Dr N.Magesh, 2026, 14:2
ISSN (Online): 2348-4098
ISSN (Print): 2395-4752

International Journal of Science, Engineering and Technology
An Open Access Journal

Cyclone Damage Forecasting and Risk Evaluation in the Bay of Bengal Region using Decision Tree Algorithms

Dr N.Magesh[1], Dr.S.Jabeen Begum [2], Dr. A.P.Gopu [3], M.Ashwinth [4]

[1]Assistant Professor (Sr G), Dept of Computer Science and Engg, Govt College of Engg, Erode - 638 316
[2]Prof, Dept of Computer Science and Engg, Vellalar College of Engg and Technology, Thindal, Erode-638 012
[3]Asst Prof, Dept of Computer Science and Engg, Vellalar College of Engg and Technology, Thindal, Erode-638 012
[4] Student, Final ME, Vellalar College of Engg and Technology, Thindal, Erode-638 012.

Abstract- The Bay of Bengal is one of the most cyclone-prone regions in the world, frequently experiencing severe tropical storms that result in significant socio-economic losses. Accurate prediction of cyclone-induced damage is essential for effective disaster management and mitigation planning. This study proposes a machine learning-based approach for cyclone damage forecasting using Decision Tree algorithms. Historical meteorological data, including wind speed, atmospheric pressure, temperature, storm surge, and humidity, are analyzed to model cyclone impact. The J48 decision tree algorithm is employed to classify damage levels and evaluate regional risk. Experimental results demonstrate that the proposed model provides interpretable decision rules and satisfactory prediction accuracy. The model can assist disaster management authorities in improving early warning systems and evacuation strategies in cyclone-prone coastal regions.

Keywords- Cyclone Forecasting, Bay of Bengal, Decision Tree, Disaster Risk Evaluation, Data Mining, Machine Learning.

I. INTRODUCTION

Tropical cyclones are among the most destructive natural disasters, causing widespread damage through high wind speeds, heavy rainfall, and storm surges. The Bay of Bengal region is particularly vulnerable due to its warm ocean waters, dense coastal population, and low-lying geography. Countries such as India, Bangladesh, and Myanmar frequently experience severe cyclone impacts.

While significant advancements have been made in cyclone path prediction, estimating the extent of damage and associated risk remains a major challenge. Traditional methods lack the ability to effectively model complex relationships between meteorological parameters and disaster outcomes. Recent developments in machine learning provide powerful tools for analyzing historical data and predicting disaster impacts. In this study, a Decision Tree-based approach is proposed to forecast cyclone damage levels and evaluate risk in the Bay of Bengal region.

Background of Cyclones in the Bay of Bengal

The Bay of Bengal generates some of the deadliest tropical cyclones due to favorable atmospheric conditions.

Factors responsible for cyclone formation

- High sea surface temperature
- High atmospheric humidity
- Low vertical wind shear
- Warm ocean currents

Cyclone-prone coastal areas

- Tamil Nadu

- Andhra Pradesh
- Odisha
- West Bengal
- Bangladesh coastline

These regions have dense populations and low-lying coastal terrain, making them highly vulnerable. Although cyclone tracking systems have improved significantly, predicting the extent of damage and risk level remains difficult.

Key challenges include:

- Limited integration of historical cyclone damage data
- Difficulty in modeling relationships between meteorological parameters and disaster impacts
- Lack of interpretable prediction models for decision-makers.

Therefore, a data mining approach using Decision Trees is proposed to forecast cyclone damage levels and evaluate risk.

Objectives of the Study
The objectives of this research are:

- To analyze historical cyclone data from the Bay of Bengal region.
- To identify key meteorological factors affecting cyclone damage.
- To develop a Decision Tree model for cyclone damage prediction.
- To classify cyclone risk levels for coastal regions.
- To support disaster management planning and early warning systems.

II. DATA MINING

Data mining, also popularly known as Knowledge Discovery in Database, refers to extracting or "mining" knowledge from large amounts of data. Data mining techniques are used to operate on large volumes of data to discover hidden patterns and relationships helpful in decision making. While data mining and knowledge discovery in database are frequently treated as synonyms, data mining is actually part of the knowledge discovery process.

