An Experimental Comparison of Different Classifiers for Predicting Tropical Cyclone Rapid Intensification Events

Hadil Shaiba[1] and Michael Hahsler[2]

***Abstract*—** Accurately predicting the intensity and track of a Tropical Cyclone (TC) can save the lives of people and help to significantly reduce damage to property and infrastructure. Current track prediction models outperform intensity models which is partially due to the existence of rapid intensification (RI) events. RI appears in the lifecycle of most major hurricanes and can be defined as a change in intensity within 24 hours which exceeds 30 knots. Improving the predicting of RI events has been identified as one of the top priority problems by the National Hurricane Center (NHC). In this paper we compare the RI event prediction performance of several popular classification methods: Logistic regression, naive Bayes classifier, classification and regression tree (CART), and support vector machine. The dataset used is derived from the data used by the Statistical Hurricane Intensity Prediction Scheme (SHIPS) model for intensity prediction which contains large-scale weather, ocean, and climate condition predictors from 1982 to 2011. 10-fold cross validation is applied to compare the models. The probability of detection (POD) and false alarm ratio (FAR) are used to measure performance. Predicting RI events is a difficult problem but initial experiments show potential for improving forecast using data mining and machine learning techniques.

***Keywords*—**Classification, Data Mining, Rapid Intensification, Tropical Cyclone.

I. Introduction

A Tropical Cyclone (TC) is a weather phenomenon that forms in tropical oceans and creates strong winds and heavy rain. Its intensity is measured by its maximum wind speed. Every year during hurricane season, tens of tropical cyclones form in the Atlantic basin and some of them become major hurricanes which severely affect populated areas. Hurricane Katrina in 2005 is an example of such a hurricane. Its maximum sustained wind reached 175 mph which was classified as category 5 in the Saffir-Simpson Hurricane Scale [2]. The hurricane made 3 landfalls and killed over 1,833 people which made it the deadliest hurricanes in the United States. It also caused billions of dollars in damages [1]. The hurricane season starts in June, ends in November and reaches its peak at the end of August. Affected areas are close to the equator [3].

Different statistical, dynamical, and statistical-dynamical models have been developed to predict the hurricane track and intensity. Some models can predict the behavior of the storm several days ahead. Intensity models have not improved as much as track models. One reason is due to the existence of rapid intensification (RI) events. A RI event can be defined as the change in intensity increase of 30 knots or greater within 24 hours [6]. Improved predictions of RI have been identified as NHC top priority [6]. The reason why predicting RI events is challenging, is because they are rare and their causes are still not very clear.

The rapid intensification index (RII) [6] is a regression-based statistical model used to forecast RI events based on large-scale weather, ocean, and climate condition predictors which are derived from the Statistical Hurricane Intensity Prediction Scheme (SHIPS) hurricane intensity prediction model. The RII model is operationally used by NHC to predict RI events in real time. The RII model was examined for each of the following intensity change thresholds: 25, 30, and 35 knots and the used predictors were chosen based on previous studies. The experiments have shown that among these predictors the change in intensity over to the previous 24hours, upper-level divergence, and vertical shear have the strongest influence on the appearance of an RI event for the Atlantic basin [6]. There is significant potential for improvements since the probability of detection (POD) in the Atlantic basin ranges between 15% and 59% whereas the false alarm ratio (FAR) is very high ranging between 71% and 85% [6].

Classification is a supervised learning technique used to classify an instance into one of a set of available classes [4]. Supervised learning means that the classes of the data points in the training dataset are known and are used for model building. For the RI problem, the value of the class attribute is either rapid intensification (RI) or not rapid intensification (NRI). Since RI is a classification problem, the goal of this paper is to apply and compare different popular classification techniques on a historic dataset.

In this paper, we discuss the RII model and experimentally compare it with a set of popular classification methods.

II. Tropical Cyclone Dataset

The TC dataset includes a set of historic TCs that happened from 1982 to 2011 in the Atlantic basin. Each TC has a unique identifier. The rows are data points that represent the lifecycle of a TC every 6 hours and include a set of predictors. The

Hadil Shaiba[1] is with Southern Methodist University, Dallas, TX 75275-0122 USA (corresponding author's phone: +1(405)410-5105; e-mail: hshaiba@smu.edu).

