

# Annual Rainfall Prediction of Maharashtra State Using Multiple Regression



Loukik S. Salvi and Ashish Jadhav

**Abstract** This paper presents a study of Indian rainfall and prediction of the annual rainfall in the state of Maharashtra and Konkan. The decreasing trends in seasonal rainfall and post-monsoon rainfall and increasing occurrence of the deficit rainfall years indicates the probable intensification of water scarcity in many states and sub divisions of India. Rainfall serves a major source of water only when it is conserved, thus a proper analysis and estimation of rainfall globally is of utmost importance. In an agricultural country like India, where the majority of agriculture is rain dependent rainfall prediction can help to understand the uncertainty in rainfall pattern which may affect the overall agricultural produce. The present study consists a descriptive analysis of annual rainfall in India from 1950–2020, this visualization may prove helpful for deciding the right model for prediction. This study is aimed at finding the most apt model for making accurate prediction for the rainfall dataset. Two machine learning model and a neural network model are implemented and their results are compared. The performance of the results was measured with MSE (mean squared error), RMSE (root mean square error), MAE (mean absolute error). The machine learning models showed high level of deviation as the time series data in use was highly inconsistent, on the other hand the neural network showed better efficiency due to the local dependency in the model which helps it to learn and perform better.

**Keywords** Rainfall forecasting · SVR · ANN · Linear regression

## 1 Introduction

The process of analyzing and predicting the probability of precipitation and forecasting of rainfall in future along with estimating the amount of rainfall in specific

---

L. S. Salvi (✉) · A. Jadhav

Ramrao Adik Institute of Technology, Nerul D.Y. Patil Deemed to be University, Navi Mumbai, India

e-mail: [loukiksalvi96@gmail.com](mailto:loukiksalvi96@gmail.com)

A. Jadhav

e-mail: [ashish.jadhav@rait.ac.in](mailto:ashish.jadhav@rait.ac.in)

regions is called Rainfall Prediction. Alongside the likelihood of precipitation in that particular district, it likewise considers the precipitation volume assessment, exactness of forecast, mistake in expectation and. It is ready by the forecasters through social occasion, researching, affirming, showing, reproducing and investigating the different meteorological data and limits available. Precipitation is one of the six intrinsic bits of environment assumption and is moreover the most basic parts in the hydrologic. Storm in India is critical as it gives shape to its farming and economy. Because of the intrinsic intricacy of the actual cycles related with the forecast of Indian Summer Storm Precipitation (ISMR), it is perhaps the most confounding logical errand. However, research related to monsoon has improved and advanced significantly due to the ever increasing amount of data made available from the satellites, improved understanding of the processes, and enhanced computing resources. As an impact of changing climate, the spatiotemporal distribution of precipitation is getting modified since the last few decades. This has resulted into frequent droughts and floods within a spatial distance of a few hundred kilometers. Quantitative forecast of every day precipitation is a difficult assignment and is huge for a considerable length of time and functional applications. Over the most recent thirty years, the precision of rainstorm gauges has improved and the methodical mistake with conjecture length in the medium reach has decreased. In any case, the worry for precipitation conjecture for the Indian rainstorm locale stays as the capacity for jungles is still lower when contrasted with mid-scopes [1]. The expectation of the precipitation can be partition into two sections one being the transient forecast which considers month to month expectation and also the drawn out forecast which considers yearly expectation which ordinarily is troublesome as a result of high unconventionality and reliance of precipitation on different boundaries of precipitation. AI models are typically known to yield better proficiency in expectation in situations where the informational collection has assorted boundaries and has critical degree of consistency, though the precipitation information being a period variation and conflicting which is the reason the AI models normally don't perform well. The coming of profound learning and neural organization has filled in as a well-suited answer for such issues. The different degree of learning and neighborhood conditions in the neural organization model assist them with performing great on such fluctuating dataset with less boundaries. Please note that the first paragraph of a section or subsection is not indented. The first paragraphs that follows a table, figure, equation etc. does not have an indent, either.

## 2 Literature Review

(Table 1).