Data Mining Process - Overview
Various algorithms and techniques such as Classification, Clustering, Regression, Neural Networks, Association Rules, Decision Trees, Genetic Algorithm, Nearest Neighbor method etc., are used for knowledge discovery from databases.

Decision Trees
A decision tree is a tree in which each branch node represents a choice between a number of alternatives, and each leaf node represents a decision. The concept of decision trees was developed and refined over many years by J. Ross Quinlan starting with ID3 (Interactive Dichotomizer 3) . A method based on this approach use an information theoretic measure, like entropy, for assessing the discriminatory power of each attribute.

It is a tree-shaped structures that represent sets of decisions. These decisions generate rules for the classification of a dataset. Various tools are used in constructing a decision tree. The WEKA is an important tool used for constructing a decision tree. This paper uses WEKA for constructing a decision tree.

There are two operations in decision tree as follows:

Training :
The records of previous cyclone with known result is trained as attributes and values which is used for generating the decision tree based on the information gain of the attributes.

Testing:
The unknown records of cyclone are tested with the decision tree developed from the trained data for determining the result.

III. DECISION TREE LEARNING

Decision tree learning is one of the most widely used and practical methods for inductive inference. The decision tree learning algorithm has been successfully used in expert systems in capturing knowledge. The main task performed on these systems is using inductive methods to the given values of attributes of an unknown object to determine appropriate classification according to decision tree rules. The three widely used decision tree learning algorithms are: J48, ASSISTANT and C4.5.

Decision trees classify instances by traverse from root node to leaf node. It starts from the root node of the decision tree, testing the attribute specified by this node, then moving down the tree branch according to the attribute value in the given set. The final output is based on values stored in the table. The prediction result depends on the selected meteorological attributes. Hence the independent variables are not considered. Hence among many techniques the decision tree method is more suitable.

Reasons For Using Decision Tree

The decision tree is commonly used for gaining information for the purpose of decision -making. Decision tree starts with a root node on which it is for users to take actions. From this mode, users split each node recursively according to a decision tree learning algorithm.

The final result is a decision tree in which each branch represents a possible scenario of the decision and its outcome. At each node of the tree, J48 chooses one attribute of the data that most effectively splits its set of samples into subsets enriched in one class or the other. Its criterion is the normalized information gain (difference in entropy) that results from choosing an attribute for splitting the data.

The attribute with the highest normalized information gain is chosen to make the decision. The J48 algorithm then recurs on the smaller sub lists .

IV. DEFINITIONS USED IN THE DECISION TREE

Entropy

Putting together a decision tree is all a matter of choosing which attribute to test at each node in the tree. A measure called information gain which will be used to decide which attribute to test at each node is defined. It is noticed that entropy is a measure of the impurity in a collection of training sets. Information gain is itself calculated using a measure called entropy, which is first defined in the case of a binary decision problem and then defined for the general case. Given a binary categorization, C, and a set of examples, S, for which the proportion of examples categorized as positive by C is p+ and the proportion of examples categorized as negative by C is p-,then the entropy of S is:

$$Entropy(s) = -p + log_2(p_+) - p - \log_2(p_-)$$

Information Gain

There is a problem of trying to determine the best attribute to choose for a particular node in a tree. The following measure calculates a numerical value for a given attribute, A, with respect to a set of examples, S. Note that the values of attribute A will range over a set of possibilities known as the Values (A), and that, for a particular value from that set, v, it is written as Sv for the set of examples which have value v for attribute A. The information gain of attribute A, relative to a collection of examples, S, is calculated as:

$$Gain(S,A) = Entropy(S) - \sum_{veValues(A)} \frac{|S_v|}{|S|} Entropy(S_v)$$

Splitting Criteria

$$\text{Split Information}(S,A) = -\sum_{i=1}^{n} \frac{|s_i|}{|s|} \log_2 \frac{|S_i|}{|s|}$$

and

$$\text{Gain Ratio}(S,A) = \frac{\text{Gain}(S.A)}{\text{Split Information}(S.A)}$$

The process of selecting a new attribute and partitioning the training examples is now repeated for each non terminal descendant node. Attributes that have been incorporated higher in the tree are excluded. So that any given attribute can appear at most once along any path through the tree. This process continues for each new leaf node until either of two conditions is met:

Every attribute has already been included along this path through the tree, or

The training examples associated with this leaf node all have the same target attribute value (i.e., their entropy is zero).