Michael Hahsler[2] is with Southern Methodist University, Dallas, TX 75275-0122 USA (e-mail: mhahsler@lyle.smu.edu).

predictors describe weather, persistence, atmosphere, ocean, and climate conditions such as the maximum intensity and wind shear. Weather and climate conditions are gathered in various ways. Satellites, buoys, and radars are examples. Different sophisticated models estimate the values of the predictors. Table 1 represents the predictors' minimum and maximum values and their description. The description is gathered from [5] and [7]. The RII model estimate of RI with a 30 knots threshold is derived from the 2010 SHIPS model and is added to the dataset as the feature RII. The TCs dataset can be found at http://lyle.smu.edu/IDA/data/storms.

TABLE I
SHIPS (2010) PREDICTORS: RANGE AND DESCRIPTION

Predictor	Minimum	Maximum	Description
LAT	7.2	51.9	Latitude in degrees North * 10 vs. time
LON	-109.3	-6	Longitude in degrees West * 10 vs. time
VMAX	15	160	Maximum intensity in knots
PER	-45	75	Previous 12-h change in maximum intensity in knots
ADAY	0	1	Absolute value of Julian day
SPDX	-29	48.034	Zonal component of initial storm motion. (-) is the wind movement from east to west whereas (+) is the opposite
PSLV	246	1124	Vertical depth
VPER	-4025	11250	VMAX x PER
PC20	0	100	Percentage of cold cloud-top (GOES brightness temperature ≤ -20°C)
GSTD	0	2822	VMAX x standard deviation of GOES brightness temperature
POT	-30.14	151.62	Along-track average of empirically estimated maximum potential intensity(MPI) minus VMAX at time 0
SHDC	0.5	74.5	850 mb shear magnitude vs. time
T200	-60.9	-45.7	Average 200 mb temperature within 1000 km of storm center
T250P	-8	0	Average 250 mb temperature within 1000 km of storm center
EPOS	0	22.35	200-800 km average theta difference between a parcel lifted and environment vs. time
RHMD	19.5	84.5	700-500 mb relative humidity
TWAT	-6	7.9	GFS model mean tangential wind
Z850	-162.5	278	850-hpa absolute vorticity
D200	-95	205	200-pHa divergence
LSHDC	0.2245	54.5737	SHDC x sine of LAT
VSHDC	32.5	5960	SHDC x VMAX
POT2	$1.000e^{-05}$	$2.30e^{04}$	Square of POT
RHCN	0	94	Ocean heat content (KJ/cm2)
SDIR	-86.37526	86.59061	Reference direction for shear direction
SHGC	-14.0022	19.1011	Generalized shear magnitude vs time

For the study in this paper we added derived features including the changes in intensity (persistence) over the previous 6, 18, and 24 hours and VMAX (intensity) multiplied by the previous changes in maximum intensity. The storm's day of the year was also added. The derived predictors are shown in Table 2.

TABLE 2
DERIVED PREDICTORS: RANGE AND DESCRIPTION

Predictor	Minimum	Maximum	Description
PER6	-65	55	Previous 6-h change in VMAX
PER18	-115	85	Previous 18-h change in VMAX
PER24	-110	95	Previous 24-h change in VMAX
VPER6	-4675	7150	VMAX x PER6
VPER18	-4275	13600	VMAX x PER12
VPER24	-4750	15200	VMAX x PER18
YDAYS	1	365	Day of the year
VPC20	0	16000	PC20 x VMAX

Once the storm hits the land the intensity of the storm decays rapidly [5]. Therefore, the overland cases were removed from the dataset.

The class attribute is derived from the changes in intensity features. It has two values; either RI or NRI. The data point is classified as RI if the change in intensity over 24 hours is 30 knots or greater otherwise the data point is classified as NRI.

III. MODEL COMPARISON

A. Cross Validation

There are different ways to test how well a predictive model performs. For hurricane prediction it is common to predict the values for a year (a hurricane season) using the data from all previous years. This mimics the real application where a model trained with all available data is used for the current season. However, to get more accurate results, we have decided to use the standard evaluation method in data mining called 10-fold cross validation.

