the northwestern pacific ocean in 2009 as well.

3.4 Leave One Year Out Cross Validation

In this experiment, the aim was to compare our method with the improved version of the method proposed by [5] in [6]. They used TC data over the period 2004-2010. Our dataset comprised data over the period 2004-2009 obtained from the GOES-12 satellite. Also, in their experiment, Ritchie et al. used only the data for TCs with a minimum speed of 34 kt as low intensity TCs are reported to adversely affect the accuracy of their method.

We have compared the performance of SVR, OLS, XGBoost and BPNN for Leave One Year Out cross-validation over all the images including both high and low intensity examples. The comparison is presented in Table 6. As evident from the results, SVR gives better prediction accuracy (RMSE of 11.1 knots) in comparison to other machine learning models. For TC data with a minimum intensity of 34 kt, all machine learning models perform better than the DAV [5] and Improved DAV techniques [6] (Table 7).

It can be seen that, including the low intensity examples did not have much effect over the performance of our method. It can, therefore, be concluded that the proposed method is more robust and has a better performance than the previously published techniques.

3.5 Comparison with Aircraft Reconnaissance Data

On restricting our error evaluation to only those points in time that are within 3 hours of an aircraft pass, we get a mean RMSE of 12.1 kts which is only slightly above the average RMSE for leave one TC out cross-validation (11.2 kts). This clearly illustrates the true generalization performance of the proposed scheme.

3.6 Center Annotation Error Analysis

The plot of prediction error in response to center annotation error is shown in Figure 9. It shows that the proposed system undergoes graceful degradation in performance with increase in center annotation error. Figure 9(a) shows the effects of random center shifts in hurricane RITA 2005 images whereas Figure 9(b) shows the change in prediction accuracy as a consequence of random center shifts for 5 different hurricanes. The average RMSE increases from 11.5 kts to 16.5 kts for these hurricanes as the pixel shift is varied from 0 to 10 pixels (corresponding to 80 km or 43.2 nautical miles). [24] showed that when only satellite data were available, the mean position uncertainty of tropical storms, hurricane, and major hurricanes was 29, 21 and 14 nautical miles, respectively. Those roughly correspond to 7, 5 and 3 pixel displacements in Figure 9. For hurricanes and major hurricanes, Figure 9 suggests that the position uncertainty would only slightly degrade the intensity estimates. For tropical storms, the impact is larger. Tests with real-time position estimates are needed to assess the accuracy of our system for operations.

3.7 Analysis of Images from Other Channels

Leave One Year Out cross validation results using our proposed features over images from near infrared (IRNIR), water vapor (IRWVP) and visible (VSCHN) channels through SVR, OLS, BPNN, and XGBoost machine learning models are presented in Tables 8-10, respectively. The best mean RMSE values for the three channels are 12.3 kts (using XGBoost), 12.3 kts (using SVR) and 17.9 kts (using XGBoost), respectively. It is important to note that although these values are higher than the RMSE obtained using IR channel (11.1 knots with SVR), the relatively small decrease in accuracy for other channels, especially the near-IR and water-vapor channels, clearly indicates the effectiveness of the features proposed in this work. The poor performance in visible channel images can be attributed to the quality of these images being dependent upon lighting conditions.

4. Conclusions and Future Work

In this paper we presented a Support Vector Regression based technique for TC intensity estimation from satellite IR images. Since the shape of the cloud patterns helps in estimation of TC intensity in manual methods, we used several statistical features to characterize the structure in circular bands around the center of a hurricane image. These features included mean, minimum, maximum, standard deviation and entropy of bands. Apart from these features, variance of deviation angle histogram of an image was also used. The method proposed in this paper gives robust and state of the art performance on a number of different experiments and can be adapted for practical use. The features proposed in the study can also be employed for other prediction tasks related to hurricane IR imagery such as path-tracking. Although the main focus of this study was hurricane intensity prediction using infrared images, we have evaluated the proposed method on images from other channels including near infrared, water vapor and visible channels. In the

future, we plan on making a single machine learning method that can learn to predict both the center of a hurricane and its intensity.

The results from this study show that the PHURIE intensity estimates are more accurate that other automated methods documented in published papers, and comparable to methods that use a consensus of several methods (such as the CIMSS SATCON). However, some assumptions such as the use of best track positions, may inflate the accuracy of the estimates. The next step is to perform completely independent tests using only input that is available in real time. That will provide a true estimate of applicability of PHURIE for operational forecast centers.

Acknowledgements

We are grateful to Dr. Charles Anderson, Colorado State University, USA and Dr. John Knaff, National Oceanic and Atmospheric Administration for discussion and suggestions. We will also like to acknowledge the internal reviewers of National Hurricane Center for their comments that helped us improve the quality of the manuscript.

AA, MD and BJ are funded via Information Technology and Telecommunication Endowment Fund at Pakistan Institute of Engineering and Applied Sciences.

Conflict of interest

The authors declare no conflict of interest.

REFERENCES

[1] R. A. Pielke Jr, J. Gratz, C. W. Landsea, D. Collins, M. A. Saunders, and R. Musulin, "Normalized hurricane damage in the United States: 1900–2005," *Nat. Hazards Rev.*, vol. 9, no. 1, pp. 29–42, 2008.

[2] V. F. Dvorak, "Tropical Cyclone Intensity Analysis and Forecasting from Satellite Imagery," *Mon. Weather Rev.*, vol. 103, no. 5, pp. 420–430, May 1975.