**Table 1** Review of relevant literature

Sr. No.	Title of the paper/journal/year	Findings
1	Annual and Non-monsoon Rainfall Prediction Modelling Using SVR-MLP: An Empirical Study From Odisha [1], IEEE Access, 2020	In this paper Odisha has been considered as the study area, the monsoon and non-monsoon rainfall prediction has been carried out using a proposed SVR-MLP model. The non-monsoon rainfall being unpredictable has been given utmost importance and proper modelled dataset has been used to ensure higher efficiency
2	Deep Learning Approach for Multi Step Ahead Daily Rainfall Prediction using GCM simulation [2]	In this paper the authors have proposed the Deep learning and neural network models for predicting the rainfall. The paper states that the DL and NN models show better efficiency over the traditional ML models for predicting the rainfall. The suggested model has also integrated the GSM simulator as the rainfall dataset has many meteorological variables
3	Forecasting of nonlinear time series using ANN [3], Future Computing and Informatics Journal (2017)	This paper imparts the importance of fuzzy logic over the common neural networks for recognizing the behavior of nonlinear or dynamic time series, in the proposed study the NARMA and ARNN are deployed and results are compared
4	Multi-model Prediction of Monsoon Rain Using Dynamical Model Selection [4], IEEE Transactions, 2016	In this paper the author has proposed a dynamic ensemble model for rainfall prediction, rainfall data being non-linear and time variant would need a dynamic model to study the irregularities
5	Summer Monsoon Rainfall Variability Over Maharashtra, India [5], Article in Pure and Applied Geophysics, 2012	This article presents a brief study of rainfall in Maharashtra for the last 30 years. The study has been conducted on the basis of the rainfall data from the meteorological stations in Maharashtra and the satellite images of cloud study. It has been observed that the rainfall has been irregular in the past ten years

### 3 Regression

Regression is module of data science which takes a statistical approach to solve the problems. Regression is categorized under the supervised learning approaches where strategies are made to predict the future variables. While calculating the future (unknown) variables, the known quantities and their relationships are taken into consideration and connections with the unknown variable are build. Simple regression is given by  $A = X + X1(B)$ , where  $A_n$  is the reliant variable, whose worth is to

be anticipated and B is the free factor whose esteem consistent worth and X1 is the improved coefficient.

### ***3.1 Support Vector Machines (SVM) for Regression***

The utilization of SVM for relapse is known as Support Vector Regression (SVR) [5]. This calculation consequently changes ostensible qualities over to numeric qualities. Input informational collection must be standardized prior to preparing start, either consequently (by apparatus arrangement or prearranging) or physically by the client (informational collection standardization). SVR observes a best fit line which decreases the mistake of the expense work. Just those occasions (Support Vectors) in the preparation informational collection are picked which are closest to the line with least expense. An edge must be made around the line for better change of forecast and afterward the information might be projected into higher layered space for better expectation and adaptability. The expense work limits the boundaries over the dataset.

Bit Functions are utilized to deal with the high dimensionality of the component space. Appropriate choice of Kernel capacity can create more powerful outcome or exactness in least time in this way expanding productivity of the model. Weka instrument utilizes different parts to accomplish this errand [2].

### ***3.2 Artificial Neural Network***

Artificial Neural Organization involves various processors working in equal and organized in levels. The principal level acknowledges the crude information. Rather than crude info, each successive level gets the result from the level going before it. The last level creates the result of the framework. The two significant conversations against ANN is that its asset concentrated and its results are difficult to decipher.

ANN is considered whenever the computational assets are not a constrictive and cost-restrictive. The ANN functions admirably when the informational collection utilized for preparing is colossal which the situation in precipitation dataset is normally. Besides the ANN model has stowed away layers and critical measure of neighborhood conditions which assists it with learning the conflicting examples and perform better on the time series information.

