International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 11 Issue: 11 | Nov 2024 www.irjet.net p-ISSN: 2395-0072

A REGRESSION-BASED PREDICTIVE MODEL FOR ESTIMATION OF RAINFALL IN WEST BENGAL

Ayushi Chakraborty
Department of ECE
Techno International New Town
Kolkata, India
chakrabortyayushi8@gmail.com

Jyoti Chowdhury
Department of ECE
Techno International New Town
Kolkata, India
jyotichowdhury888@gmail.com

Vivekananda Mukherjee
Department of ECE
Techno International New Town
Kolkata, India
vivekananda.mukherjee@tict.edu.in

Ardhendu Shekhar Biswas
Department of ECE
Techno International New Town
Kolkata, India
a.s.biswas@tint.edu.in

Md Anoarul Islam
Department of ECE
Techno International New Town
Kolkata, India
md.anoarul.islam@tint.edu.in

Manabendra Maiti
Department of ECE
Techno International New Town
Kolkata, India
dr.manabendra.maiti@tict.edu.in

***Abstract*—Understanding the long-term trends and variations in annual rainfall is crucial for effective crop planning and water resource management in West Bengal. Our study delved into a period of rainfall data (from 2004 to 2023), meticulously analyzing the patterns and changes in rainfall across the districts of this region. This paper enhances a a non-linear piecewise linear slope estimation technique for rainfall prediction combined with an automated method of estimation usually considered in a regression framework. Additionally, the approach has been modified to account for initialization errors, further improving the accuracy of the estimates.**

Keywords— rainfall estimation, water resource management, regression based method, machine learning, slope estimation

I. INTRODUCTION

India's economy depends heavily on agriculture, and the country's success is largely dependent on its agricultural production. However, rainfall is a major factor in agricultural output, therefore forecasting rainfall in advance is critical to economic growth and stability. Rainfall forecasting has proven to be extremely difficult globally, especially in recent years.

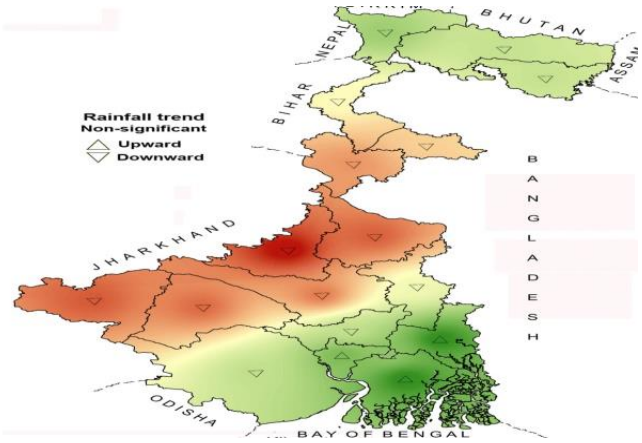

Figure_1. Variation of Average rain fall in West Bengal

The proper estimation of rainfall is a highly relevant research problem in the present weather scenario. Because of global warming, the rise in average temperature every year and late present of rainy season, heavy rainfall, flood ultimately affects the country's crop production. The depletion of natural water sources and growing environmental concerns has prompted many researchers to focus on demand-matched water resource management. Therefore, determining the precise amount of rainfall needed in different geographic areas—whether they be cities, states, or nations—is crucial. The creation of estimate methods and algorithms, particularly those that use machine learning and artificial intelligence techniques, has been the focus of current research in this area. Based on past data, these methods seek to produce precise predictions of future rainfall. The design of water distribution systems and policy-level choices about the management of water resources are two examples of how such estimates might be used to manage water resources economically. This study examines precise estimating algorithms for rainfall forecasting and suggests innovative methods in this field.

