International Journal of Engineering Trends and Technology
ISSN: 2231–5381 / https://doi.org/10.14445/22315381/IJETT-V73I7P138

Volume 73 Issue 7, 493-504, July 2025

Original Article

Discriminant Granger and Camargo Index Jensen Shannon Boosting Classifier for Enhanced Marine Weather Forecasting

Soumya Unnikrishnan[1], Smitha Vinod[2]

[1,2]*Christ (Deemed to be University), Bengaluru, Karnataka, India.*

[1]*Corresponding Author : soumyaunnikrishnan@res.christuniversity.in*

Received: 05 March 2025 Revised: 24 June 2025 Accepted: 16 July 2025 Published: 30 July 2025

Abstract - *Predicting the weather is essential for people's everyday demands and tasks. Furthermore, a number of industries, like agriculture, irrigation, etc., depend on precise weather forecasting. Many people may experience issues as a result of weather forecast deviations. Therefore, making accurate weather predictions is a crucial issue that requires attention. Enormous data assessment is a process for looking into enormous data in order to find hidden patterns that may lead to improved results. Big data has gained attention in a number of societal sectors in recent years. The evaluation of big data helps forecasters make more accurate weather predictions and produces better results while doing so. Therefore, reliable weather prediction is beyond the capabilities of standard computer intelligence models. The ceaseless evolution of big data technology necessitates preferable learning methods to discover the data value. However, the prevailing divergence and redundancy in the data acquired from a series of buoys make it both laborious and cumbersome to accurately predict future information. Motivated by these challenges, an ensemble paradigm called Discriminant Granger Causality and Camargo-Index Jensen Shannon Boosting Classifier (DGC-CJSBC) is proposed for enhanced marine weather forecasting. Here, enhanced refers to the marine weather forecasting employing big marine data using a boosting classifier model. The DGC-CJSBC method for enhanced marine weather forecasting is split into two sections. First, a linear combination of features characterizing two classes is modeled. DGC-CJSBC method deploys a novel change detection (i.e., changes observed between previous day oceanographic and surface meteorological readings) based on Jensen–Shannon divergence to record changes throughout the equatorial Pacific. Moreover, the DGC-CJSBC method considers the Camargo Evenness Index Quadratic function. In this way, classification performance improved. The El Nino Big dataset was applied to train the proposed model. Contrary to conventional algorithms, the DGC-CJSBC method outcome offers better accuracy, time, error and space complexity.*

Keywords - *Big marine data, Discriminant, Granger causality, Camargo-index, Jensen shannon, Boosting classifier.*

1. Introduction

In an atmospheric situation, weather forecasting is specifically utilized to calculate arbitrary alterations. Improving the utilization of electronic devices has led to an enormous production of high-volume data. This information is being transmitted to meteorological centers for predicting future conditions employing clustering and classification techniques from large databases. With the objective of acquiring both spatial and temporal associations of numerous meteorological characteristics, a novel Spatial Feature Attention-LSTM (SFA-LSTM) method was proposed.

With the employment of an attention mechanism, spatial features were captured. Via LSTM, temporal independencies were derived utilizing an encoder-decoder, therefore ensuring prediction accuracy with minimum error. Several mechanisms are employed to carry out weather forecasting. Nevertheless, it failed to deal with huge amounts of information for several features. A method called Tanimoto Correlation-based Combinatorial MAP Expected Clustering and Linear Program Boosting Classification (TCCMECLPBC) was designed in [2] to improve the prediction accuracy with lesser time consumption. In the dataset, data and features were collected. Significant features identified by Tanimoto Correlation. Weather information grouped with MAP expected clustering. But it failed to decrease the time.

A data distribution model was introduced in [3] with cloud computing and volunteer computing environment in a hybrid manner for Big Data analytics. But it considers a higher error rate. The Numerical Weather Prediction method was introduced in [4], which depends on artificial intelligence, and takes less time. However, it failed to minimize space complexity.

1.1. Problem Definition

Recent developments in marine weather forecasting have led to the growth of various new artificial intelligence, boosting and ensemble-based techniques. In the boosting-based techniques, by identifying similarity between spatial and temporal relations, which can differentiate between spatial and temporal interpretations [1], short-term weather forecasting was made. Also, by employing ensemble classification [2], relevant features were selected, and via expectation maximization, a boosting classifier, the prediction performance was improved. However, with the increase in the size and nature of data found to be constrained in nature, there was a negative influence between previously established oceanographic and surface meteorological readings. Though the existing boosting and ensemble-based techniques performed prediction, due to the high volume and velocity, timely detection was not ensured. Therefore, they are complex and slow in nature. Therefore, there is a requirement for enhanced marine weather forecasting involving big data.

1.2. Proposed Solution

To solve the above-mentioned problem, this paper presents a boosting-based ensemble classification method that first identifies a linear combination of features characterizing two classes, which in turn ensures dimensionality reduction even in the case of big marine data. This paper presents an ensemble classifier that joins the results of the weak learners to find accurate weather forecasting results with minimal error.

1.3. Objective

The study aims to provide a boosting classification-based ensemble method for addressing conventional marine weather forecasting approaches with enhanced precision in big data.

An enhanced marine weather forecasting method has been developed for handling state-of-the-art ensemble boosting models.

- To achieve accurate classification, regression-based weather forecasting feature selection was introduced for correlated time instances of univariate as well as bivariate analysis, compared to conventional works.
- The Camargo Index Jensen-Shannon Boosting Classifier is for enhanced marine weather forecasting by ensemble weak learners compared to existing boosting classifiers.
- A fine-tune method based on both previous and current meteorological readings is used to predict marine weather.
- The performance of previous and current meteorological readings is estimated with the existing boosting classifiers.

1.4. Contributions

The following are the major contributions of our paper:

- A multivariate feature selection model is introduced to choose meteorological features as well as time-varying oceanographic and surface meteorological variables. In this paper, it is crucial to measure the influences of oceanographic and surface meteorological variables on marine weather forecasting.
- Jensen–Shannon divergence evaluates the allocation between the reference and current window for the change detection issue.
- Latitude and longitude from the approximate location are employed by ensemble classification to preserve classifiers.
- The E1-nino database was used to validate the proposed method. The outcome of DGC-CJSBC accomplishes precise prediction along with the minimization of prediction error compared to existing conventional marine weather forecasting methods.

1.5. Organization Paper

The article is arranged as follows: marine weather forecasting methods using optimization and learning techniques are described in Section 2. Section 3 proposes our Discriminant Granger Causality and Camargo-Index Jensen Shannon Boosting Classifier (DGC-CJSBC) for enhanced marine weather forecasting. Section 4 presents the experimental results of E1 Nino Big datasets and compares with existing methods. Conclusion presented in Section 5.