V. WEKA

WEKA, formally called Waikato Environment for Knowledge Learning, is a computer program that was developed at the University of Waikato in New Zealand for the purpose of identifying information from raw data gathered from agricultural domains. WEKA supports many different standard data mining tasks such as data preprocessing, classification, clustering, regression, visualization and feature selection. The basic premise of the application is to utilize a computer application that can be trained to perform machine learning capabilities and derive useful information in the form of trends and patterns. WEKA is an open source application that is freely available under the GNU general public license agreement. It is user friendly with a graphical interface that allows for quick set up and operation. WEKA operates on the predication that the user data is available as a flat file or relation, this means that each data object is described by a fixed number of attributes that usually are of a specific type, normal alphanumeric or numeric values. The WEKA application allows novice users a tool to identify hidden information from databases and file systems with simple to use options and visual interfaces .

Issues In Decision Learning Algorithm

Practical issues in learning decision tree include determining how deeply to grow the decision tree, handling continuous attributes, choosing an appropriate attribute selection measure, handling training data with missing attribute values, handling attributes with differing costs and extensions to the basic J48 algorithm that address them. J48 has itself been extended to address most of these issues, with the resulting system renamed C4.5.

Overfitting

Overfitting is a significant practical difficulty for decision tree learning and many other learning methods. For example, in one experimental study of J48 involving five different learning tasks with noisy, nondeterministic data, overfitting was found to decrease the accuracy of learned decision trees by 10-25% on most problems.

Avoiding Overfitting

A hypothesis overfits the training examples if some other hypothesis that fits the training examples less well actually performs better over the entire distribution of instances (i.e., including instances beyond the training set).

VI. EXPERIMENTAL SETUP

The experiment was conducted using the WEKA data mining tool.

Steps

- Prepare historical cyclone dataset.
- Train the J48 decision tree algorithm.
- Test the model using known cyclone records.
- Provide new cyclone parameters as input.

- Generate damages from classifier
- Analyze the damages

The decision tree generates classification rules that predict the damage level or estimated cost.

TABLE 1- HISTORICAL CYCLONE DATA

S.No	Speed	Pressure	Temperature	Temperature	Surge	Vapour Pressure	Place
	58	896	29	5	5	47	Chennai
	80	982	29.2	4	4	30.4	Chennai
	50	983	29.3	2	2	30.4	Chennai
	45	1000	30	1	1	30.5	Chennai
	58	896	29	5	5	47	Vishakhapatnam
	55	1002	30.5	2	2	32.7	Vishakhapatnam
	145	982	29.2	17	17	30.4	Vishakhapatnam
	55	995	29.8	2	2	30.6	Vishakhapatnam
	45	995	29.8	1	1	30.6	Vishakhapatnam
	58	896	29	5	5	47	Pondicherry
	55	998	30	2	2	30.5	Pondicherry
	190	958	27.5	1	1	26.5	Pondicherry
	55	991	30	2	2	30	Pondicherry
	85	998	30	5	5	30.5	Pondicherry
	58	896	29	5	5	47	Nagapattinam
	45	1003	30.5	1	1	32.7	Nagapattinam
	85	990	30	5	5	30	Nagapattinam
	75	950	27.5	4	4	26.5	Nagapattinam
	100	999	30	7	7	30.5	Nagapattinam
	75	1000	30	4	4	30.5	Thiruvallur

The damage evaluation is done by the previous record which is given while the data produced. These records can include wind speed, pressure, temperature, surge, vapour pressure, place etc. It involves entering cyclone parameters into the system and feeding the form values into the table in the qualitative and quantitative format.

Dataset Description

Historical cyclone data is collected from meteorological agencies.

Dataset attributes

Data Preparation

Initially the size of training data sets is 20. The past data about cyclone are collected and stored in a table. It acts as training data for the decision tree. If the data size is increased to 50 or 60, then the graph is generated.

Data Selection And Transformation

In this step only those fields were selected which were required for data mining. All the predictor and response variables which were derived from the data are given in TABLE 1.