[3] C. S. Velden, T. L. Olander, and R. M. Zehr, "Development of an Objective Scheme to Estimate Tropical Cyclone Intensity from Digital Geostationary Satellite Infrared Imagery," *Weather Forecast.*, vol. 13, no. 1, pp. 172–186, Mar. 1998.

[4] T. L. Olander and C. S. Velden, "The Advanced Dvorak Technique: Continued Development of an Objective Scheme to Estimate Tropical Cyclone Intensity Using Geostationary Infrared Satellite Imagery," *Weather Forecast.*, vol. 22, no. 2, pp. 287–298, Apr. 2007.

[5] M. F. Piñeros, E. A. Ritchie, and J. S. Tyo, "Estimating tropical cyclone intensity from infrared image data," *Weather Forecast.*, vol. 26, no. 5, pp. 690–698, 2011.

[6] E. A. Ritchie, G. Valliere-Kelley, M. F. Piñeros, and J. S. Tyo, "Tropical cyclone intensity estimation in the North Atlantic basin using an improved deviation angle variance technique," *Weather Forecast.*, vol. 27, no. 5, pp. 1264–1277, 2012.

[7] E. A. Ritchie, K. M. Wood, O. G. Rodríguez-Herrera, M. F. Piñeros, and J. S. Tyo, "Satellite-Derived Tropical Cyclone Intensity in the North Pacific Ocean Using the Deviation-Angle Variance Technique," *Weather Forecast.*, vol. 29, no. 3, pp. 505–516, Dec. 2013.

[8] G. Fetanat, A. Homaifar, and K. R. Knapp, "Objective Tropical Cyclone Intensity Estimation Using Analogs of Spatial Features in Satellite Data," *Weather Forecast.*, vol. 28, pp. 1446–1459, Dec. 2013.

[9] N. Jaiswal, C. M. Kishtawal, and P. K. Pal, "Cyclone intensity estimation using similarity of satellite IR images based on histogram matching approach," *Atmospheric Res.*, vol. 118, no. Supplement C, pp. 215–221, Nov. 2012.

[10] K. R. Knapp and J. P. Kossin, "New global tropical cyclone data from ISCCP B1 geostationary satellite observations," *J Appl Remote Sens*, vol. 1, p. 13505, 2007.

[11] Y. Zhao, C. Zhao, R. Sun, and Z. Wang, "A Multiple Linear Regression Model for Tropical Cyclone Intensity Estimation from Satellite Infrared Images," *Atmosphere*, vol. 7, no. 3, p. 40, Mar. 2016.

[12] K. R. Knapp, M. C. Kruk, D. H. Levinson, H. J. Diamond, and C. J. Neumann, "The international best track archive for climate stewardship (IBTrACS) unifying tropical cyclone data," *Bull. Am. Meteorol. Soc.*, vol. 91, no. 3, pp. 363–376, 2010.

[13] B. Craven and S. M. Islam, *Ordinary least squares regression*. Sage Publications, 2011.

[14] D. Basak, S. Pal, and D. C. Patranabis, "Support vector regression," *Neural Inf. Process.-Lett. Rev.*, vol. 11, no. 10, pp. 203–224, 2007.

[15] P. J. Rousseeuw and A. M. Leroy, *Robust regression and outlier detection*, vol. 589. John wiley & sons, 2005.

[16] N. Aronszajn, "Theory of reproducing kernels," *Trans. Am. Math. Soc.*, vol. 68, no. 3, pp. 337–404, 1950.

[17] H. Q. Minh, P. Niyogi, and Y. Yao, "Mercer's theorem, feature maps, and smoothing," in *COLT*, 2006, vol. 6, pp. 154–168.

[18] M. Ring and B. M. Eskofier, “An Approximation of the Gaussian RBF Kernel for Efficient Classification with SVMs,” *Pattern Recogn Lett*, vol. 84, no. C, pp. 107–113, Dec. 2016.
[19] S. S. Haykin, *Neural Networks and Learning Machines*. Prentice Hall, 2009.
[20] T. Chen and C. Guestrin, “XGBoost: A Scalable Tree Boosting System,” *ArXiv160302754 Cs*, pp. 785–794, 2016.
[21] F. Chollet and others, *Keras*. 2015.
[22] T. Chai and R. R. Draxler, “Root mean square error (RMSE) or mean absolute error (MAE)?,” *Geosci. Model Dev. Discuss.*, vol. 7, pp. 1525–1534, 2014.
[23] S. CIMSS, “The CIMSS Satellite Consensus,” *The CIMSS Satellite Consensus*, 2018. [Online]. Available: http://tropic.ssec.wisc.edu/misc/satcon/info.html. [Accessed: 05-Mar-2018].
[24] C. W. Landsea and J. L. Franklin, “Atlantic Hurricane Database Uncertainty and Presentation of a New Database Format,” *Mon. Weather Rev.*, vol. 141, no. 10, pp. 3576–3592, Oct. 2013.