2. Literature Survey

Weather forecasting is referred to as a process that comprises temperature, humidity, wind speed, and direction. Huge volumes of data are required for forecasting the weather. Moreover, the data are found to be disorganized. Hence, weather prediction is said to be a complicated task owing to a large number of changeable factors. These factors differ depending on the weather conditions, which are said to be very swift. A framework featuring long short-term memory deep learning for spatiotemporal extreme forecasting in the SouthPacific region was presented in [5]. Here, an indicator called the Effective Drought Index (EDI) was utilized for understanding the dominant feature, therefore ensuring model accuracy. Though accuracy was ensured, the reliability factor involved in predicting Sea Surface Temperature (SST) was not focused on. To address this aspect, several deep learning models were trained in [6], capturing the major meteorological and oceanic features governing SST variability. With this, a reliable SST prediction was said to be ensured. However, the time factor was not analyzed.

A platform risk assessment mechanism by utilizing dual machine learning techniques with the objective of forecasting in the U.S. federal waters of the Gulf of Mexico (GoM) employing Gradient Boosted Regression Tree (GBRT) and an Artificial Neural Network (ANN) was presented in [7]. With this type of boosting mechanism, the environmental risk involved in predicting offshore platforms was said to be reduced extensively. A systematic review on big data analytics for weather forecasting was conducted in [8]. Advanced machine learning techniques like Seasonal Auto-regressive Integrated Moving Average and Long Short Term Memory were analyzed in depth [9], therefore ensuring accuracy. But it failed to focus on the time factor. As far as differences in climatic conditions are concerned, ocean parameters are of rising interest in ocean-related fields. Most of the prevailing methods only insisted on the utilization of a single parameter, such as Sea Surface Temperature (SST).

In [10], a deep learning technique like Multi-Variant Convolutional (MVC) High Speed (HS) Long and Short-Term Memory(HM-LSTM) method was presented with the objective of predicting four distinct metrics, like temperature, pressure, salinity and density. By means of utilizing these four metrics for analysis, the error involved in prediction was said to be reduced significantly. A critical review of wind and solar power forecasting employing accurate short-term predictive models was presented in [11].

However, the unpredictability and arbitrary facets of wind power had an adverse influence on grid planning and operation. A significant wind power forecasting method was presented in [12] to address these concerns. With the utilization of Artificial Intelligence (AI) techniques, high precision was said to be achieved. AI-based hybrid methods were proposed here to solve error factors. Though precision and accuracy factors were analyzed above, the error rate was not focused on.

One of the significant elements that highly influences agricultural production is weather. Over the past few years, weather forecasting has entered the Big Data era. Therefore, conventional computational intelligence techniques are not adequate for accurately predicting the weather. A survey of weather forecasting models based on deterministic factors that can learn and make predictions in a more efficient manner on the basis of past data was investigated in [13]. Two types of machine learning algorithms, multiple regression and multilayer perceptron, were applied in [14] for determining sensitivity analysis. In weather forecasting, false alarms play a vital role. To address this aspect, a convolutional neural network and long short-term memory were presented in [15] that, in turn, not only reduced the false alarm but also improved the stability to a greater extent.

In [16], a review of materials and methods for assessing water quality contributing to sustainable management of marine environments was designed. To be more specific, the Deep Learning (DL) technique was employed for estimating and forecasting water quality. Several techniques have been proposed with the objective of enhancing the forecasting reliability level. Amongst them, Long Short-Term Memory (LSTM) is a frequent method for making predictions on the basis of timeseries data. A machine learning named Adaptive Dynamic Particle Swarm Algorithm (AD-PSO) integrated with Guided Whale Optimization Algorithm (Guided WOA) for forecasting wind speed ensemble was proposed in [17].

It is well-known that the prediction of weather using numerical techniques necessitates substantial computing power to answer complicated mathematical equations for forecasting on the basis of the prevailing weather conditions. In [18], a novel lightweight data-driven weather forecasting method by navigating temporal approaches of Long Short-Term Memory(LSTM) and Temporal Convolutional Networks (TCN) were presented. The proposed LSTM and TCN layers ensure accurate fine-grained weather forecasting. In [19], location-specific Sea Surface Temperature forecasts were produced by integrating deep learning with numerical valuators at five distinct locations for three numerical time horizons. An improved decision tree method was proposed in [8] to enhance the time and accuracy of prediction with big data. In this article, a Discriminant Granger Causality and Camargo-Index Jensen Shannon Boosting Classifier (DGC-CJSBC) is proposed for enhanced marine weather forecasting. DGC-CJSBC explained the sections below.

3. DGC-CJSBC for Enhanced Marine Weather Forecasting

Weather forecasting, as a significant and essential plan of action in people's everyday lives, measures variations experienced in the prevailing atmospheric situation. On the other hand, big data refers to examining huge amounts of information to extract hidden patterns that can produce better results. Over the past few years, various elements of society have been concerned, and the meteorological agency is no exception. For precise weather prediction, big data produces superior outcomes. We plan to develop Discriminant Granger Causality and Camargo-Index Jensen Shannon Boosting Classifier (DGC-CJSBC) for enhanced marine weather forecasting of marine big data with higher accuracy and lesser time consumption. DGC-CJSBC method of weather forecasting includes feature selection using Discriminant Granger Causality Regression function and classification employing Camargo Index Jensen-Shannon Boosting Classifier. The DGC-CJSBC method of architecture is illustrated in Figure 1.

As shown in the above Figure, first, with the purpose of reducing the dimensionality involved, Discriminant Granger Causality Regression-based Feature Selection is performed, which in turn obtains a linear combination of features, therefore characterizing them into two classes. Second, the Camargo Index Jensen-Shannon Boosting Classifier is utilized for weather forecasting. Jensen-Shannon Boosting Classifier is an ensemble technique. Testing as well as training data were examined by the Camargo Index Quadratic Classifier. Finally, the results of weak learners are joined by an ensemble classifier to identify accurate weather forecasting results with minimal error.