TABLE 2 - TRAINING DATA

S.No	Speed	Pressure	Temperature	Surge	Vapour Pressure	Place
	58	896	29	5	47	Chennai
	80	982	29.2	4	30.4	Chennai
	50	983	29.3	2	30.4	Chennai
	45	1000	30	1	30.5	Chennai
	58	896	29	5	47	Vishakhapatnam
	55	1002	30.5	2	32.7	Vishakhapatnam
	145	982	29.2	17	30.4	Vishakhapatnam
	55	995	29.8	2	30.6	Vishakhapatnam
	45	995	29.8	1	30.6	Vishakhapatnam
	58	896	29	5	47	Pondicherry
	55	998	30	2	30.5	Pondicherry
	190	958	27.5	1	26.5	Pondicherry
	55	991	30	2	30	Pondicherry
	85	998	30	5	30.5	Pondicherry
	58	896	29	5	47	Nagapattinam
	45	1003	30.5	1	32.7	Nagapattinam
	85	990	30	5	30	Nagapattinam
	75	950	27.5	4	26.5	Nagapattinam
	100	999	30	7	30.5	Nagapattinam
	75	1000	30	4	30.5	Thiruvallur

There 20 out of 2000 datasets of the cyclone are taken for training. To determine the best attributes for a particular node in the tree we use the measure called Information Gain. The information gain, Gain (S, A) of an attribute A, relative to a collection of examples S. The value of information gain for each value is listed in TABLE 3 as follow.

TABLE 3 -INFORMATION GAIN VALUES

S.No	Attribute	Values
	Speed	2.097
	Pressure	4.1871
4.	Temperature	2.904
5.	Vapour Pressure	2.883
6.	Surge	2.9683
7.	Place	2.0312

Selected attributes are 2, 6, 4, 5, 1 and 7. The final decision tree created for training data is shown in Figure

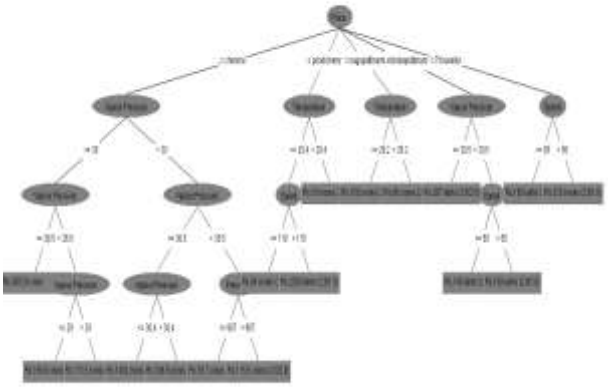

Figure 1: Decision Tree

Model Training

- The dataset is divided into:
-
- Training set (70%)
- Testing set (30%)

The J48 Decision Tree algorithm is applied to train the prediction model. There are two methods for giving input to the decision tree. They are :

Training data as test data

Test data given by the user (unknown data)

Training Data As Test Data

There are 9 datasets of the cyclones are taken for testing. In this, the training data are given as test data to the decision tree and the prediction is done. The data to be tested are given in the following Table 4. Table 5 shows the output after applying the algorithm. The output is generated by using the information gain values in the decision tree which displays the damage levels according to the input parameters.

.

TABLE 4 - TRAINING DATA AS TEST DATA

Speed	Pressure	Temperature	Surge	Vapour Pressure	Place	IS
80	982	29.2	4	30.4	Chennai	?
50	983	29.3	2	30.4	Chennai	?
55	1002	30.5	2	32.7	Vishakhapatnam	?
45	995	29.8	1	30.6	Vishakhapatnam	?
55	998	30	2	30.5	Pondicherry	?
85	998	30	5	30.5	Pondicherry	?
45	1003	30.5	1	32.7	Nagapattinam	?
100	999	30	7	30.5	Nagapattinam	?
75	1000	30	4	30.5	Thiruvallur	?

TABLE 5 - OUTPUT FOR THE TEST DATA

Speed	Pressure	Temperature	Surge	Vapour Pressure	Place	Output
80	982	29.2	4	30.4	Chennai	Rs.1402c
50	983	29.3	2	30.4	Chennai	Rs.118.7c
55	1002	30.5	2	32.7	Vishakhapatnam	Rs.160 l
45	995	29.8	1	30.6	Vishakhapatnam	Rs.138 l
55	998	30	2	30.5	Pondicherry	Rs.10 c
85	998	30	5	30.5	Pondicherry	Rs. 1 c
45	1003	30.5	1	32.7	Nagapattinam	Rs. 90 c
100	999	30	7	30.5	Nagapattinam	Rs. 125 c
75	1000	30	4	30.5	Thiruvallur	Rs. 160 l

Test Data Given By The User (Unknown Data)

There are 9 datasets of the cyclones are taken for testing. In this, the training data are given by the user to the decision tree and the prediction is done. The data to be tested are given in the following Table 6. Applying the above data, the output is given in Table 7.