LIST OF TABLES

501 Table 1- Intensity distribution of images used in the study (C1 to C5 correspond to category of the
502 hurricane).

Category	**Number of Images**
Pre-Developmental (< 20 kt)	82
Tropical Depression (20-34 kt)	1,617
Tropical Storm (35-64 kt)	2,088
Hurricane: C1	399
Hurricane: C2	183
Hurricane: C3	210
Hurricane: C4	95
Hurricane: C5	2
Total	**5,531**

503

Table 2- Formulae for computation of statistical features

Statistic	**Formula**
Mean	$\bar{v} = \frac{1}{n}\left(\sum_{i=1}^{n} v_i\right)$
Standard Deviation	$s = \sqrt{\frac{\sum_{i=1}^{n}(v_i - \bar{v})}{n-1}}$
Entropy	$H(v) = -\sum_{i=1}^{n} p(v_i)\log_{10} p(v_i)$ $p(v_i)$ is the probability of v_i based on its relative frequency or counts of occurrence.

Table 3- Comparison between RMSE values for Leave One Hurricane Out cross validation for different machine learning models used in this work with statistical and Deviation Angle Variance features and zero-order predictors with and without post-processing.

Method	**Mean RMSE (kt)**	**Mean RMSE after smoothing**
PHURIE: SVR	**11.2**	**9.5**
PHURIE: OLS	12.8	10.5
PHURIE: BPNN	12.0	10.1
PHURIE: XGBoost	11.3	9.8
Baseline Predictor (Mean)	24.3	-

Table 4- Comparison of results using our method and deviation angle variation based method for the same hurricanes as in (Piñeros et al., 2011). This table shows results on leaving certain hurricanes out for testing and training on the remaining ones from 2004-2008.

Method	**RMSE (kt)**	**RMSE (kt) after smoothing**
PHURIE: SVR	**11.5**	**9.8**
PHURIE: OLS	12.2	10.2
PHURIE: BPNN	11.5	10.0
PHURIE: XGBoost	11.6	9.9
Deviation Angle Variation Technique[5]	14.7	-

Table 5- Comparison of results using our method and deviation angle variation based method (Piñeros et al., 2011). In this experiment, hurricane data of years 2004-2008 was used for training and data of year 2009 was used for testing.

Method	**RMSE (kt)**	**RMSE after smoothing**
PHURIE: SVR	13.6	12.1
PHURIE: OLS	**13.4**	**11.5**
PHURIE: BPNN	13.2	11.6
PHURIE: XGBoost	13.5	12.0
Deviation Angle Variation Technique [5]	24.8	-

Table 6- Comparison among different methods for Leave One Year Out cross-validation

Year ↓	**Method**			
	SVR	**OLS**	**BPNN**	**XGBoost**
2004	12.7	14.2	15.0	**12.6**
2005	**10.2**	11.3	10.3	11.5
2006	10.3	10.4	11.1	**10.2**
2007	**9.7**	11.1	10.8	9.9
2008	11.6	**11.4**	12.3	11.8
2009	12.1	**11.5**	11.9	12.1
Mean →	**11.1**	11.7	11.9	11.4

526 Table 7- Comparison with DAV and improved DAV technique for Leave One Year Out

527 Experiment for intensities higher than 34 kt.

Year	Method					
	DAV [5]	**Improved DAV** [6]	**SVR**	**OLS**	**BPNN**	**XGBoost**
2004	15.6	13.3	13.9	13.7	**12.0**	12.9
2005	17.3	14.1	**9.8**	10.6	9.9	11.6
2006	11.7	**10.3**	11.1	11.1	11.2	10.9
2007	12.8	11.4	10.5	11.5	11.4	**10.0**
2008	12.2	12.0	10.3	**9.9**	10.2	10.5
2009	17.9	10.6	12.7	**10.9**	11.4	11.4
Mean →	14.6	12.0	11.3	11.3	**11.0**	11.2

528

529

Table 8- RMSE values for Leave One Year Out cross validation over images from near IR channel

Years	SVR	OLS	BPNN	XGBoost
2004	14.3	15.7	15.2	**14.2**
2005	**11.1**	13.0	12.2	11.6
2006	10.8	**10.5**	11.5	11.1
2007	**10.3**	12.3	11.8	11.3
2008	14.5	13.9	**12.1**	11.5
2009	13.9	**12.5**	15.0	14.3
Mean	12.5	13.0	13.0	**12.3**

Table 9- RMSE values for Leave One Year Out cross validation over images from Water Vapor Channel (IRWVP)

Years	SVR	OLS	BPNN	XGBoost
2004	15.6	**14.6**	17.8	18.0
2005	**11.6**	13.7	15.9	12.0
2006	**9.1**	10.8	11.1	10.4
2007	11.2	12.9	14.9	**11.1**
2008	12.0	11.3	**11.0**	11.6
2009	14.5	15.6	13.5	**12.3**
Mean	**12.3**	13.15	14.03	12.6

534 Table 10- RMSE values for Leave One Year Out cross validation over images from visible channel

Years	SVR	OLS	BPNN	XGBoost
2004	23.2	23.8	22.4	**21.6**
2005	17.9	18.5	17.9	**16.7**
2006	12.9	12.6	12.8	**12.1**
2007	**17.6**	19.6	20.5	20.5
2008	21.4	16.2	**15.5**	15.8
2009	24.4	23.2	45.7	**18.5**
Mean	19.6	19.00	22.5	**17.6**