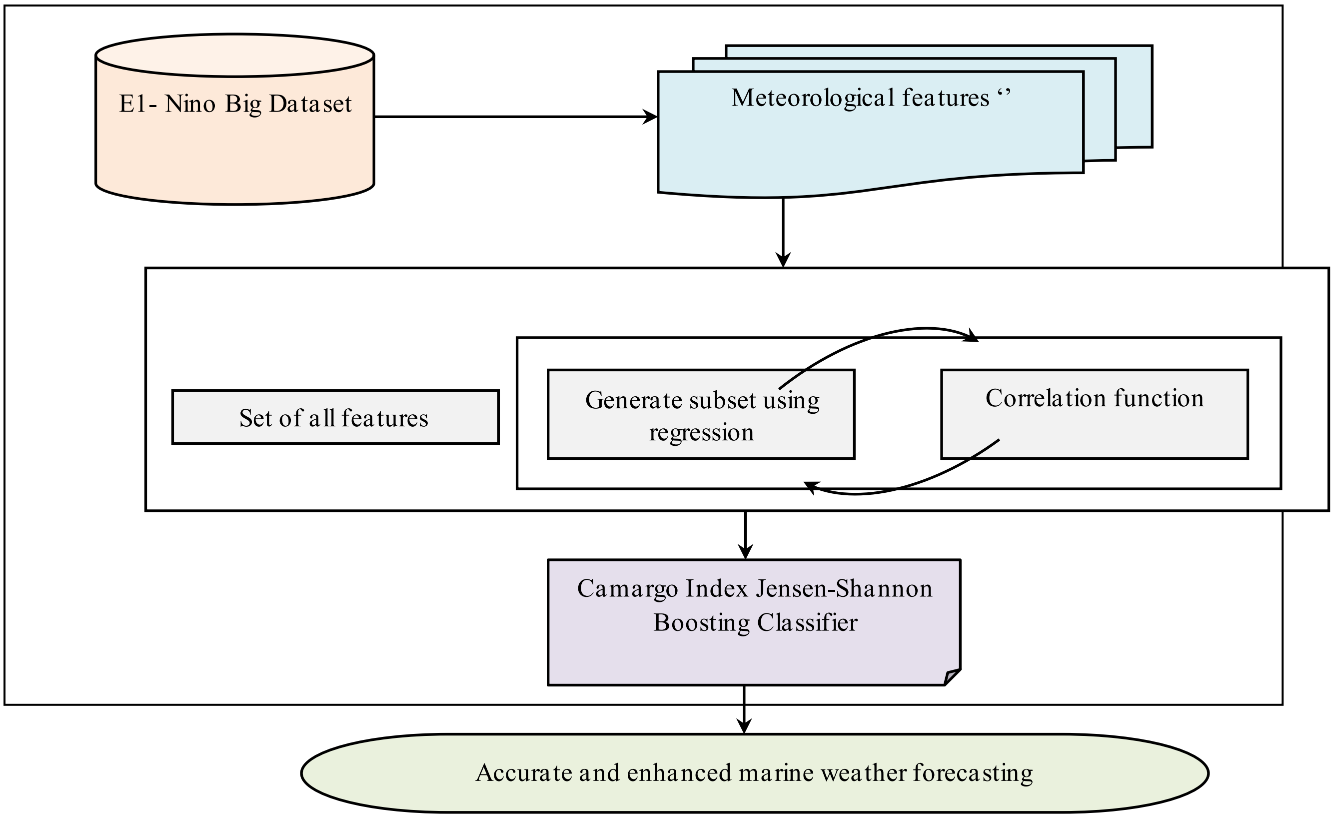

Fig. 1 Block diagram of DGC-CJSBC method

3.1. Dataset Details

E1 Nino big data dataset includes both oceanographic and surface meteorological readings obtained in the equatorial Pacific. The Tropical Atmosphere Ocean array employed consisted of 70 moored buoys spanning the equatorial Pacific.

These recordings measured the oceanographic and surface meteorological variables that were the most critical for detections. Features mentioned in Table 1.

Table 1. E1 nino big data set

Sl. No	Features
1	Year
2	Month
3	Date
4	Latitude
5	Longitude
6	Zonal Winds
7	Humidty
8	Air Temperature
9	Sea Surface

In a certain position, all the above data were considered as of buoys as of 1980. Other information was also acquired from several locations. With latitude and longitude information, the buoy shifted among distinct positions. Moreover, the value of the latitude was found to be positioned, and the longitude value was even recorded. Finally, each reading was obtained at a similar time of day.

3.2. Discriminant Granger Causality Regression-based Feature Selection

Due to certain convoluted atmospheric surroundings, oceanographic and surface meteorological readings needed in the sequence of buoys are affected. Therefore, most correlated features were selected by multifaceted feature selection. As far as the temporal dimension is concerned, the majority of correlated time instances toward target time instances can be analyzed, therefore minimizing the input variable dimensionality by eliminating irrelevant noisy information. The traditional spatial correlation analysis methods, such as Mutual Correlation [1], can only analyze either linear or non-linear correlation between features and then eliminate the features that are found to be irrelevant for further processing. Nevertheless, it failed to examine relevant features with Mutual Correlation in marine weather prediction. Correlations amongst oceanographic and surface meteorological readings are investigated by Discriminant Granger Causality Regression. This regression among features ‘P’and ‘Q’ is referred as when marine weather forecasting result via joint past data of features ‘P’and ‘Q’ better to employ feature ‘Q’ alone, then feature variable ‘P’aids to describe the transpose at future, ‘P’ said to be granger cause of ‘Q’ and vice versa. Figure 2 shows the structure of the Discriminant Granger Causality Regression-based Feature Selection model.

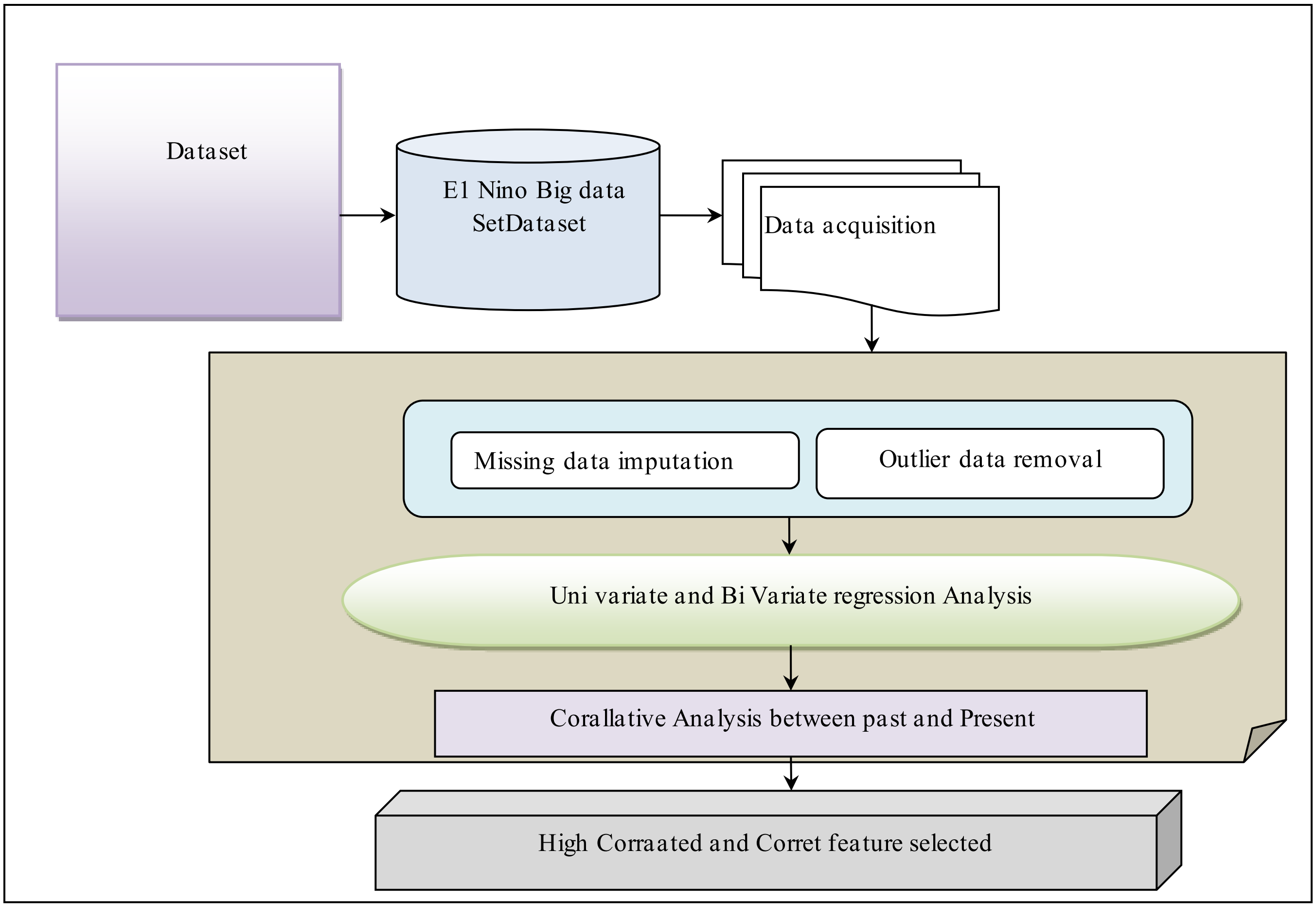

Fig. 2 Structure of discriminate granger causality regression-based feature selection model