## **4 Study Area**

The area of Maharashtra lies on the western side of India. Maharashtra lies between  $19^{\circ} 36' 4.2984''$  N scope and  $75^{\circ} 33' 10.7244''$  E longitude. The state is bifurcated

into 35 area and four meteorological locales. These sub divisional bifurcations are (1) Vidarbha (2) Madhya Maharashtra (3) Marathwada and (4) Konkan. The Konkan sub-division lies on the windward side of the Ghats and the other sub-divisions lie on the leeward side. Within a piece of the state is semi-very dry. Colossal assortments in precipitation in different areas of the state achieve a wide extent of climatic conditions. In light of geological Maharashtra gets most outrageous precipitation in July (33% of SW storm precipitation) followed by August (28% of SW rainstorm precipitation). 89% of yearly precipitation gets during southwest tempest precipitation (June–September). Most outrageous precipitation gets during the SW storm season over the districts in Konkan region (2361–3322 mm) while bits of Madhya Maharashtra and Marathwada get least precipitation (454–600 mm). During the entire year there is an immense development in Blustery days in Nandurbar, Jalgaon, Raigarh, Kolhapur and Bhandara regions. However there is a basic reducing in Turbulent days in Pune, Solapur, Kolhapur, Ahmednagar, Aurangabad, Jalna, Beed, Hingoli, Nanded, Yavatmal, Wardha districts. During the period June to September there is an enormous development in Profound precipitation days in Nandurbar, Jalgaon, Raigarh, Kolhapur and Bhandara area. While it is also observed that there is an enormous decrease in significant rainy days in Pune, Solapur, Kolhapur, Ahmednagar, regions. According to the meteorological and climatic variations in.

- (1) Pre-monsoon seasonal rainfall (March–May).
- (2) South-west monsoon season or monsoon season rainfall (June–September).
- (3) Post monsoon season rainfall (October–November).
- (4) Rainfall in winter season (December-February).

In our study we have considered the annual rainfall for the state of Maharashtra from 1950 to 2020, with the annual rainfall we have also taken into consideration the normal monsoon season period which consists of June–September (4 months) since this is the period when the state of Maharashtra receives the majority of rainfall. Along with the above-mentioned period, a time period of further 8 months comprising of 5 months of January–May and 3 months of October–December is considered as the non-monsoon rainfall period.

## 5 Data Collection

The primary data collected for this study consists of the annual rainfall data for India for a time period between 1950–2020 from IMD. Data-set in use has 36 sub divisions and 19 attributes (individual months, annual, combinations of 3 consecutive months according to the monsoon and non-monsoon period). All the attributes have the amount of rainfall in mm.

## 6 Analysis of Rainfall Pattern

Many key parts of earth and human existence are subject to precipitation straightforwardly or in a roundabout way. The progressions in worldwide environment have a solid relationship with the yearly precipitation. It has been seen that the precipitation design off of late has been exceptionally unpredictable which has upset the agrarian and water the board universally. It is vital to investigate the precipitation design to comprehend and moderate the conditions of the boundaries which are straightforwardly or by implication impacted by the yearly precipitation. In the introduced study a dataset of Yearly precipitation of India from 1950–2020 has been dissected and the investigation has been addressed in the structure Reference diagrams showing appropriation of measure of precipitation, Conveyance of measure of precipitation yearly, month to month, gatherings of months, Circulation of precipitation in developments, areas structure every month, gatherings of months. The visualization of rainfall helps to make observations which can be used for creating or choosing the right predictive model for further experimentation.

Observations made from the visualization:

1. Histograms in Fig. 1 show the circulation of precipitation over months. It is Seen that there is expansion in measure of precipitation over months July, August, September all through
2. The two charts in Fig. 2 shows that how much precipitation is sensibly great in the long stretches of Spring, April, may in eastern India.
3. Scatter graph in the Fig. 3 shows the conveyance of precipitation on yearly premise, high measure of precipitation is Seen in 1950s.
4. Heat-Map in Fig. 4 shows the co-relation (dependency) between the measures of precipitation over months.
5. From above representations obviously on the off chance that measure of precipitation is high in the long periods of July, august, September then how much precipitation will be high yearly.
6. It is additionally seen that assuming measure of precipitation in great in the long stretches of October, November, December then the precipitation will be great in the general year.

## 7 Methodology

In the presented study firstly the analysis of rainfall was carried out followed by the prediction of annual rainfall. Three models for prediction were used and tested over the annual rainfall dataset.