535

536

537

538 **LIST OF FIGURES**

554

555

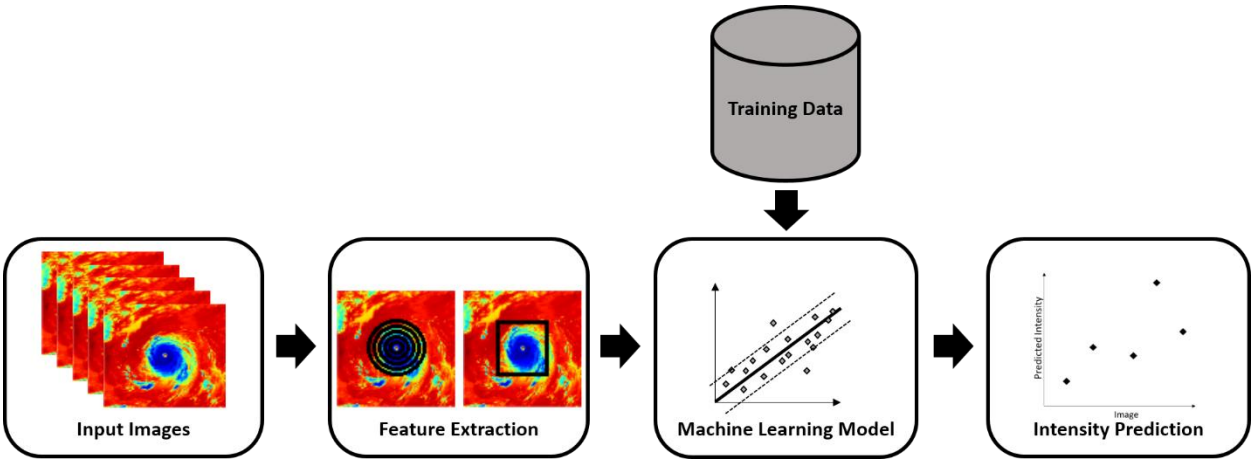
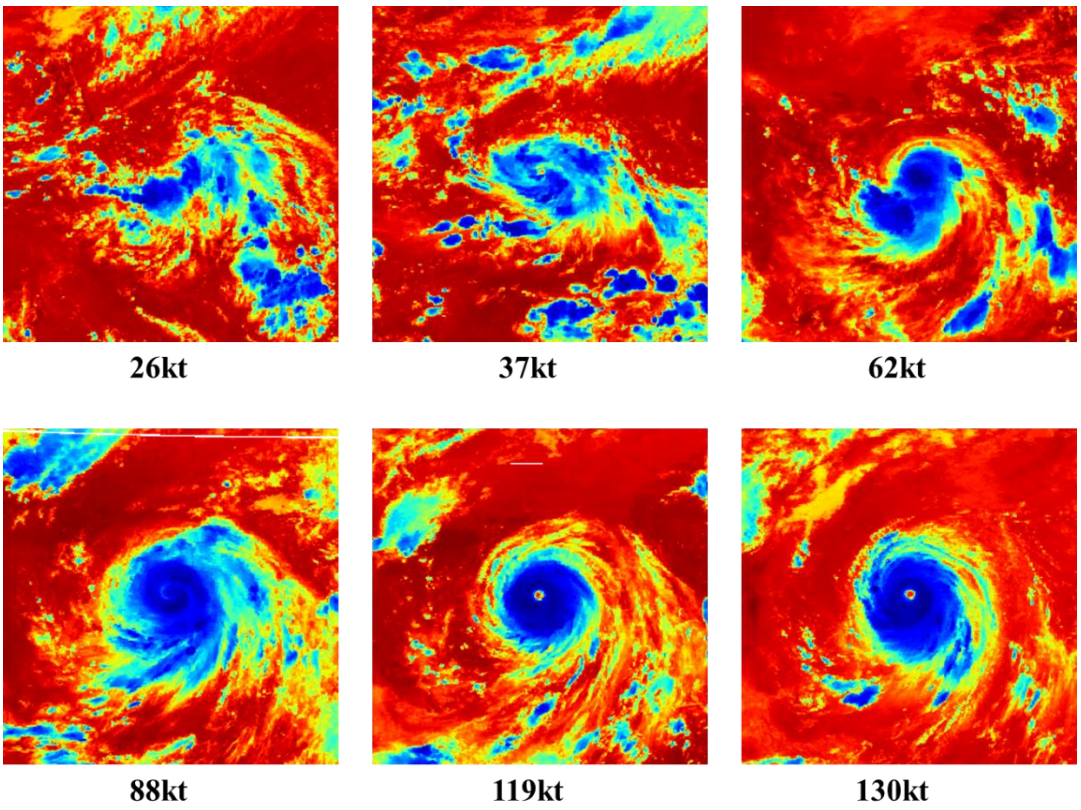