From table 1 and Figure 2, with 12 distinct features or weather parameters present in the E1 Nino big data dataset at a given time instance, 't', '$F_t = F_1, F_2, \ldots, F_{12}$', with the big data dataset consideration, the objective remains in predicting the resultant value, 'Y_t' at time 't'. With this assumption, the boosting modeling network is defined as a function $Fun: F_{t+1} \rightarrow Y_{t+1}$ ', generating a mapping function, '$Y_1, Y_2, \ldots, Y_t = Fun\ (F_1, F_2, \ldots, F_t)$', due to the big size of data. The objective of the marine weather forecast in the proposed work remains in identifying a function 'Fun' that reduces the error 'Err' between the actual marine prediction outputs and the estimated marine predictions. Then, the multivariate time series input vector matrix for a large amount of data employing the E1 Nino big data dataset is mathematically formulated as given below.

$$IVM = \begin{bmatrix} S_1F_1 & S_1F_2 & \ldots & S_1F_n \\ S_2F_1 & S_2F_2 & \ldots & S_2F_n \\ \ldots & \ldots & \ldots & \ldots \\ S_mF_1 & S_mF_2 & \ldots & S_mF_n \end{bmatrix} \quad (1)$$

With the above input vector matrix 'IVM' in (1), the univariate and bivariate regression using Granger Causality functions are formulated as given below.

$$Q_t^U = IVM\ \left[\alpha_0 + \sum_{i=1}^{a} \alpha_i Q_{t-1} + \sum_{i=1}^{b} \beta_i Q_{t-1}\right] \quad (2)$$

$$Q_t^B = IVM\ \left[\alpha_0 + \sum_{i=1}^{a} \alpha_i Q_{t-1}\right] \quad (3)$$

From the above Equations (2) and (3), multivariate analysis (i.e., univariate and bivariate) is evaluated separately with respect to the input vector matrix 'IVM' discriminately. Finally, with the multivariate time series regression results, the association among features present value as well as corresponding features past value is evaluated by Serial Auto Correlation.

$$RF = SAC = \sum_{i=1}^{m} \frac{\left[s_i - (Q_t^U)'\right]\left[s_i - (Q_t^B)'\right]}{\sqrt{\left[s_i - (Q_t^U)'\right]^2 \left[s_i - (Q_t^B)'\right]^2}} \quad (4)$$

$$IRF = 1 - \{SAV\} \quad (5)$$

From the above Equations (4) and (5), based on the relationship between features via the Serial Auto Correlation function, highly correlated features are selected using the samples. S_i' involved in simulation, mean value of univariate regression '$(Q_t^U)'$' and mean value of bivariate regression '$(Q_t^B)'$' respectively.

//Algorithm 1:
Input: Datasets 'Ds', tweet samples $TS = \{TS_1, TS_2, TS_3, \dots TS_n\}$, large lexical database 'LLD'
Output: Improve depression detection Accuracy

Begin
1: Collect the number of tweets $TS = \{TS_1, TS_2, TS_3, \dots TS_n\}$ from dataset
2: Input the tweets 'TS' to the input layer
3: For each tweet 'TS'
4: Compute the neuron probability using (1)
5: Perform word tokenization using (2)
6: Apply the Laplace kernel to find the stop words using (3)
7: Apply the Lovins Stemmer for word stemming
8: End For
9: For each pre-processed tweet
10: Initialize word populations $W_j = W_1, W_2, \dots W_b$
11: for each word
12: Measure the word frequency using (6)
13: Measure fitness using (7) based on censored regression
14: End for
15: End for
16: Select the current best word
17. While (t < Max_iter) do
18. If $\left(F\ (X_j) > F\ (X_i)\right)$ then
19. Update the position using (8)
20. End if
21. for each current best word
22. Execute the Dispersion and Ruthless behaviour using (9) (10)
23. End for
24. t = t+1
25. Go to step 17
26: End while
27: Obtain the global best keywords
28: End for
29. Return (optimal keywords)
30. For each keyword
31. Extract similar words from LLD'
32. Construct the vector model using (12)
33. End for
34. For each vector model
35. for each word in the testing set
36. Compute the pattern matching score using (13)
37. $If\ (\varphi_{pm} = 1)$ then
38. tweet is classified as 'depression'
39. else
40. tweet is classified as ' no depression'
41. End if
42. End for
43. End for
44. Obtain final classification results using the sigmoid activation function at the output layer
End

In this way, applicable as well as unrelated features are said to be characterized. Algorithm 1 exposed multivariate feature selection. Granger Causality functions are applied to the input vector matrix to select the highly correlated oceanographic and meteorological features from the given El Nino Dataset as input, and then the Serial Auto Correlation function is applied to the correlated features with the objective of identifying the highly correlated time range with respect to the temporal aspect.

With the multivariate feature selection proposed in the above algorithm, dimensionality reduction is said to be achieved by selecting the most correlated features. As far as the temporal dimension is concerned, the highly correlated time instance to the target time instance is analyzed. With this, not only is the input feature variable's dimensionality said to be reduced, but it also discards the irrelevant noise data, which in turn can result in the marine weather prediction accuracy in a timely manner.

3.3. Camargo Index Jensen-Shannon Boosting Classifier

Several of the ensemble classifiers concentrate [2] on observing meteorological readings in the error of the classifiers without taking into consideration changes in the distribution of data. In the proposed method, a dual window concept is adopted that compares the distribution of data between two successive windows. In this work, with the selected relevant features, the Camargo Index Jensen-Shannon Boosting Classifier is utilized for enhanced marine weather forecasting. The Jensen-Shannon Boosting Classifier is an ensemble technique where the Camargo Index Quadratic Classifier is used as a weak learner. Accurate weather forecasting results are identified to join the results of a weak learner via an ensemble classifier with minimal error.