1. The input data set originally had monthly and annualized rainfall for each year from 1950–2020. Data integration was performed by adding the Monsoon and Non Monsoon rainfall parameters in form of the Monthly divisions namely Jan–Feb, Mar–May, Jun–Sep, Oct–Dec.

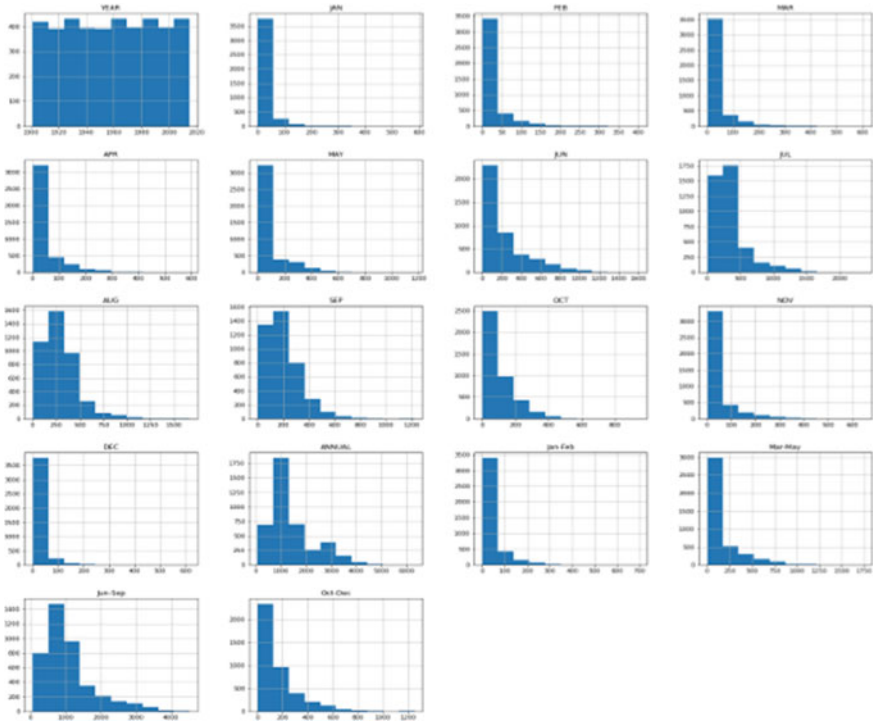


Fig. 1 Total rainfall distribution month wise from 1950–202

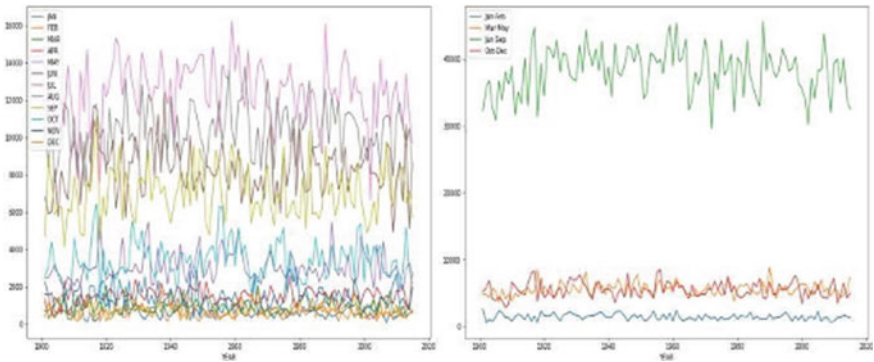


Fig. 2 Monthly rainfall in various sub-divisions of India

2. All the parameters were verified and the the missing data was augmented wherever necessary. The available data was then analyzed to create a visualization of the rainfall trends over the years. The dataset was then split into 80% train and 20% test data.