We randomly split the dataset into 10 distinct sets (folds), each containing the same number of hurricanes. In each round, one fold is used for testing and the remaining 9 folds are used for training. This guarantees that all data is used at least once for testing. Finally, the ten results are averaged.

B. Performance Metric

The predictions of a classification model are used to compute true positives (TP), true negatives (TN), false positives (FP) and false negatives (FN) as shown in confusion table in Table 3. Based on the counts we calculate the probability of detection (POD) and false alarm ratio (FAR) and plot them similar to a Receiver Operating Characteristic (ROC) [14] curve which plots sensitivity versus specificity. We use POD and FAR because the RII model uses them for evaluation and we compare our results to their results.

We use prediction models which produce a probability of class membership so we can use different thresholds to adjust the tradeoff between POD and FAR.

TABLE 3
CONFUSION TABLE

Predicted/Actual	RI	NRI
RI	TP	FP
NRI	FN	TN

POD is the ratio of the correct forecasts of RI occurrences to the actual number of RI occurrences while FAR is the number of incorrect forecasts of RI divided by the total number of RI forecasts [6]. The greater the value of POD and the lower the value of FAR the better the model performs. The POD and FAR values are calculated as shown in (1) and (2) [4]. Sensitivity used in the ROC curve is the true positive rate (TPR) and is equivalent to POD whereas specificity is the false positive rate (FPR) and is computed as shown in (3) [14].

$$POD = \frac{TP}{(TP + FN)} \tag{1}$$

$$FAR = \frac{FP}{(FP + TP)} \tag{2}$$

$$FPR = \frac{FP}{(FP + TN)} \tag{3}$$

An example of a storm with an RI event is shown in Figure 1 (from time 120 through 168). The way the true class is found is shown in Table 4. All observations for with the intensity increases over the next 24 hours by 30 or more knots (see marked values in row Change in Table 4) are part of the RI event. In addition 3 observations (i.e., 24 hours) after the last change that exceeded 35 knots are also part of the event since during that time the intensity also increased by more than 30 knots. The classification algorithm will predict a class label. Based on Table 3 and the true class we determine if the prediction is a TP, TN, FP, or a FN.

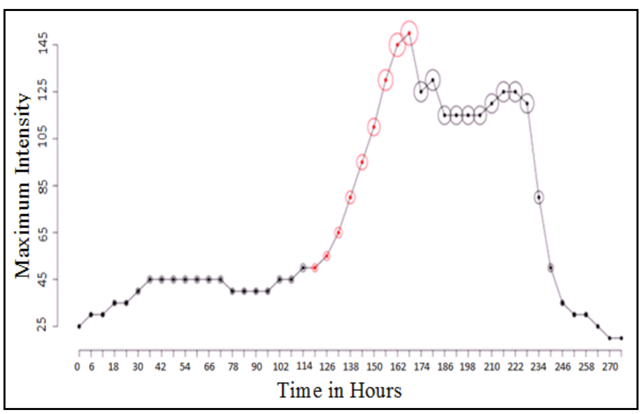

Fig. 1 An example of a RI event maximum intensity versus time where the size of the circle expresses the maximum intensity.

IV. Rapid Intensification Index (RII)

A. Description

An updated version of the RII model is currently used by the NHC. The predictors (PER, SHRD, RHLO, and POT) are identical to those used in earlier versions but the way some were measured has been slightly altered. The predictors D200, PX30, SDBT, and OHC were added. Some predictors were averaged from time t=0 to t=24 instead of only using time t=0 [6]. The set of additional predictors that were not mentioned in Table 1 are described in Table 5 [6].