Figure 3 shows the structure of the Camargo Index Jensen-Shannon Boosting Classifier model. As shown in the above Figure, the proposed model deploys novel change detection on the basis of Jensen–Shannon divergence. To predict marine weather according to latitude and longitude from the approximate location, oceanographic as well as surface meteorological variables computed. For detecting variances between previously established marine as well as exterior climatic readings and current distributions, Jensen–Shannon divergence is utilized. Let 'P' represent the input feature space, '$P \in R^r$' and '(p_t, q_t)' denotes the meteorological reading instance at time point 't' where '$p_t = (p_t^1, p_t^2, \dots, p_t^r) \in R^r \sqsubseteq RF$' allocated by labels set '$q_t = (q_t^1, q_t^2, \dots, q_t^{:}) \in \{0,1\}^L$', here '$q_i^j = 1$' if the '$j - th$' label is relevant and '$q_i^j = 0$'' When '$j - th$' label immaterial. The objective of this classification model is to predict marine weather forecasting. Let 'RW' and 'CW' be references as well as the current window, with the size of windows being 'n'. Null hypotheses '$H0$' against alternative hypotheses '$H1$' are decided for modifying the observation issue in meteorological readings detection.

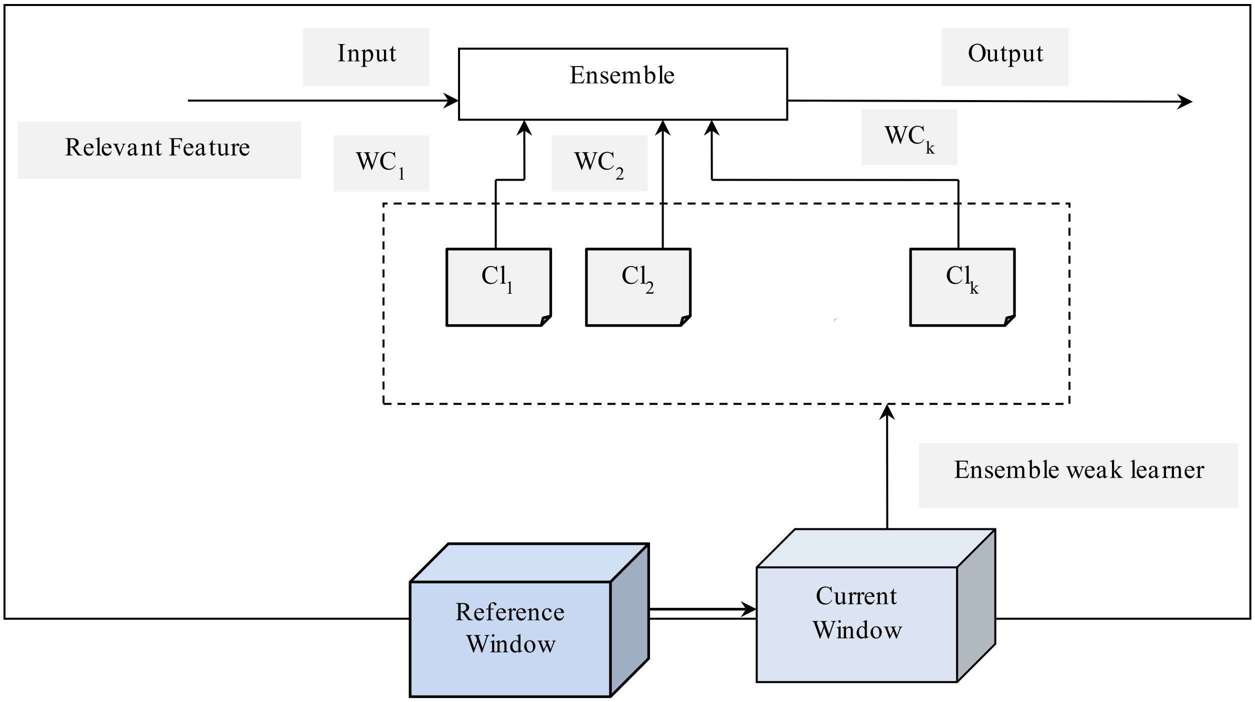

Fig. 3 Structure of camargo index jensen-shannon boosting classifier model

$$H0\ Dis\ (RW, CW)\ \leq\ Si\ [lat]\ +\ Sj\ [5°][long] \quad (6)$$

$$H0\ Dis\ (RW, CW)\ >\ S_1[lat]\ +\ S_1[5°][long] \quad (7)$$

From the above Equations (6) and (7), the distance function '$Dis(RW, CW)$' is to estimate the difference between two windows. Here, values in the reference window denote the previous day's meteorological readings, and the current window denotes the current day's meteorological readings. Based on these two window values, a prediction can be made efficiently. This is performed with a threshold factor of latitude as well as longitude data. First, the Camargo Evenness Index Quadratic function employed to obtain ensemble results with the hypotheses (7) is mathematically expressed as given below.

$$WC = CE = 1 - \left[\sum_{i=1}^{S}\sum_{j=i+1}^{S}\left(\frac{Cl_i - Cl_j}{S}\right)\right] \quad (8)$$

From the above Equation (8), 'Cl_i' represents the proportion of classes 'i' in the sample 'S', 'Cl_j' denotes the proportion of classes 'j' in the sample 'S' with 'S' denoting the overall sample used for simulation. Here, the Camargo Evenness Index Quadratic function forms the weak classifier.

Finally, Jensen-Shannon divergence is employed to identify the meteorological readings between two discrete distributions '$Dis1(a)$' and '$Dis2(a)$'.

$$JS(Dis1||Dis2) = \sum\left[Dis1(a)\log\frac{2Dis1(a)}{Dis1(a)Dis2(a)} + Dis2(a)\log\frac{2Dis2(a)}{Dis1(a)Dis2(a)}\right] \quad (9)$$

In general, buoys move around to different locations to predict weather, which can be determined based on two factors, i.e., latitude 'lat' and longitude '$long$'. The input of the classifier consists of relevant features according to distinct samples and the window size 'n'.To be more specific, if the previous day's meteorological readings and the current day's meteorological readings are the same, then the corresponding weather on that day repeats for the next day and vice versa. If any changes are said to be detected, then an abnormality is found to occur. A distinct classifier was constructed for relevant feature samples. Set of incoming instances forecasted. As of a weak classifier 'WC_i', error prediction 'Err_i' measured given below.

$$Err_i = \sum_{i=1}^{n} W_i . Point\ [H_i(WC_i)] \quad (10)$$

Where, pointer function is'$Point[.]$' to produce result is'1' when innermost expression outcome is true. Otherwise, '0' while outcome false with weight 'W' at first put to '1'.

The pseudo-code representation of the Camargo Index Jensen-Shannon Boosting Classifier for enhanced marine weather forecasting. An ensemble boosting classifier using the Camargo Index Jensen-Shannon is designed. First, with the designed hypothesis, to obtain ensemble results, the Camargo Evenness Index Quadratic function is applied to the relevant feature samples. Second, with the identified results, Jensen-Shannon divergence is utilized for identifying meteorological readings between two discrete distributions. Finally, error prediction is made in the case of a weak learner, therefore obtaining enhanced results.

4. Simulation Settings

Accuracy, time, and error rate are employed to estimate performance. Outcomes are measured, such as (1) performance of dissimilar categorization algorithms in terms of space complexity, (2) the time and accuracy in terms of marine weather forecasting time and marine weather forecasting accuracy and (3) the error rate with respect to distinct numbers of marine data. Simulations conducted in Python.