Figure 1- Illustration of workflow of the proposed system

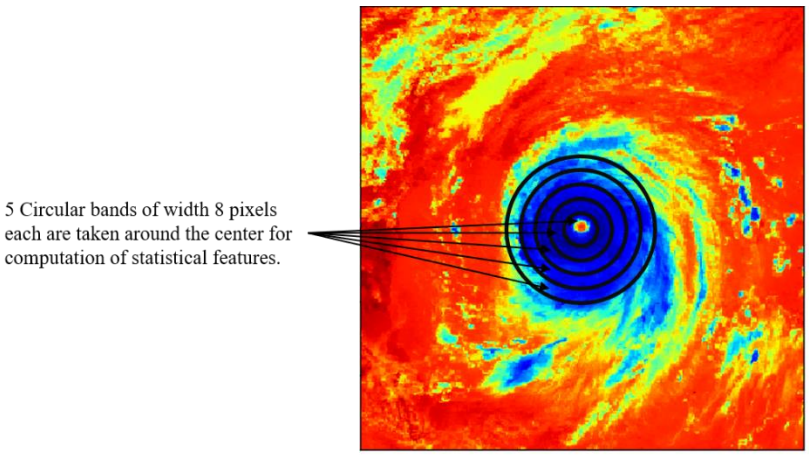


559

560 Figure 2- Images for Hurricane Katrina (2005).

561 It can be seen that the cloud gets organized to a circular structure as the intensity increases.

562

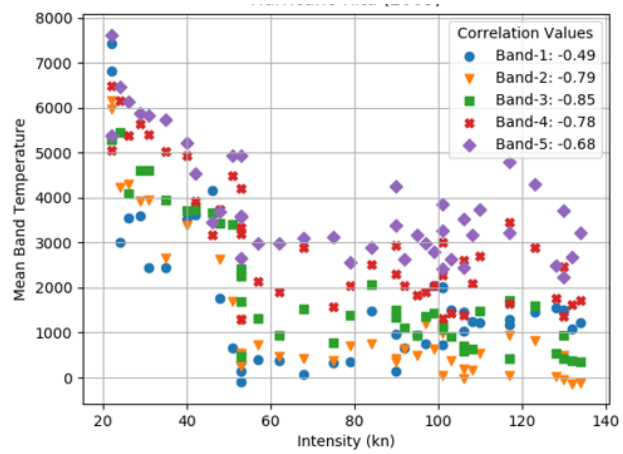


563

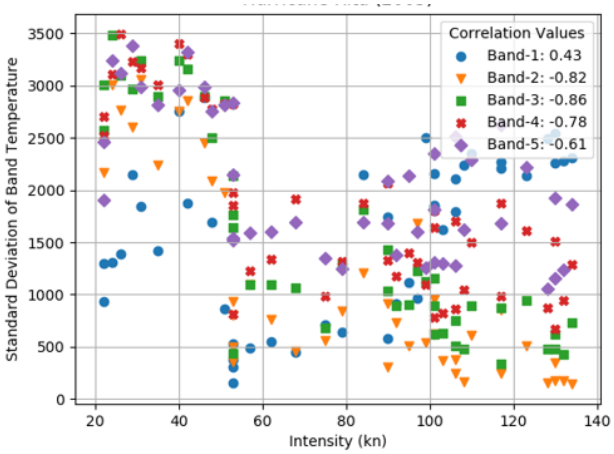
564 Figure 3- Central region of an image is divided into circular bands for computing statistical

565 features.

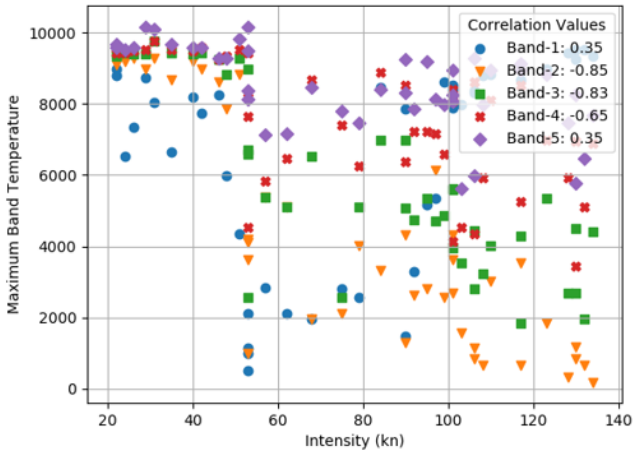
566


Correlation Values
Band-1: -0.49
Band-2: -0.79
Band-3: -0.85
Band-4: -0.78
Band-5: -0.68
Mean Band Temperature
Intensity (kn)

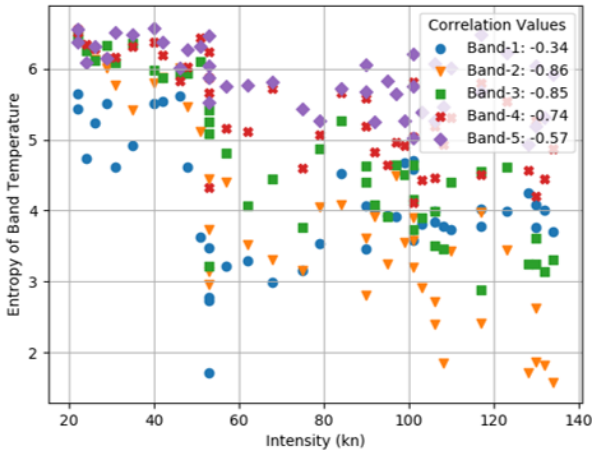
(a)


Correlation Values
Band-1: 0.43
Band-2: -0.82
Band-3: -0.86
Band-4: -0.78
Band-5: -0.61
Standard Deviation of Band Temperature
Intensity (kn)

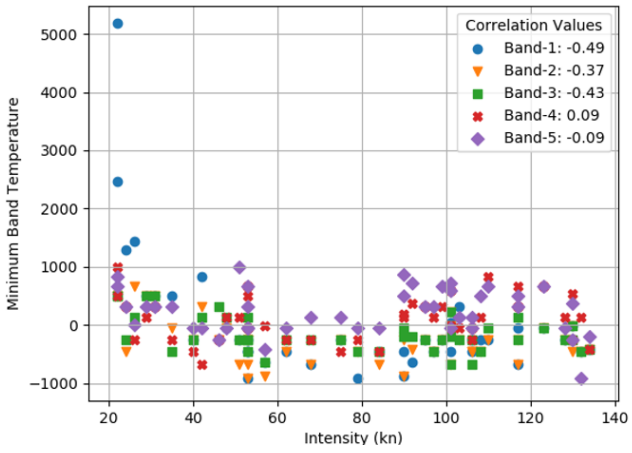
(b)