TABLE 4

RI EVENT EXAMPLE

Time in Hours	108	114	120	126	132	138	144	150	156	162	168	174	180
VMAX	45	45	50	50	55	65	80	95	110	130	145	150	125
Change	5	10	**30**	**45**	**55**	**65**	**65**	**55**	15	-45	-55	-60	-45
Class	NRI	NRI	RI	RI	RI	RI	RI	RI	RI	RI	RI	NRI	NRI

TABLE 5

RII MODEL PREDICTORS

Predictor	Description
SHRD	850-200 mb shear magnitude vs. time
RHLO	850-700 mb relative humidity vs. time
PX30	Percentage of cold cloud-top (GOES brightness temperature ≤ -30°C)
SDBT	Standard deviation of cold cloud-top brightness temperature
OHC	Ocean heat content

B. Training

The magnitude of each predictor is scaled between 0 and 1 where 0 indicates no or minimal influence on RI and 1 indicates maximal influence. The weighted sum of the scaled magnitude values S_p for the eight predictors p is computed to produce a RI score R_S for each observation [4]:

$$R_S = \sum_{i=1}^{n} W_i S_{P_i}, \tag{4}$$

where n is the number of predictors and W is the weight of each predictor obtained from linear discriminant analysis.

The scores are split into 4 quartiles such that each quartile has an equal number of RI events regardless of the number of NRIs. For each quartile, the number of RI events is divided by the total number of RIs+NRIs that appeared in that particular quartile to estimate the probability of RI in each quartile [6].

C. Testing Result

To test the model, the score (i.e., discriminant value) R_s of each data point in the test set is computed. The matching quartile is found and the probability is estimated using linear interpolation within the quartile. We use the SHIPS code (version 2010) provided by Dr. Mark DeMaria for the calculation. For classification we use a threshold on the score, regarding scores above the threshold as RI and below NRI. The results of 10-fold cross validation with varying thresholds are shown in Figure 2. Each line connects the classification results obtained at different thresholds for one of the 10 validation runs. The thicker blue line in the center represents the mean of the 10 runs.

I. Classification models

For all classification models we only use the original and derived features discussed in section II and not the features added for the RI index discussed in the previous section.

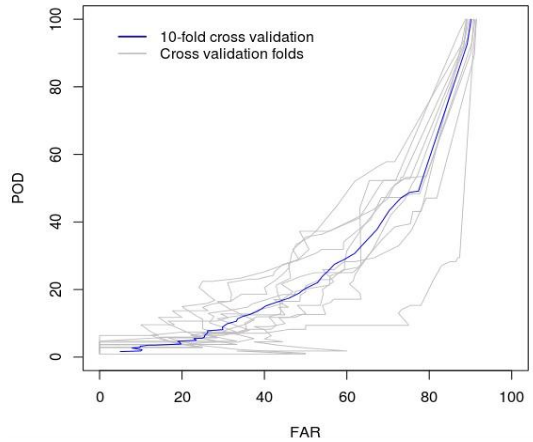

Fig. 2 Results of RII model POD and FAR using 10-fold cross validation.

A. Logistic Regression

1) Description

Logistic regression is used to predict a categorical variable such as the class label. Logistic regression models the log of the odds ratio (logit) of a class label as a linear regression on a set of covariates in the following form [16]:

$$\ln\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_1 x_1 + \dots + \beta_i x_i \tag{5}$$

This formulation guarantees that p_i is between 0 and 1 and can be used as a probability of a class label (in our case RI). A threshold on this probability can be used to extract a classification.

2) Training

We used R [15] for logistic regression.

3) Testing Results

Figure 3 is a representation of the results of POD and FAR using 10-fold cross validation.

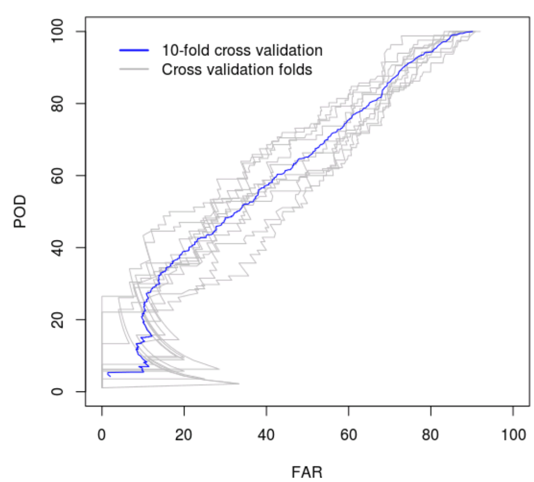

Fig. 3 Results of regularized logistic regression POD and FAR using 10-fold cross validation.