4.1. Performance Analyzis of Marine Weather Prediction Accuracy

The first parameter of significance for enhanced marine weather forecasting is accuracy. Method efficiency can be validated using the accuracy rate.

$$MWP_{acc} = \sum_{i=1}^{n} \frac{S_{AP}}{S_i} \tag{11}$$

From the above Equation (11), 'MWP_{acc}' is marine weather prediction accuracy, 'S_i' indicates samples or observations, 'S_{AP}'is samples accurately predicted. 'MWP_{acc}' determined as a percentage (%). Marine weather prediction accuracy is shown in Table 2. Figure 4, given above, shows the accuracy of marine weather prediction using 170000 samples or observations on the x-axis and their corresponding accuracy rate on the y-axis. From the above graphical representation, a steady flow is observed using all three methods. With simulations performed for 17000 samples with oceanographic and surface meteorological readings obtained from buoy series positioned throughout the equatorial Pacific, 16535 samples were correctly predicted using DGC-CJSBC, 16025 samples were correctly predicted using [1], and 15835 samples were correctly predicted using [2]. As a result, the overall marine weather prediction accuracy for 17000 samples using the three methods was observed to be 97.26%, 94.26% and 93.14% respectively.

Table 2. Marine weather prediction accuracy (%)

Number of sample instances (number)	Marine weather prediction accuracy (%)		
	DGC- CJSBC	SFA-LSTM	TCPBCCM ECL
17000	97.26	94.26	93.14
34000	96.35	93.15	92.15
51000	96	93.05	91
68000	95.85	92.55	89.5
75000	95.25	92	88
102000	94.75	91.55	87.25
119000	94	91	87
136000	93.75	90	86.35
153000	93.25	88.35	85
170000	92	86	83

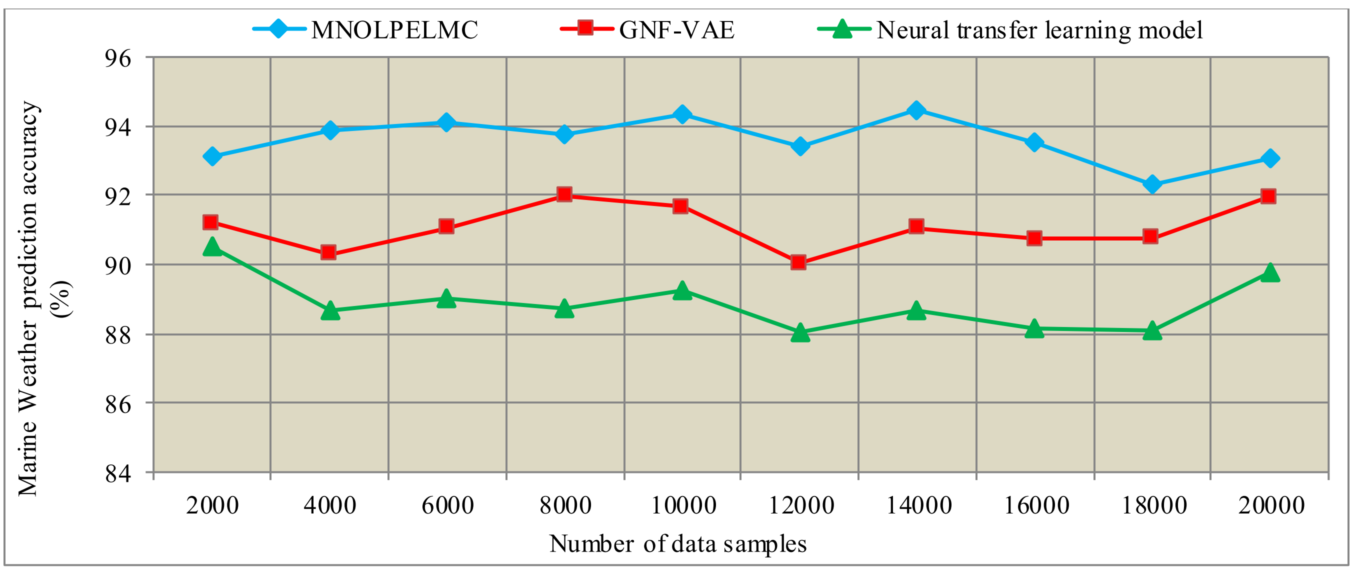

Fig. 4 Marine weather prediction accuracy

With this result, the accuracy of marine weather prediction using the DGC-CJSBC method is found to be comparatively better than [1, 2]. Correlations between oceanographic and surface meteorological readings were evaluated using Discriminant Granger Causality Regression.

As a result, highly correlated features were selected for further processing. This, in turn, resulted in the improvement of marine weather prediction accuracy using the DGC-CJSBC method, which was 4% compared to [1] and 8% compared to [2], respectively.

4.2. Performance Analysis of the Marine Weather Prediction Time

The prediction time has a second significance metric. This metric is of high significance because early the prediction is made, the faster mechanisms can be taken in case of an emergency.

$$MWP_{time} = \sum_{i=1}^{n} S_i * Time\ [prediction] \qquad (12)$$

From the above Equation (12), 'MWP_{time}' is marine weather prediction time, '$Time\ [prediction]$' is time needed for actual prediction. 'MWP_{time}' compute milliseconds (ms). Experimental results measured in terms of marine weather prediction time are shown in Table 3.

Table 3. Weather prediction time

Number of data samples	Weather prediction time		
	MNOLPELMC	GNF-VAE	Neural transfer learning model
2000	93.1	91.2	90.5
4000	93.85	90.32	88.65
6000	94.11	91.05	89.03
8000	93.75	91.98	88.73
10000	94.36	91.65	89.23
12000	93.40	90.05	88.05
14000	94.47	91.05	88.65
16000	93.55	90.73	88.13
18000	92.31	90.77	88.11
20000	93.05	91.94	89.78