Correlation Values
Band-1: 0.35
Band-2: -0.85
Band-3: -0.83
Band-4: -0.65
Band-5: 0.35
Maximum Band Temperature
Intensity (kn)

(c)


Correlation Values
Band-1: -0.34
Band-2: -0.86
Band-3: -0.85
Band-4: -0.74
Band-5: -0.57
Entropy of Band Temperature
Intensity (kn)

(d)

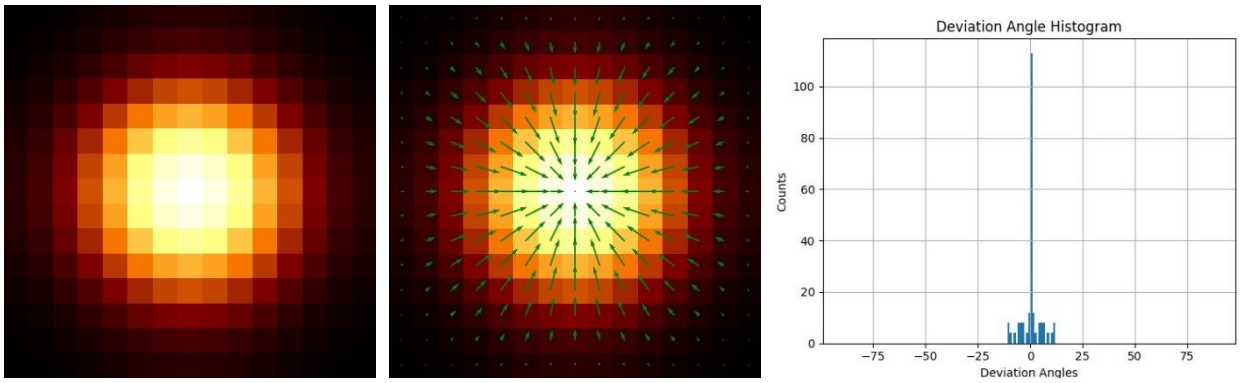

Correlation Values
Band-1: -0.49
Band-2: -0.37
Band-3: -0.43
Band-4: 0.09
Band-5: -0.09
Minimum Band Temperature
Intensity (kn)

(e)

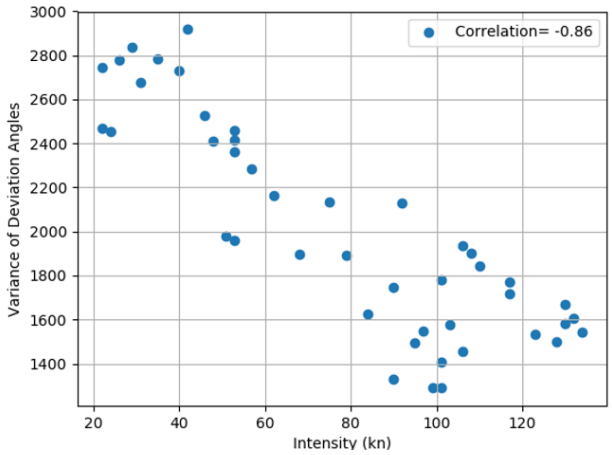
567

Figure 4- Statistical features plotted against intensity values for images from Hurricane Rita (2005).

Mean (a), Standard Deviation (b), Maximum (c), Entropy (d) and Minimum (e) of the band temperatures have been used as features. A high correlation for most of the bands in (a)-(d) can be seen. The correlation between minimum band temperatures (e) and intensities is low, showing this feature may not be very informative.