B. Naïve-Bayes Classifier

1) Description

The Bayes Theorem shown in (6) is used to estimate a posterior probability given the class-conditional probability *P(X|Y)* and new evidence *P(X)*:

$$P(Y \mid X) = \frac{P(Y)P(X \mid Y)}{P(X)} \tag{6}$$

The class-conditional probability in the Bayes Theorem can be computed using different Bayesian classifiers such as naïve-Bayes and belief network classifiers. Naïve-Bayes is a simple classifier which assumes that the predictors are conditionally independent whereas belief networks are more complicated and are directed acyclic graphs that include nodes representing the relationships between features [8].

The naïve Bayes classifier is chosen for its simplicity, capability of efficient computation, and ability to perform well even if the predictors are not independent in reality and especially when the number of predictors is large [9]. It works with categorical and, continuous features. Under the assumption of independence, the class-conditional probability for each feature is computed as shown in (7) [8].

$$P(X \mid Y = y) = \prod_{i=1} P(X_i \mid Y = y) \tag{7}$$

If the values are categorical then, the class-conditional probability $P(X_i|Y=y)$ is estimated by counting the number of times X_i and class y co-occur divided by the total number of times class y occurs. Whereas if the values are continuous, they can either be transformed into categories then the probability is estimated as described above or a distribution is chosen to estimate the probability. Typically a Gaussian distribution is chosen where the mean μ and variance σ^2 are computed and the probability is estimated as shown in (8) [8].

$$P(X_i = x_i \mid Y = y) = \frac{1}{\sqrt{2\Pi}\,\sigma_{ij}} \exp^{-\frac{(x_i - \mu_{ij})^2}{2\sigma_{ij}^2}} \tag{8}$$

After calculating the class-conditional probability, the posterior probability can be easily computed using (9) [8].

$$P(Y \mid X) = \frac{P(Y)\prod_{i=1}^{d} P(X_i \mid Y)}{P(X)} \tag{9}$$

2) Training

We used a package in R called **e1071** [10] that includes a naive-Bayes classifier. The prior probabilities *P(X)* of RI and NRI are shown in Table 6 which shows that there is strong class imbalance between RI and NRI where most of the records belong to class NRI. This is because RI cases are rare events. There are several available methods that can be applied to solve this problem (e.g., resampling), but this is outside the scope of this paper.

TABLE 6
RI AND NRI PRIOR PROBABILITIES

P(RI)	P(NRI)
0.90076046	0.09923954

3) Testing Result

Figure 4 is a representation of the results of POD and FAR using 10-fold cross validation.

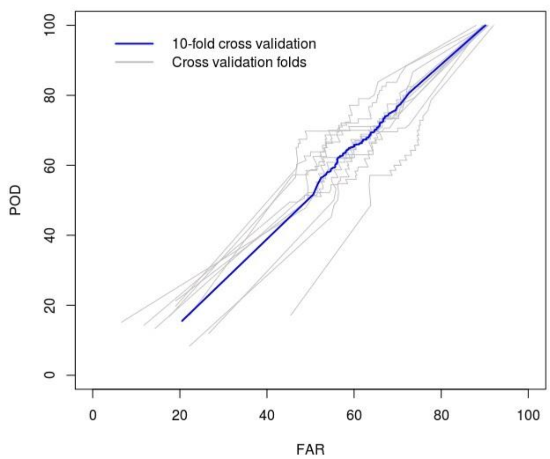

Fig. 4 Results of naïve Bayes classifier POD and FAR using 10-fold cross validation.

C. Classification and regression tree (CART)

1) Description

A decision tree is a simple technique that is used widely and has proven to work well for many problems. They can be used for classification (classification) or regression (regression tree) [8]. We use a regression tree to predict the probability of RI.

To decide about which class a certain object belongs to, the object has to go through several test conditions. The tree's root and internal nodes represent these test conditions while the leaf nodes represent the classes. The test conditions are based on the features such as the wind speed and shear. The tree is built based on the training dataset and to classify a new data point a series of decisions are made through the tree path starting from its root and ending at a leaf node specifying the class [8].