II. LITERATURE SURVEY

Rainfall is a vital lifeline for communities dependent on its nourishing touch, making accurate predictions essential for their survival and prosperity. In recent years, the art and science of forecasting rainfall have captured the interest of government bodies, industries, risk assessment agencies, and researchers alike. Knowing when and how much it will rain is not solely for weather preparedness as it serves the purpose of saving people's life and property as well as the economy. As per the article A Survey on Rainfall Prediction Techniques published by Hirani (D.) and Mishra (N.) [1], rainfall is predicted beforehand employing several techniques including certain cognitive techniques and various machine learning approaches. The study focus also

makes use of dynamical and empirical model approaches.

Likewise, in their study "Prediction of Rainfall Using Data Mining Techniques Over Assam," Dutta Saikia and Tahbilder Hitesh [2] elaborate on data mining techniques as auxiliary towards the Monthly Rainfall Prediction using Multiple Linear Regression, along with the normal exercise.

According to a research by V. Brahmananda Rao and K. Hada [3], "An Experiment with Linear Regression for Forecasting Spring Rainfall Over South Brazil." This study employs a Root Mean Square Error (RMSE) method to forecast rainfall, especially in South Brazil.

Likewise, Panchani, Ankita, and collaborators [4] in their study "Prediction of Rainfall Using Image Processing" provide a technique for rainfall prediction based on digital cloud photographs. They contend that when economic and security considerations are taken into account, computerised cloud images—as opposed to satellite images—can forecast rainfall with greater accuracy. K-Means Clustering is used to identify the kind of cloud, and a Cloud Mask Algorithm is used to establish the cloud status. By examining the colour and density of the cloud photos, which are saved in JPEG format, they are able to determine the kind of rainfall cloud. In their research Mulubrhan Amare et al. [5] quantify the effects of rainfall variability using a negative rainfall shock index. In their investigation of the impacts of climatic variability in rural Mexico, Conroy, Skoufias, and Vinha [6] define weather shocks as situations in which rainfall or growing degree days diverge by more than one standard deviation from their respective averages.

They used historical data from several meteorological stations around Mexico to create a measure for rainfall variance. Using a power regression method, N. Sen [8] has created a long-range forecast model for summer monsoon rainfall prediction. El Niño, Eurasian snow cover, northwest European temperatures, the European pressure gradient, South Indian Ocean temperatures, Arabian Sea surface temperatures, East Asian pressure, and the 50 hPa wind pattern from the previous year are all included in this model. The experimental findings showed a 4% model inaccuracy.

The development of a statistical forecasting technique for Thailand's summer monsoon rainfall (SMR) was detailed by S. Nkrintra and associates [9]. They use nonparametric techniques based on local polynomials and multiple linear regression. The El Niño Southern Oscillation Index (ENSO), wind speed, sea surface temperature (SST), sea level pressure (SLP), and the Indian Ocean Dipole (IOD) are important predictions. The results of their trials indicated that the predicted and actual rainfall had a 0.6 correlation coefficient.

A prediction model for high rainfall occurrences in South Korea was created by T. Sohn and colleagues [10] using artificial neural networks, decision trees, multiple linear and logistic regression, and more. A numerical model produced 45 synoptic factors, which they took into consideration as possible predictors.Data-driven (empirical) methods have become more popular recently than knowledge-driven (physical) methods. This change has created new opportunities, especially in time series analysis with the introduction of deep neural networks. A mainstay of time series forecasting for a long time is the venerable autoregressive integrated moving average model (ARIMA), sometimes referred to as the Box–Jenkins model.