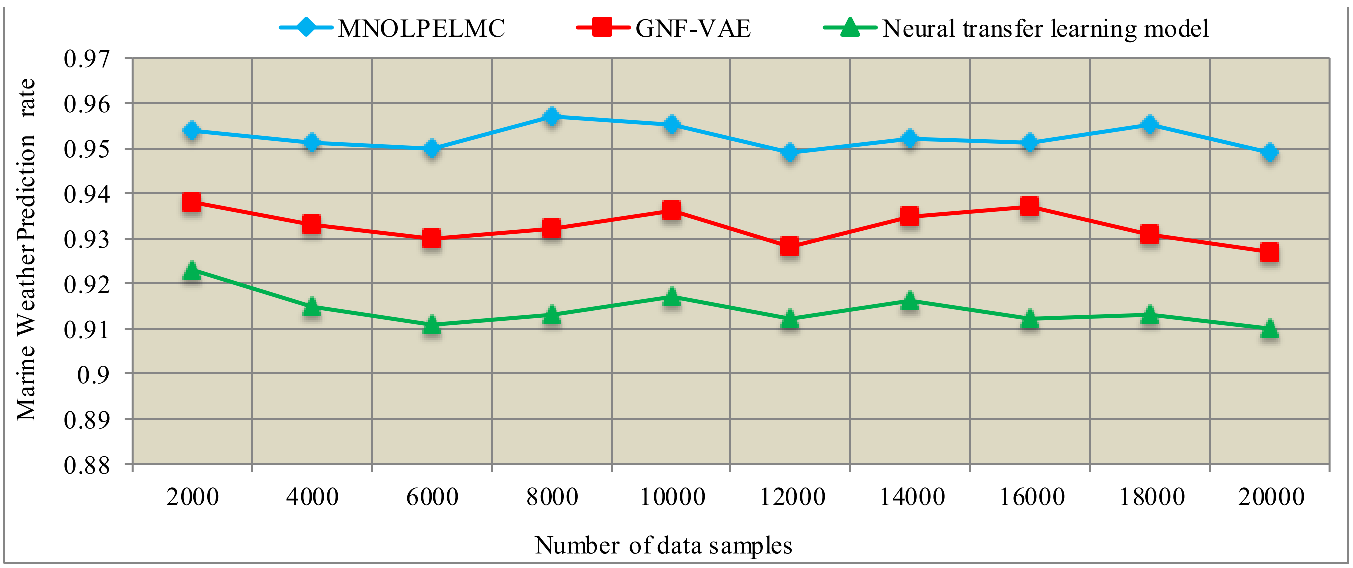

Fig. 5 Marine weather prediction rate

Figure 5, given above, shows the graphical representation of marine weather prediction time on the y-axis with respect to 170000 samples for simulation in the x-axis. From the above Figure, using all three methods, DGC-CJSBC, SFA-LSTM [1] and TCCMECLPBC [2], increasing the number of sample instances resulted in an increase in the amount of data available, which is also dependent on the buoy, as certain buoys were commissioned earlier than others.

This, in turn, results in an increase in marine weather prediction time when the number of sample instances increases. However, with simulations performed with 17000 samples, 0.35ms was said to be consumed for prediction using DGC-CJSBC, 0.50ms was said to be consumed using [1], and 0.55ms was said to be consumed for prediction using [2]. With these inferences, the marine weather prediction time using the DGC-CJSBC method was said to be comparatively reduced than [1, 2] by Discriminant Granger Causality Regression, with the temporal dimension taken into consideration, wherein only highly correlated time instance readings were utilized for further processing.

As a result, by reducing the feature variables dimensionality, also irrelevant noise data were eliminated, which in turn reduced the marine weather prediction time using the DGC-CJSBC method by 21% compared to [1] and 30% compared to [2], respectively.

4.3. Marine Weather Prediction Error Rate

During marine weather prediction, a small error is said to occur owing to the presence of noise. Therefore, predicting error rate becomes the most significant aspect while readings are considered from a sequence of floats located during the equatorial Pacific.

$$MWP_{Err} = \sum_{i=1}^{\Sigma i} \frac{S_{WP}}{S_i} \qquad (13)$$

From the above Equation (13), 'MWP_{Err}' is the marine weather prediction error rate, 'S_{WP}' represents samples wrongly predicted. 'MWP_{Err}' evaluated by percentage (%). Marine weather forecast error rates are shown in Table 4.

Table 4. Weather prediction error rate

Number of data samples	Weather Prediction		
	MNOLPELMC	GNF-VAE	Neural transfer learning model
2000	0.968	0.953	0.937
4000	0.96	0.951	0.932
6000	0.962	0.945	0.925
8000	0.963	0.949	0.928
10000	0.958	0.935	0.917
12000	0.965	0.942	0.922
14000	0.955	0.938	0.916
16000	0.963	0.947	0.929
18000	0.968	0.945	0.926
20000	0.965	0.942	0.925

The above shows the marine weather prediction error rate observed using sample instances ranging between 17000 and 170000. However, with simulations performed for 17000 samples, 115 samples were wrongly predicted using DGC-CJSBC, 185 samples were wrongly predicted using [1], and 235 samples were wrongly predicted using [2]. Marine weather prediction error rate was observed to be 0.67%, 1.08%, and 1.38%, using the three methods.

Owing to the application of the Jensen-Shannon Boosting Classifier, which combines the results of the weak learners. With this ensemble classifier, a pool of classifiers is maintained, and upon identification of changes in the consecutive oceanographic and surface meteorological readings down to a depth of 500 meters they are added to the pool. As a result, the marine weather prediction error rate using the DGC-CJSBC method is said to be reduced by 42% compared to [1] and 55% compared to [2].

4.4. Performance Analysis of Space Complexity

Finally, space complexity refers to the space occupied during ensemble learning. This is because while performing an ensemble, a certain amount of space is said to be occupied, which is referred to as space complexity. Space complexity is mathematically stated below.

$$SC = \sum_{i=1}^{n} S_i * Mem\ [Cl(results)] \tag{14}$$

From the above Equation (14), space complexity 'SC' is measured as well as '$Mem\ [Cl(results)]$' is memory required during overall prediction 'SC' discovered by kilobytes (KB). Space complexity results are shown in Table 5.

Table 5. Weather prediction complexity

Number of Data Samples	Weather Prediction Complexity		
	MNOLPELMC	GNF-VAE	Neural Transfer Learning Model
2000	3.08	3.93	4.24
4000	3.88	6.11	7.17
6000	4.55	6.93	8.49
8000	5.59	7.16	10.07
10000	5.64	8.35	10.77
12000	7.22	10.89	13.08
16000	8.15	11.71	15
18000	10.30	12.37	15.94
20000	9.82	11.39	14.45

Finally, Figure 6 illustrates the graphical representation of space complexity involved in enhanced marine weather forecasting. While forecasting marine weather, the intermittent results of the reference window and current window (i.e., oceanographic and surface meteorological readings) have to be stored in the stack for further processing. While doing this storage, a portion of space is said to be consumed for storing the results of this intermittent window. However, with simulations performed for 17000 samples, 2550KB was consumed when applied with DGC-CJSBC, 3400KB using [1], and 4250KB using [2]. With this result, the space complexity using the DGC-CJSBC method was found to be better than [1, 2]. Ensemble boosting classifier employing the Camargo Index Jensen-Shannon is formulated. Second, the Camargo Evenness Index Quadratic function was applied to the relevant feature samples as input. Third, Jensen-Shannon divergence was utilized for obtaining meteorological readings between discrete distributions, in turn reducing the error prediction using the DGC-CJSBC method by 24% compared to [1] and 36% compared to [2].