A decision tree is chosen for its simplicity where the tree is easy to understand. Also, it works well with large datasets and does not need much data preprocessing. Furthermore, it handles continuous and categorical data types.

CART uses a recursive technique that is applied on each child node to build the tree. It divides the tree into smaller and smaller subsets by adding new test conditions until each leaf node's members belong to the same class. The set of features chosen as test conditions provides information about their importance [11]. CART's test conditions are binary which means that it creates a binary tree. One of the advantages of CART is that it provides a post-pruning strategy to prevent overfitting after the complete tree is built.

2) Training

The R package **rpart** [12] was used to build the model. Table 7 represents a summary of the features the model identified as the most important ones. It is interesting that only a small set of features have high importance. PER (persistence) was chosen to be the most important feature. Most of the other important features were also related to the change in intensity.

TABLE 7
THE MOST IMPORTANT FEATURES CHOSEN BY CART AND THEIR IMPORTANCE

Feature	Importance	Feature	Importance	Feature	Importance
PER	22	PER18	15	VPER	14
VPER18	11	PER6	10	VPER6	9
PER24	3	SHDC	2	VPER24	2
LSHDC	2	VSHDC	1	LOC	1
LAT	1	YDAYS	1	SPDX	1
ADAY	1	POT	1	Z850	1

The produced complete tree is very large with a high testing error. To reduce the tree’s size we pruned the tree using the complexity parameter. The complexity parameter that minimizes the error after cross validation was chosen.

The 11 features that were used in splits in the final tree are ADAY, PER18, LOC, LSHDC, PER, POT, PSLV, SHDC, VPER6, YDAYS, and Z850. Figure 5 shows one of the final CART models (there is one tree built per cross validation run).

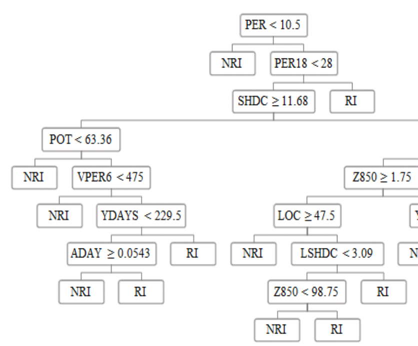

Fig. 5 A representation of a CART decision tree.

3) Testing Result

Figure 6 is a representation of the results of POD and FAR using 10-fold cross validation.

D. Support Vector Machine (SVM)

1) Description

SVM is a very important and successful classification and regression technique. Support vectors are points that are selected from the training data and are used to specify the best decision boundary given as a hyperplane that produces the maximum margin. The margin is the distance between two hyperplanes used to separate the classes. The idea is to maximize the space in order to have the largest possible separation between classes. A new data point will be assigned a class depending on which side of the separating hyperplane it falls [8].

If a hyperplane that separates the two classes in the dataset perfectly can be found, then the dataset is linearly separable and the linear discriminant function shown in (10) can be used to separate the classes [13].

$$f(x) = w^T x + b \tag{10}$$

where *w* and *b* are parameters of the hyperplane and $w^T x$ is the dot product.

SVM also works with data where classes cannot be separated linearly. To deal with such a problem, the training data can be transformed into a high dimensional space in which the data is linearly separable. This non-linear discriminant function is shown in (11) [13].

$$f(x) = w^T \Phi(x) + b \tag{11}$$

where $\mathbf{\Phi}(x)$ maps *x* into the transformed space.

When a non-linear transformation is performed, the dimensionality of the feature space can be very large and the transformation can be very expensive. However, this can be avoided by using the kernel trick which transforms the dataset into an inner product space [17]. This step computes the dot product between the pair of points to measure the similarity between the objects in the transformed space using the original space using (12) [8].

$$K(x, y) = \Phi(x)^T \Phi(y) \tag{12}$$

Polynomial, Gaussian, radial basis, and sigmoid are some kernel examples used to transform the data.