In hydrological forecasting, models like as the Autoregressive (AR), Autoregressive Moving Average (ARMA), and ARIMA have become essential tools due to their development; ARIMA is particularly renowned for its dependability and resilience. The effectiveness of statistical models, machine learning algorithms, and deep learning techniques—specifically, Long Short-Term Memory (LSTM) models—in rainfall prediction is assessed by S. D. Latif and colleagues [11]. Because of their limited research, it highlights the benefits of machine learning across a variety of climates and timelines and recommends more research into remote sensing and hybrid models. The study addresses the integration of satellite, radar, and ground-based data for enhanced prediction accuracy and emphasises the use of RMSE, R^2, and MAE metrics for evaluating model correctness. To anticipate weather events, Nolan et al. [12] specifically predicted wet and dry days in Australia for the next day. They identified the primary factors influencing the weather by using a DTM (Decision-Tree Model) with capstone analysis. The capstone decision-tree model forecasted the weather with the highest accuracy grade of 87.9%. Furthermore, the accuracy rating of 75.6% for the combined-city model suggests that the approach may be useful in a wide range of geographic areas. To find out how effectively machine learning and statistical methods predict rainfall. Balamurugan and Manoj Kumar [13] carried out a comparison research. According to the Indian Meteorological Department (IMD), they found that the percentage of rainfall departure for the month of June 2019 varied from 46% to 91%. However, their research showed that machine learning approaches performed better in rainfall prediction than statistical methods. The total accuracy of logistic regression was 84.6%, while the ROC statistical approach and the Decision Tree statistical strategy were 72.6% and 77.6%, respectively. These findings highlight how well machine learning algorithms capture the intricate correlations and patterns seen in rainfall data.

The use of logistic regression modelling to forecast rainfall for the next day was investigated by Ejike et al.

[14]. They made use of a year's worth of weather data from Canberra, Australia, which included wind speed, direction, temperature, pressure, humidity, sunshine, evaporation, and cloud cover. The findings showed that rainfall for the next day can be predicted using logistic regression with an accuracy of 87% when suitable meteorological parameters are included. This result emphasises how important it is to include pertinent elements in the modelling process in order to get precise rainfall forecasts.

Neural network models were created by Kumarasiri and Sonnadara [15] to predict rainfall on various time scales. They developed a model that predicted the rainfall for the following day with an accuracy of 74.30%. They also created a model for yearly rainfall depth estimates one year in advance, which produced an accuracy of 80.0% within a 5% error margin. Furthermore, forecasts for several time steps into the future were made using these models. The results point to neural networks' potential for accurately forecasting rainfall patterns across a range of time periods and capturing their temporal dynamics.

Using machine learning algorithms and a dataset taken from the Bangladesh Jatiyo Tottho Batayon website, Ria et al. [16] carried out a study on rain prediction. To forecast rainfall, they trained and assessed five distinct models. Before being used to validate rainfall estimates, each model was trained using eight weather-related input attributes. The study's findings showed that, when compared to other models, the Random Forest classifier had the best accuracy of 86%, indicating how well it captured the intricate correlations between input factors and rainfall patterns.

Using real-time data sets, Neelakandan and Paulraj [17] suggested a prediction model based on CA-SVM for rainfall forecasting. To verify their model, they used a 12-month period of yearly rainfall data from RMC, Chennai. The study employed two learning models: the local learning model and the dynamic learning model. When compared to practical methods, the proposed technique showed that it could accurately estimate rainfall by merging several data points without overlap, obtaining an accuracy of 88.9%. In order to forecast floods based on a number of variables, including temperature, precipitation, water velocity, water level, humidity, and ANN (Artificial Neural Networks) with a single hidden layer, Sankaranarayanan et al. [18] used machine learning techniques. For flood forecasting, they used a deep neural network that included stream flow. In addition to other pertinent factors, the research gathered a significant amount of rainfall data. The DNN (deep neural network) beat the benchmark, attaining a better accuracy of more than 90%, according to the findings of a comparison of the accuracy attained using four distinct techniques. These results demonstrate the promise of DNN methods for precise flood forecasting.

In this study, a piecewise linear machine learning-based regression model for rainfall prediction is compared and contrasted with an estimating approach based on linear regression. In order to further reduce prediction error, it also uses power series-based estimation to improve the piecewise linear model.