(a) **(b)** **(c)**



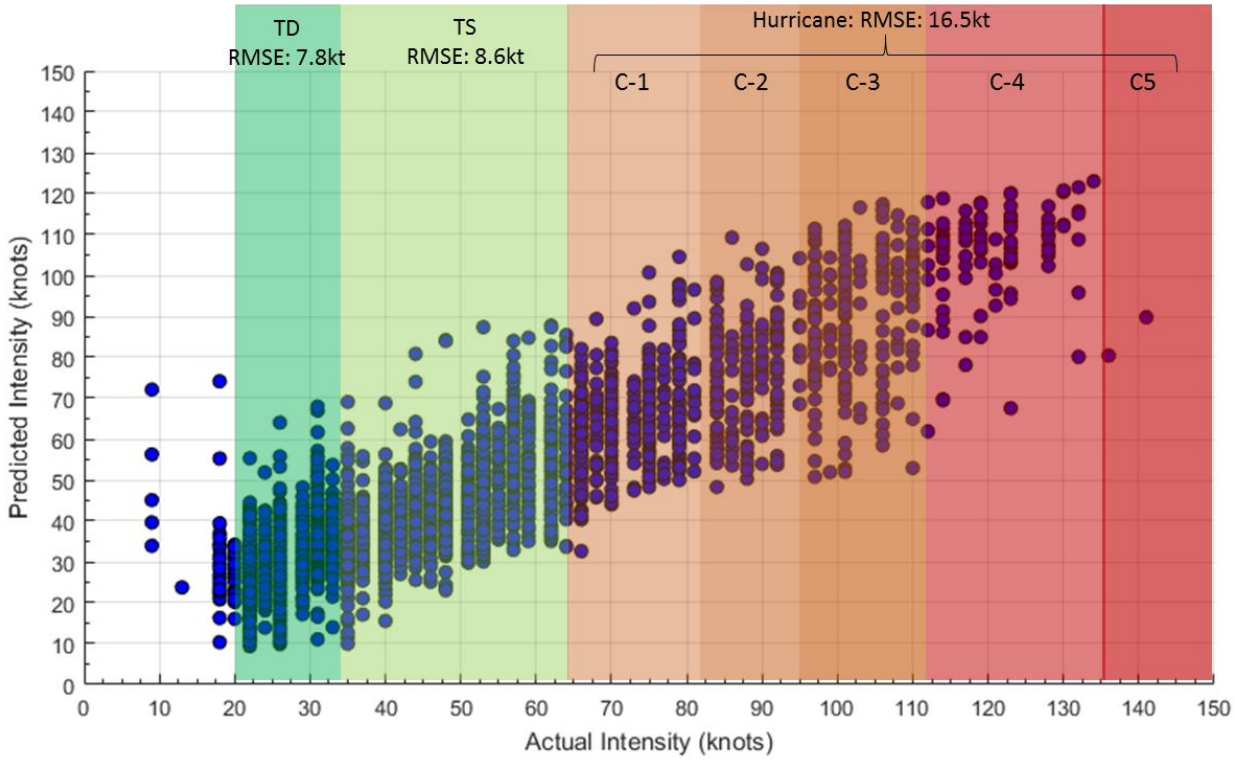
(d)

576 Figure 5- Illustration of concept of Deviation Angles.

577 (a) shows a test image exhibiting a circular structure. (b) shows gradient vectors for each pixel.
578 Most of the vectors are directed towards the center, hence the angles between the gradient vectors
579 and the lines joining other pixels with the center are mostly zero. (c) shows a histogram of deviation
580 angles for the image shown in (a). (d) presents a plot of deviation angle variance against intensity
581 values for Hurricane Rita (2005). A high correlation can be seen for deviation angle variance,
582 making it an informative feature.

583

584



585

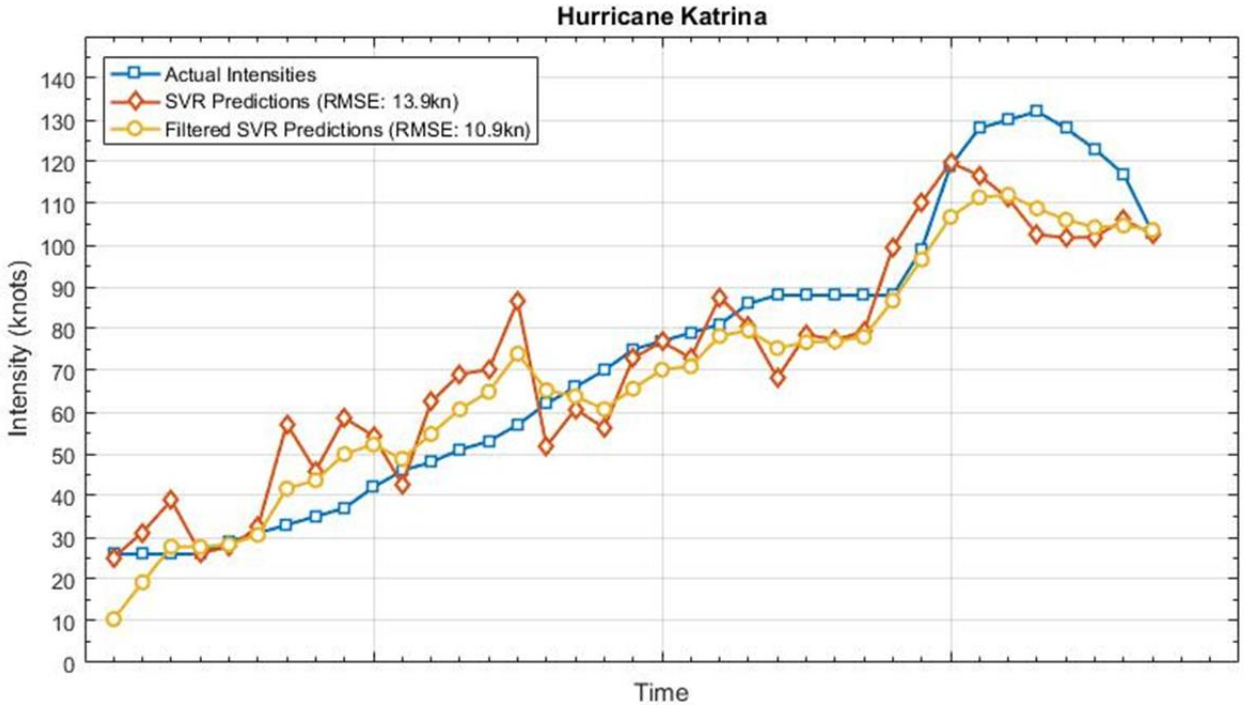
586 Figure 6- Plot of actual vs. SVR-predicted intensities of all test hurricanes in leave one hurricane

587 out cross validation using SVR.