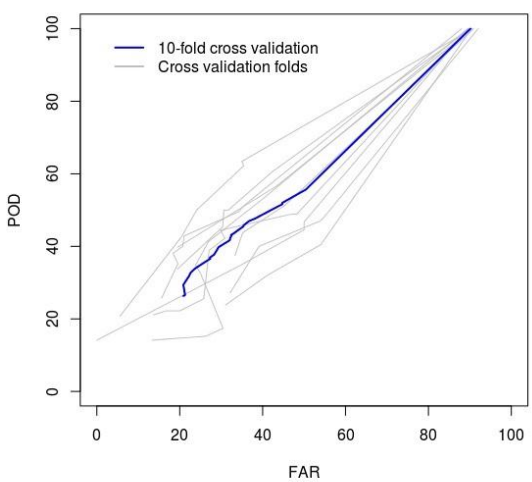

Fig. 6 Results of CART model POD and FAR using 10-fold cross validation.

2) Training

The **e1071** R package includes a SVM classifier that we used to build our model. The package provides the following kernel functions: linear, polynomial, radial basis, and sigmoid. We built our model using the above kernels and chose the one with the least testing errors.

***3)** Testing Result*

The results of POD and FAR of each kernel are shown in Table 8.

TABLE 8
POD AND FAR RESULTS USING DIFFERENT KERNELS

Kernel	POD	FAR
Radial	39.24	6.061
Linear	41.46	15
Sigmoid	39.02	50.77
Polynomial	24.39	9.09

As we can see from Table 8, the radial kernel produced the least testing errors while POD is only slightly below the linear kernel. Therefore, we build the model using the radial kernel. The results of POD and FAR using 10-fold cross validation are shown in Figure 7.

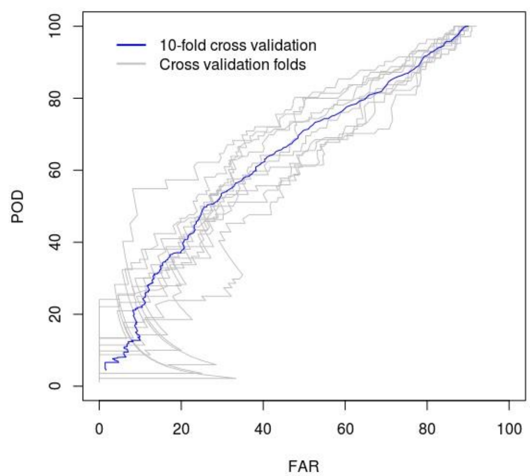

Fig. 7 Results of SVM model POD and FAR using 10-fold cross validation using a radial kernel.

II. Comparison

By looking at Figure 8, we can see the results of the different models using 10-fold cross validation. It shows that the support vector machine (SVM) model using a radial kernel outperforms the other models. The performance of logistic regression (LR) is very close to SVM where naïve-Bayes and classification and regression tree (CART) do not perform as well.

The results of the experiments show an improvement in the prediction of rapid intensification which can potentially improve the intensity forecast of tropical cyclones (TC).

III. Conclusion

Predicting RI is very important since it can improve the intensity forecast which can prevent loss of human life and reduce damages that are caused during landfall. Predicting RI events is a very hard problem since it is a rare event and the causes are related to climatological properties that are still not very clear. By comparing different classification techniques, we found that predicting RI events can be enhanced using popular data mining classification techniques.

The results can be improved in the future by addressing the class imbalance problem. Feature selection algorithms can also be applied in order to add weights so that important predictors will have more influence on the classifiers.

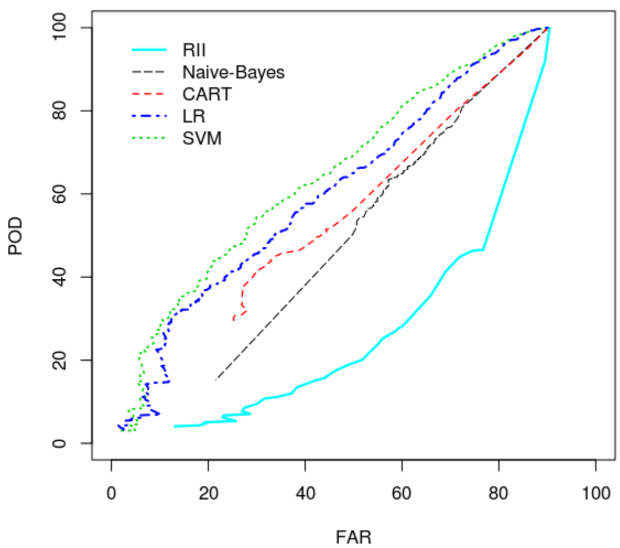

Fig. 8 A comparison of the models' POD and FAR using 10-fold cross validation.