III. Regression Model

The Gross Domestic Product (GDP) per unit of agricultural output is calculated using a data set of West Bengal's yearly rainfall from 1994 to 2023. A LRM (linear regression model) is then suggested for the same period. The regression equation 1 below, where "a(y)" stands for GDP for a year "y," contrasted with the average rainfall (r(y)). Slope is M(y), and the constant value for the year y is C(y). In conventional linear regression, every C(y) result is the same.

$$r(y) = m(y) * a(y) + C(y) \quad (1)$$

Equation 2 below displays the dataset's real regression equation, where rl(y) is the estimate of the actual rainfall a(y) for every given year y based on linear regression.

$$rl(y) = 122.3 * a(y) - 293.2 \quad (2)$$

The predicted values derived from the linear regression model are compared with the dataset for the years 2004 through 2014. Figure 2 illustrates how agricultural yield varies with average rainfall.

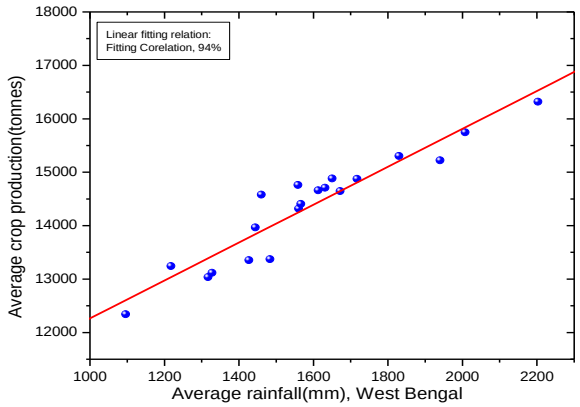

Figure_2. Variation of crop production with Average rain fall

Table 1 displays the dataset together with its projected values from 2004 to 2014.

Table 1: Annual rain fall data (2004-2023) comparison with linear regression estimate

Year	Annual Rain fall (mm)	Crop Production (tonns)
2003	1526.7	14389.3
2004	1612.9	14662.3

2005	1650.1	14884.8
2006	1631.4	14510.8
2007	2007.5	14745.9
2008	1560.1	14719.5
2009	1317.2	15037.3
2010	1096.0	14340.7
2011	1671.7	13045.9
2012	1566.0	14605.8
2013	1939.9	15023.7
2014	1483.5	15370.7
2015	1717.0	14677.2
2016	1427.0	15953.9
2017	1830.0	15302.5
2018	1444.1	14967.0
2019	1217.3	16242.2
2020	1460.0	15881.4
2021	2202.7	16520.0
2022	1558.8	16760.0
2023	1327.6	15419.2

A piecewise linear regression model has been suggested by the authors as a way to lower the estimation error. For slope estimation, the model employs a straightforward regression technique based on machine learning and the previously observed estimation error [4]. In equation 3, the key regression relation associated with this model is displayed. The year preceding y is denoted by (y-1) in this case, and the piecewise linear approximation of r(y) is rp(y).

$$\text{rp(y)} = \frac{rl(y)-rl(y-1)}{a(y)-a(y-1)} * a(y) - C(y) \qquad (3)$$

The aforementioned formula is used by a software application to calculate the estimate of average rainfall using the constant, the previous and current GDP, and the linear estimations. The approach shown above uses the variance in the expected slope to obtain the regression estimate for each year. Table 2 below shows the comparable data for estimates based on piecewise linear regression that were produced in this way.

.

Table 2: Annual rain fall data (2004-2023) compared with the estimate of piecewise linear regression

Year	Crop production (tonnes)	Annual rainfall (mm)	Estimated rainfall using linear regression(mm)	Estimation Error (%)
2004	14662.3	1612.9	1627.2	-0.89
2005	14884.8	1650.1	1669.98	-1.20
2006	14510.8	1631.4	1637.85	-0.40
2007	14745.9	2007.5	2038.30	-1.53
2008	14719.5	1560.1	1515.56	2.85
2009	15037.3	1317.2	1120.02	14.97
2010	15037.3	1096	631.42	42.39
2011	15037.3	1671.7	1692.14	-1.22
2012	15037.3	1566	1533.41	2.08
2013	15037.3	1939.9	1974.71	-1.79
2014	15037.3	1483.5	1384.57	6.67
2015	15037.3	1717	1731.72	-0.86
2016	15037.3	1427	1288.15	-9.73
2017	15037.3	1830	1869.28	-2.15
2018	15037.3	1444.1	1318.42	8.7
2019	15037.3	1217.3	1229.61	1.15
2020	15037.3	1460	1367.94	6.31
2021	15037.3	2202.7	2246.39	-1.98
2022	15037.3	1558.8	1467.8	5.84
2023	15037.3	1327.6	1295.25	2.5