588 Different shades represent different categories of storms based on their intensities: Tropical

589 Depression (TD), Tropical Storm (TS) and Categories 1-5 Hurricanes.

590



591

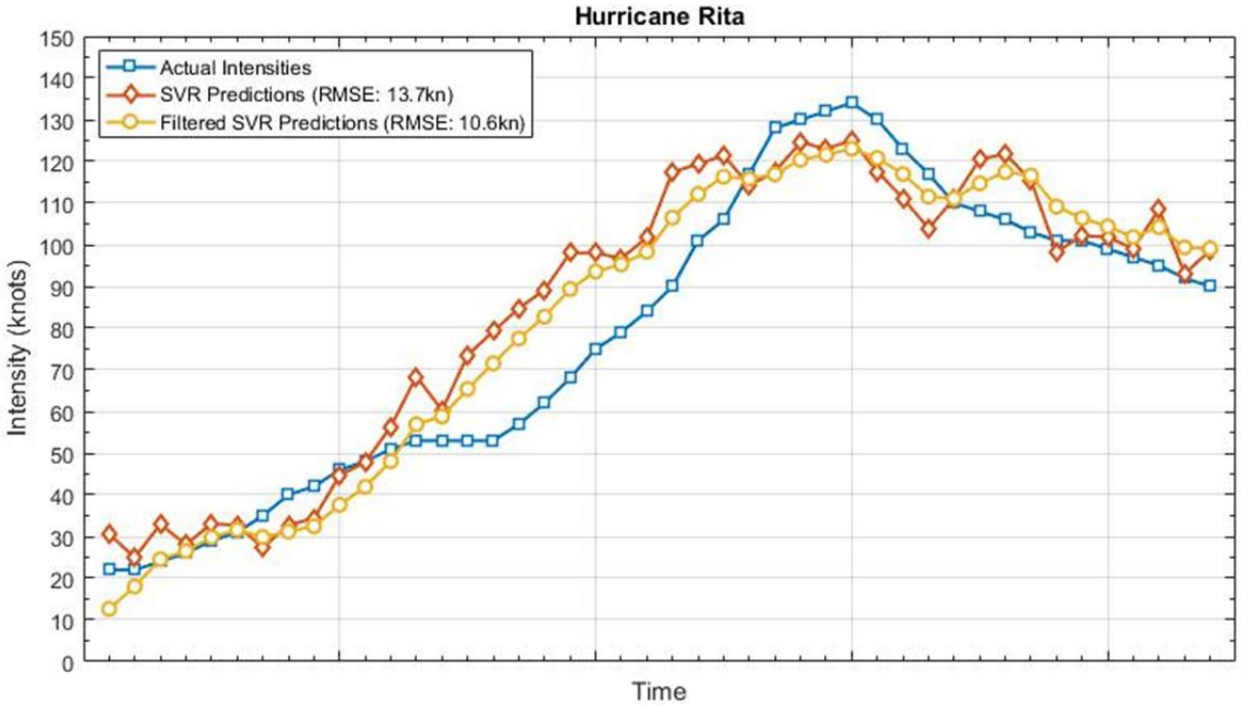
592 Figure 7- Actual and predicted intensity values for Hurricane Katrina (2005).

593 The RMSE values obtained for SVR predictions before and after filtering are 13.9 kt and 10.9kt

594 respectively.

595

596



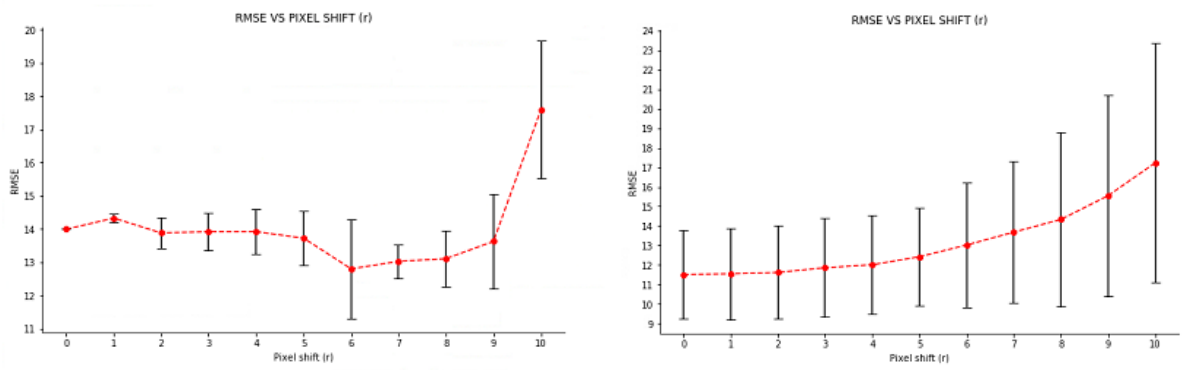
597

598 Figure 8- Actual and predicted intensity values for Hurricane Rita (2005).

599 The RMSE values obtained for SVR predictions before and after filtering are 13.7 kt and 10.6 kt

600 respectively.

601



(a) (b)

602 Figure 9 Effect of hurricane center annotation errors on root mean square error (RMSE) in
603 predicted intensity using leave one TC out cross-validation. (a) Plot for pixel shift vs. RMSE for
604 Rita (2005) and, (b) Combined plot for pixel shift vs. RMSE for Alex (2004), Rita (2005), Gordon
605 (2006), Felix (2007), Bertha (2008) and Ana (2009).

606