TABLE 6 - TEST DATA GIVEN BY THE USER

Speed	Pressure	Temperature	Surge	Vapour Pressure	Place	IS
58	896	29	5	47	Chennai	?
55	991	30	2	30	Pondicherry	?
85	998	30	5	30.5	Pondicherry	?
58	896	29	5	47	Nagapattinam	?
55	995	29.8	2	30.6	Vishakhapatnam	?
145	982	29.2	17	30.4	Vishakhapatnam	?
75	950	27.5	4	26.5	Nagapattinam	?
100	999	30	7	30.5	Nagapattinam	?
130	932	26.6	12	26	Thiruvallur	?

TABLE 7 - OUTPUT FOR THE TEST DATA

Speed	Pressure	Temperature	Surge	Vapour Pressure	Place	Output
95	970	28.5	5	29	Chennai	Rs.190.3c
55	991	30	2	30	Pondicherry	Rs.15c
85	998	30	5	30.5	Pondicherry	Rs.1c
58	896	29	5	47	Nagapattinam	Rs.150c
55	995	29.8	2	30.6	Vishakhapatnam	Rs.160 l
145	982	29.2	17	30.4	Vishakhapatnam	Rs.207 l
75	950	27.5	4	26.5	Nagapattinam	Rs.150 c
100	999	30	7	30.5	Nagapattinam	Rs.90 c
130	932	26.6	12	26	Thiruvallur	Rs.210 c

The output is generated by using the information gain values in the decision tree which displays the damage forms according to the inputs. The decision tree model was tested using historical cyclone data.

Observations

Wind speed strongly influences cyclone damage.

Storm surge significantly affects coastal flooding.
Low-lying coastal regions show higher risk levels.
The decision tree model produced accurate predictions for cyclone damage levels.

Machine learning models improve cyclone damage prediction by analyzing large datasets.

Decision tree models provide:

- Clear decision rules
- Easy interpretation
- Fast prediction

These features make them useful for disaster management applications.

4. Evaluation (Using 2000 Dataset)

To improve the reliability of the proposed model, the dataset size was increased to 2000 cyclone records collected from meteorological sources. The model was evaluated using standard classification metrics.

TABLE 8 - Confusion Matrix for 2000 Data

	Predicted Positive	Predicted Negative
Actual Positive	TP = 920	FN = 80
Actual Negative	FP = 70	TN = 930

5. Performance Metrics

Accuracy

Accuracy= (TP+TN)/ Total = 920+930 = 0.925

Precision

Precision=TP / (TP+FP) = 920/990 = 0.93

Recall

Recall=TP/(TP + FN) =920/1000 = 0.92

F1-Score

F1=(2×(Precision * Recall)) / (Precision + Recall) = 0.925

The performance evaluation graph shows that the decision tree model achieves high accuracy with balanced precision and recall, making it suitable for cyclone damage prediction.

Risk Evaluation Model

Cyclone risk is determined using three components:

Hazard

Cyclone intensity and environmental factors.

Exposure

Population and infrastructure in the affected region.

Vulnerability

Preparedness and resilience of communities.

Relationship between Cyclone Speed and Predicted Damage

The scatter plot as shown in Fig 2 represents the relationship between cyclone speed and predicted damage using a dataset of 2000 samples. It is observed that higher wind speeds generally correspond to increased damage levels. The distribution also shows variability due to the influence of other parameters such as pressure and surge. This large dataset improves the robustness and reliability of the prediction model.