The above table shows that, for the provided dataset, the piecewise linear model differs only marginally from the ordinary linear regression model. For more precise estimate, a better method must be suggested. For this reason, the authors have put forth a model that includes precise estimation of the change in the quasi-constant Cv(y). The piecewise linear estimation model is modified to employ the regression equation in this machine learning-based estimation approach. The model's average rainfall is expressed as rpc(y), while the linear regression model's prior prediction error is expressed as error(y-1) as a percentage. The model is formally represented by the matching equation 4.

.

$$\text{rpc(t)} = \frac{rl(y) - rl(y-1)}{a(y) - a(y-1)} * a(y) - Cv(y) \qquad (4)$$

$$\text{Cv(y)} = C(y) * \left\{\frac{100 - error(y-1)}{100}\right\} n - C(y) \qquad (5)$$

C(y) =293.2 is the value assumed for the current estimation. n is a positive number in this model.

In order to reduce the subsequent inaccuracy in prediction using prior historical data, the machine learning model iteratively attempts to determine the best-fit n. Additionally, it should be remembered that the model converges to the earlier piecewise linear estimating model for n=0. The following table 3 shows the related simulation data.

Table 3: Annual rain fall data (2004-2023) includes initialization error correction and a piecewise linear regression estimate

Year	Crop production (tonnes)	Annual rainfall (mm)	Estimated rainfall using piecewise linear regression(mm)	Estimation Error (%)
2004	14662.3	1612.9	1629.5	-1
2005	14884.8	1650.1	1626.4	1.4
2006	14510.8	1631.4	1631.6	0.01
2007	14745.9	2007.5	1628.3	18.9
2008	14719.5	1560.1	1628.7	-4.4
2009	15037.3	1317.2	1624.2	-23.3
2010	14340.7	1096.0	1633.9	-49.1
2011	13045.9	1671.7	1652.0	1.2
2012	14605.8	1566.0	1630.2	-4.1
2013	15023.7	1939.9	1624.4	16.3
2014	15370.7	1483.5	1619.6	-9.2
2015	14677.2	1717.0	1629.2	5.1
2016	15953.9	1427.0	1611.4	-12.9
2017	15302.5	1830.0	1620.5	11.4
2018	14967.0	1444.1	1625.2	-12.5
2019	16242.2	1217.3	1607.4	-32
2020	15881.4	1460.0	1612.5	-10.4

2021	16520.0	2202.7	1603.6	27.2
2022	16760.0	1558.8	1600.2	-2.7
2023	15419.2	1327.6	1618.9	-21.9

The next part explains the comparative analysis of the collected data.

IV. RESULT

When estimating mean rainfall, both machine learning-based simulation results perform better than conventional linear regression estimators. A comparison of rainfall estimates obtained from linear regression and actual rainfall is presented in Figure 3 below. The improved linear regression and partial linear regression models with initialised error correction plots are displayed in Figures 4 and 5.

.

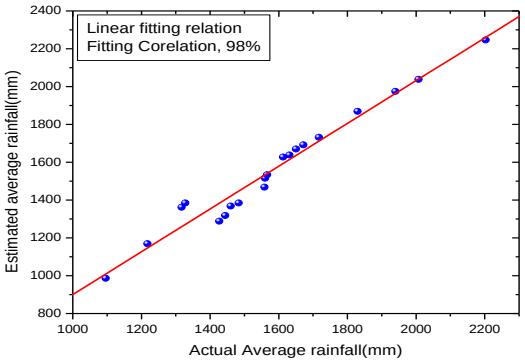

Figure_3. Actual rainfall compared to an estimate based on linear regression (2004-2023)