References

[1] Wikipedia, "Hurricane Katrina," Wikipedia, The Free Encyclopedia, 2013. [Online]. Available: http://en.wikipedia.org/w/index.php?title=Hurricane_Katrina&oldid=573086191. [Accessed 17 9 2013].

[2] Wikipedia, "Saffir-Simpson hurricane wind scale," Wikipedia, The Free Encyclopedia, 2013. [Online]. Available: http://en.wikipedia.org/w/index.php?title=Saffir%E2%80%93Simpson_hurricane_wind_scale&oldid=571645050. [Accessed 17 9 2013].

[3] Wikipedia, "Tropical cyclone," Wikipedia, The Free Encyclopedia, 2013. [Online]. Available: http://en.wikipedia.org/w/index.php?title=Tropical_cyclone&oldid=572968333. [Accessed 17 9 2013].

[4] M. H. Dunham, Data Mining: Introductory and Advanced Topics, India: Pearson Education , 2006.

[5] M. DeMaria, M. Mainelli, L. Shay, J. A. Knaff and J. Kaplan, "Further improvements to the statistical hurricane intensity prediction scheme (SHIPS)," *Weather and Forecasting*, vol. 20, no. 4, pp. 531-543, 2005.

[6] J. Kaplan, M. DeMaria and J. A. Knaff, "A revised tropical cyclone rapid intensification index for the Atlantic and eastern north Pacific basins," *Weather and Forecasting*, vol. 25, no. 1, pp. 220-241, 2010.

[7] RAMMB, "Statistical tropical cyclone intensity forecast technique development," 2013. [Online]. Available: http://rammb.cira.colostate.edu/research/tropical_cyclones/ships/developmental_data.asp. [Accessed 30 9 2013].

[8] P.-N. Tan, M. Steinbach and V. Kumar, Introduction to Data Mining, Pearson Education, 2006.

[9] G. Shmueli, N. R. Patel and P. C. Bruce, Data Mining for Business Intelligence: Concepts, Techniques, and Applications in Microsoft Office Excel With XLMiner, Wiley, 2008.

[10] D. Meyer, E. Dimitriadou, K. Hornik, A. Weingessel and F. Leisch, "e1071: misc functions of the department of statistics (e1071), Tu Wein," 12 9 2012. [Online]. Available: http://CRAN.R-project.org/package=e1071. [Accessed 8 9 2013].

[11] J. Fridley, "Tree models in R," 2010. [Online]. Available: http://plantecology.syr.edu/fridley/bio793/cart.html. [Accessed 1 10 2013].

[12] T. Therneau, B. Atkinson and B. Ripley, "rpart: recursive partitioning," 2012. [Online]. Available: http://CRAN.R-project.org/package=rpart. [Accessed 5 10 2013].

[13] A. Ben-Hur and J. Weston, "A user's guide to support vector machines," in *Data Mining Techniques for the Life Sciences*, Springer, 2010, pp. 223-239.

[14] T. Fawcett , "An introduction to ROC analysis," Pattern recognition letters, 27, 861-874, 2006.

[15] R Core. Team, "R: a language and environment for statistical computing," R Foundation fo Statistical Computing, 2011. [Online]. Available: http://www.R-project.org/. [Accessed 17 11 2013].

[16] T. Hastie, R. Tibshirani, and J. Friedman. "The elements of statistical learning," Springer, New York, NY, 2001.

[17] Wikipedia, "Kernel trick," Wikipedia, The Free Encyclopedia, 2013. [Online]. Available: http://en.wikipedia.org/w/index.php?title=Kernel_trick&oldid=570535370. [Accessed 13 10 1013].