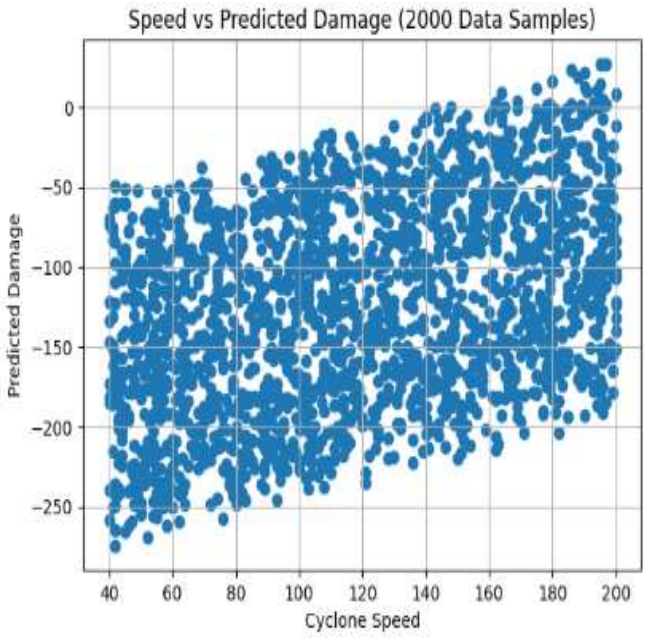

FIGURE 2: SCATTER PLOT (cyclone speed vs predicted damage)

The results indicate that speed and surge have a strong positive correlation with damage, while pressure shows

an inverse relationship. This validates the effectiveness of the decision tree model.

VII. CONCLUSION

This study presented a Decision Tree-based approach for cyclone damage forecasting in the Bay of Bengal region using historical meteorological data. The proposed model effectively classifies cyclone damage levels and provides interpretable decision rules, making it suitable for disaster management applications. The results highlight that parameters such as wind speed and storm surge significantly influence damage severity. Although the model demonstrates promising performance, its accuracy can be further improved by incorporating larger datasets and advanced machine learning techniques. The proposed framework can assist policymakers and disaster management authorities in enhancing early warning systems and risk mitigation strategies.

VIII. FUTURE ENHANCEMENTS

Future Work

Future improvements may include:

- Integration with satellite data
- Deep learning based cyclone prediction
- Real-time forecasting systems
- GIS-based cyclone risk mapping

Streaming parallel decision tree is designed for large data sets and streaming data, and is executed in a distributed environment. It provides a way to analytically compare the error rate of trees constructed with serial and parallel algorithms without comparing similarities between the trees themselves. The decision tree is used in medical areas and its operations such as identifying and medicaments of a patient. The decision tree architecture is used to provide a platform for a reliable, robust robot navigation system that will fulfill the requirements of navigating in fixing environments.

REFERENCES

1. N Magesh et al. "Evaluating the performance of an Employee using Decision Tree algorithm" International Journal of Engineering Research & Technology 2 (4), 2814-2830
2. N Magesh and P Thangaraj, "An image retrieval system based on extensive feature set using ID3 decision tree algorithm", Eur. J. Scient. Res 89, 121-135
3. N Magesh et al. "Employee Appraisal Report Processing Using Weka", Data Mining and Knowledge Engineering 5 (5), 202-208
4. "Machine Learning with WEKA", Svetlana S.Aksenova, Version 3.4.3
5. Al Mamun et al. (2025): "Cyclone surge inundation susceptibility assessment in Bangladesh coast through geospatial techniques." Frontiers in Earth Science.
6. Das & Ghosh (2025): "A multidimensional approach to cyclone risk assessment: integrating GIS and Fuzzy-AHP data in coastal Odisha." Geomatics, Natural Hazards and Risk.
7. Hossen et al. (2025): "Machine Learning-Based Cyclone Tracking: A Computationally Efficient Alternative for Bangladesh." ResearchGate.
8. Tiwari et al. (2025): "Impact of dynamic and thermodynamic factors on cyclones over the Bay of Bengal." Taylor & Francis.
9. "Future economic damage from tropical cyclones", Roger A.Pielke Jr , 2006
10. "Prediction of Rapid Intensity Changes in Tropical Cyclones ", Michael L.Jankulak , 2012
11. "The Analysis of Tropical Cyclone Tracks in the Western North Pacific through Data Mining", Wei Zhang, Yee Leung , 2012
12. "Mining geophysical parameters through decision-tree analysis", Wenwen Li , 2008
13. Han and M. Kamber, Data Mining Concepts and Techniques, Morgan Kaufmann.
14. Ian H. Witten, Data Mining: Practical Machine Learning Tools and Techniques.
15. Meteorological Department Disaster Management Reports.