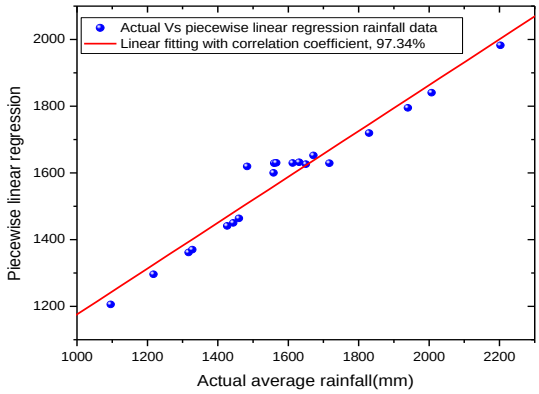

Figure_4. Actual rainfall compared to a prediction determined by piecewise linear regression (2004-2023)

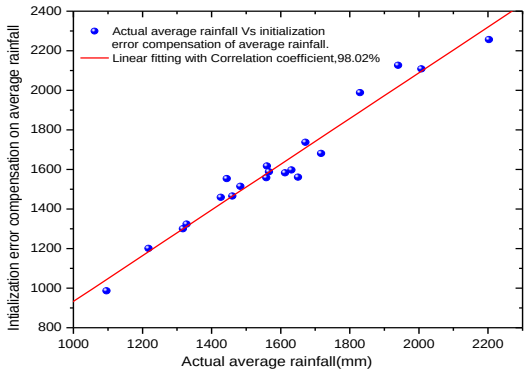

Figure_5. Average rainfall (2004–2023) as compared to a piecewise linear regression-based estimate with initialisation error adjustment

The below figure 6 presents the results of comparing the forecasting error percentages for the three methods.

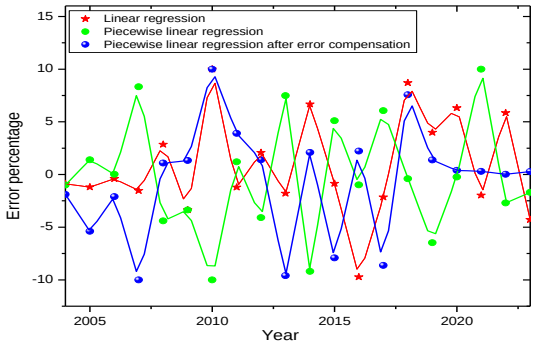

Figure_6. Analysis of linear regression, linear regression with initialisation error correcting errors, and piecewise linear regression (2004–2023)

Figure 6 makes it evident that out of the three methods, the modified piecewise linear methodology with initialisation error correction performs the best.

V. CONCLUSION

This paper's introduction of the piecewise linear regression approach should be improved by refining the slope estimation process. For instance, incorporating radius of curvature-based adaptive regression analysis methods within the machine learning framework could improve the accuracy of our results. Our findings demonstrate that the models presented are effective in predicting average rainfall, particularly when the data follows a relatively smooth polynomial trend. However, to achieve more precise predictions in economies characterized by oscillatory growth patterns, the model requires further adjustment. Looking ahead, Significant advancements in this field may also be possible by using cutting-edge techniques like deep learning algorithms and artificial neural networks.

ACKNOWLEDGMENT

Sincere thanks are extended by the authors to Techno International New Town's Department of Electronics and Communication Engineering for providing the research facilities that made this study possible.

REFERENCES

[1] Dhawal Hirani and Dr. Nitin Mishra "A Survey on Rainfall Prediction Techniques" International Journal of Computer Application (2250-1797) Volume 6– No.2, March- April 2016.

[2] Dutta, Pinky Saikia. "PREDICTION OF RAINFALL USING DATAMINING TECHNIQUE OVER ASSAM." (2014).

[3] V. Brahmananda Rao , K. Hada 1994: An experiment with linear regression in forecasting of spring rainfall over south Brazil.