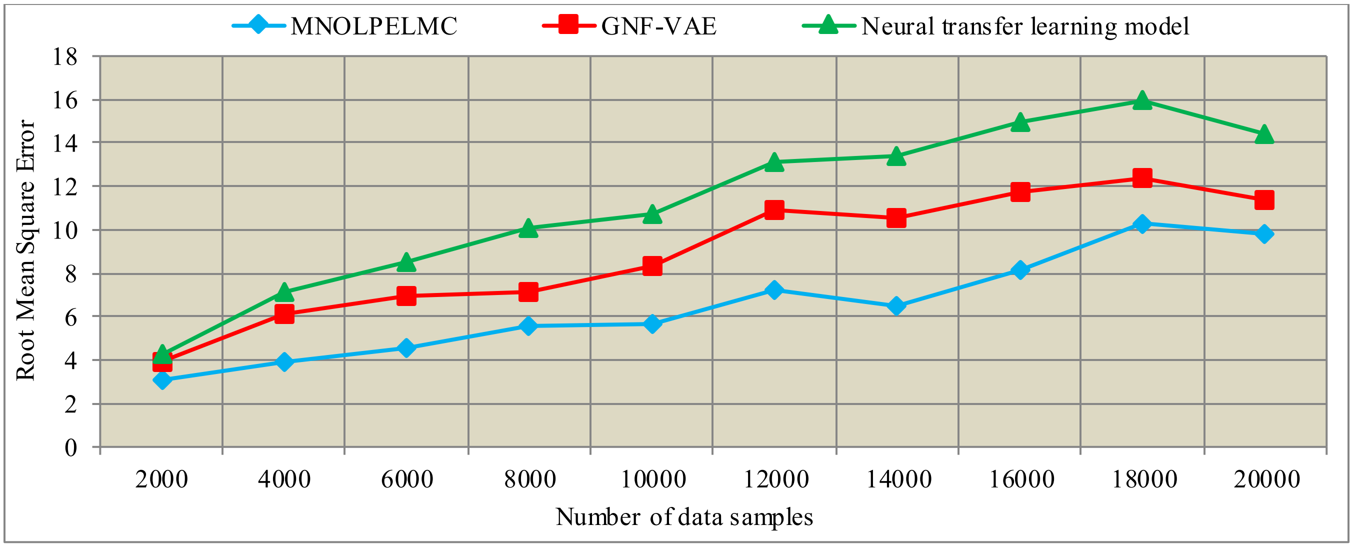

Fig. 6 Root mean square error

5. Conclusion

In this paper, in order to enhance marine weather prediction, a novel method called Discriminant Granger Causality and Camargo-Index Jensen Shannon Boosting Classifier (DGC-CJSBC) is proposed. Univariate and bivariate regression analysis were employed to choose highly correlated relevant features and eradicate immaterial features.

The Serial Auto Correlation function picks relevant features. For obtaining a strong classifier outcome, Camargo Index Jensen-Shannon Boosting Classification using Jensen-Shannon divergence likelihood function. Contrary to conventional approaches, the proposed DGC-CJSBC method is determined in terms of marine weather prediction accuracy, marine weather prediction time, error rate, and space complexity. The outcome of DGC-CJSBC provides superior performance to existing methods. The proposed method failed to consider the data pre-processing method for eliminating the noisy data. In future work, the data pre-processing method will be introduced to forecast weather to remove noisy data and reduce the dimensionality of the dataset with less time and space complexity.

References

[1] Masooma Ali Raza Suleman, and S. Shridevi, "Short-Term Weather Forecasting Using Spatial Feature Attention Based LSTM Model," *IEEE Access*, vol. 10, pp. 82456-82468, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[2] Pooja S.B et al., "Techniques Tanimoto Correlated Feature Selection System and Hybridization of Clustering and Boosting Ensemble Classification of Remote Sensed Big Data for Weather Forecasting," *Computer Communications*, vol. 151, pp. 266-274, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[3] Julio Cesar Santos Dos Anjos et al., "Data Processing Model to Perform Big Data Analytics in Hybrid Infrastructures," *IEEE Access*, vol. 8, pp. 170281-170294, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[4] Velumani Thiyagarajan, "An Intelligent Early Detection of Melanoma Using Fuzzy Neural Networks with Java Optimization (FNN-JO) Classifier," *International Journal of Engineering Trends and Technology*, vol. 72, no. 7, pp. 75-82, 2024. [CrossRef] [Publisher Link]

[5] Anshuka Anshuka et al., "Spatio Temporal Hydrological Extreme Forecasting Framework using LSTM Deep Learning Model," *Stochastic Environmental Research and Risk Assessment*, vol. 36, no. 10, pp. 3467-3485, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[6] Kalpesh R. Patil, and Masaaki Iiyama, "Deep Learning Models to Predict Sea Surface Temperature in Tohoku Region," *IEEE Access*, vol. 10, pp. 40410-40418, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[7] Alec S. Dyer et al., "Applied Machine Learning Model Comparison: Predicting Offshore Platform Integrity with Gradient Boosting Algorithms and Neural Networks," *Marine Structures*, vol. 83, pp. 1-16, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[8] Velumani Thiyagarajan et al., "Adaptive Feature Learning for Robust Pathogen Detection in Plants," *2024 4th International Conference on Sustainable Expert Systems (ICSES)*, Kaski, Nepal, pp. 1347-1353, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[9] Cheng Yin et al., "Advanced Machine Learning Techniques for Predicting Nha Trang Shorelines," *IEEE Access*, vol. 9, pp. 98132-98149, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[10] D. Menaka, and Sabitha Gauni, "Prediction of Dominant Ocean Parameters for Sustainable Marine Environment," *IEEE Access*, vol. 9, pp. 146578-146591, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[11] V. Prema et al., “Critical Review of Data, Models and Performance Metrics for Wind and Solar Power Forecast,” *IEEE Access*, vol. 10, pp. 667-688, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[12] Kathiresh Murugesan et al., “Homomorphic Encryption, Privacy-Preserving Feature Extraction, and Decentralized Architecture for Enhancing Privacy in Voice Authentication,” *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 15, no. 2, pp. 2150-2160, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[13] K.U. Jaseena, and Binsu C. Kovoor, “Deterministic Weather Forecasting Models based on Intelligentpredictors: A Survey,” *Journal of King Saud University - Computer and Information Sciences*, vol. 34, no. 6, pp. 3393-3412, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[14] Nicolas Guillou, and Georges Chapalain, “Machine Learning Methods Applied to Sea Level Predictions in the Upper Part of a Tidal Estuary,” *Oceanologia*, vol. 63, no. 4, pp. 531-544, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[15] Lei Zhang et al., “Weather Radar Echo Prediction Method based on Convolution Neural Network and Long Short-Term Memory Networks for Sustainable E-Agriculture,” *Journal of Cleaner Production*, vol. 298, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[16] Thiyagarajan Velumani, “Comprehensive Analysis of Cognitive Radio Technologies and its Applications,” *International Conference on Soft Computing and Signal Processing*, Hyderabad, India, vol. 2, pp. 103-115, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[17] Abdelhameed Ibrahim et al., “Wind Speed Ensemble Forecasting Based on Deep Learning Using Adaptive Dynamic Optimization Algorithm,” *IEEE Access*, vol. 9, pp. 125787-125804, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[18] Pradeep Hewage et al., “Deep Learning-based Effective Fine-grained Weather Forecasting Model,” *Pattern Analysis and Applications*, vol. 24, no. 1, pp. 343-366, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[19] Partha Pratim Sarkar, Prashanth Janardhan, and Parthajit Roy, “Prediction of Sea Surface Temperatures using Deep Learning Neuralnetworks,” *SN Applied Sciences*, vol. 2, no. 8, pp. 1-14, 2020. [CrossRef] [Google Scholar] [Publisher Link]