[4] Panchani, Ankita et al. "PREDICTION OF RAINFALL USING IMAGE PROCESSING." (2014).

[5] Mulubrhan Amare, Nathaniel D. Jensen, Bekele Shiferaw, Jennifer Denno Cissé, Rainfall shocks and agricultural productivity: Implication for rural household consumption, Agricultural Systems, Volume 166, 2018, Pages 79-89, ISSN 0308-521X.

[6] Skoufias, Emmanuel and Vinha, Katja and Conroy, Hector V., The Impacts of Climate Variability on Welfare in Rural Mexico (February 1, 2011). World Bank Policy Research Working Paper No. 5555, Available at SSRN: https://ssrn.com/abstract=1754348.

[7] Meza-Pale, Pablo and Antonio Yúnez-Naude. "The Effect of Rainfall Variation on Agricultural Households: Evidence from Mexico." (2015).

[8] N. Sen, " New forecast models for Indian south-west Monsoon season Rainfall", in Current Science, vol. 84, No. 10, May 2003, pp.1290- 1291.

[9] S. Nkrintra, et al., "Seasonal Forecasting of Thailand Summer Monsoon Rainfall", in International Journal of Climatology, Vol. 25, Issue 5, American Meteorological Society, 2005, pp. 649-664.

[10] T. Sohn, J. H. Lee, S. H. Lee, C. S. Ryu, "Statistical Prediction of Heavy Rain in South Korea", in Advances in Atmospheric Sciences, Vol. 22, No. 5, 2005, pp.703-710.

[11] S. D. Latif, N. A. B. Hazrin, C. H. Koo, J. L. Ng, B. Chaplot, Y. F. Huang, A. El-Shafie, and A. N. Ahmed, "Assessing rainfall prediction models: Exploring the advantages of machine learning and remote sensing approaches," *Alexandria Engineering Journal*, vol. 82, pp. 16-25, 2023, doi: 10.1016/j.aej.2023.06.015.

[12] A. G. Nolan and W. J. Graco, "Using the results of capstone analysis to predict a weather outcome" in Advances in Data Mining. Applications and Theoretical Aspects, Singapore:Springer, pp. 269-277, 2017.

[13] M. S. Balamurugan and R. Manojkumar, "Study of short term rain forecasting using machine learning based approach", *Wireless Netw.*, vol. 27, no. 8, pp. 5429-5434, Nov. 2021.

[14] O. Ejike, D. L. Ndzi and A.-H. Al-Hassani, "Logistic regression based next-day rain prediction model", *Proc. Int. Conf. Commun. Inf. Technol. (ICICT)*, pp. 262-267, Jun. 2021.

[15] A. D. Kumarasiri and U. J. Sonnadara, "Performance of an artificial neural network on forecasting the daily occurrence and annual depth of rainfall at a tropical site", *Hydrological Processes*, vol. 22, no. 17, pp. 3535-3542, Aug. 2008.

[16] N. J. Ria, J. F. Ani, M. Islam and A. K. M. Masum, "Standardization of rainfall prediction in Bangladesh using machine learning approach", *Proc. 12th Int. Conf. Comput. Commun. Netw. Technol. (ICCCNT)*, pp. 1-5, Jul. 2021.

[17] S. Neelakandan and D. Paulraj, "RETRACTED ARTICLE: An automated exploring and learning model for data prediction using balanced CA-SVM", *J. Ambient Intell. Humanized Comput.*, vol. 12, no. 5, pp. 4979-4990, May 2021.

[18] S. Sankaranarayanan, M. Prabhakar, S. Satish, P. Jain, A. Ramprasad and A. Krishnan, "Flood prediction based on weather parameters using deep learning", *J. Water Climate Change*, vol. 11, no. 4, pp. 1766-1783, Dec. 2020.

[19] A. M. Bagirov and A. Mahmood, "A comparative assessment of models to predict monthly rainfall in Australia", *Water Resour. Manag.*, vol. 32, no. 5, pp. 1777-1794, Mar. 2018.