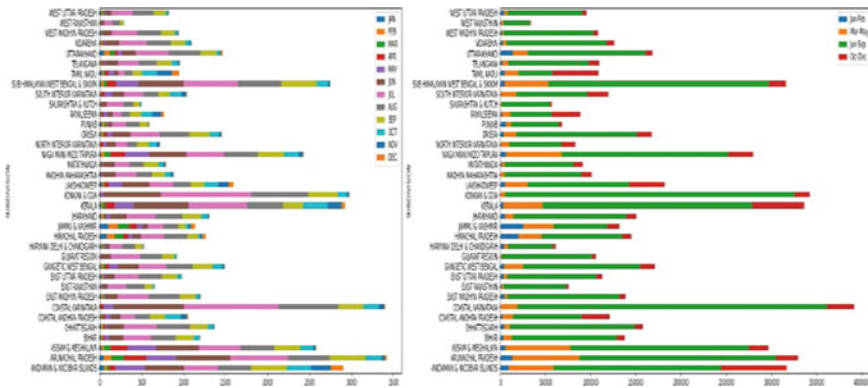


Fig. 3 Scatter plot of rainfall over the months

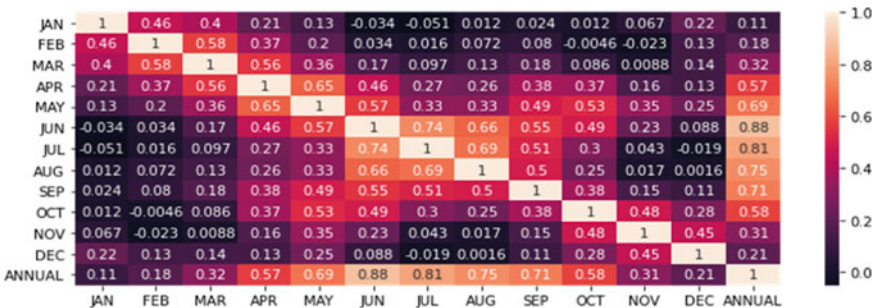


Fig. 4 Heat map for monthly and annual rainfall

3. The Linear regression model, SVR and ANN are used for making annual predictions on the dataset Two types of trainings were done once training on complete dataset and other with training on only Maharashtra and Konkan.
4. The ANN model used has been run with 10 epoch so as to yield better efficiency as it is known to have better accuracy for inconsistent data if executed with more number of epoch than usual.
5. All three models' parameters, MAE, MSE, RMSE and R2\_Score are compared to find the best rainfall prediction model for annual rainfall in Maharashtra and Konkan. The models were evaluated on the basis of above-mentioned metrics.

### 8 Results Analysis and Discussion

The study aims at long term prediction of rainfall in Maharashtra state, which is done with the help of regression modeling. It was found that due to the long-term prediction modelling the accuracies of the models leveraged showed distinct results.

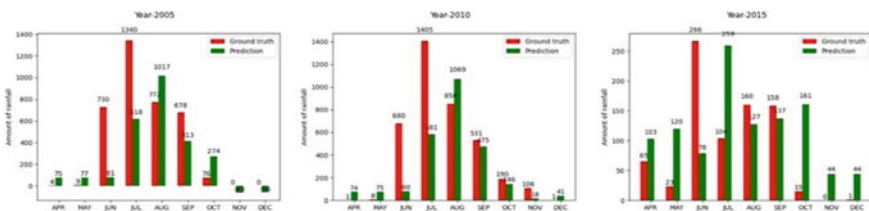


It has been observed that the long-term prediction usually is uncertain because of the data size and time accumulation and lack of information. In the long term time series prediction there are many limitations ranging from the dearth of data to less accurate results, but the deep learning techniques can be leveraged to introduce localization and improve the impact of the long term prediction. The deep learning models leveraged were able to learn directly the complex and arbitrary mapping in the input side and support and supplement the hypothesis on the output end (Table 2).

The tabular comparative result analysis mentioned in the following table shows the parametric performance of the three algorithms. The predictions made were on the basis of the annual rainfall, monsoon and non-monsoon sub-divisions The MAE values for the three models show the deviation of the predicted values from the actual value, which is higher for linear regression and SVR whereas ANN has showed comparatively less deviation. Similarly, the RMSE value tells us that the predictions made by LR and SVR are discrete and away from the best fit line. ANN has a better efficiency of prediction as the even the R2\_Score for this model is nearer to 1 which is considered ideal for any predictive model. The rainfall data is discrete and nonlinear which is why the two machine learning models have showed less efficiency in prediction, the efficiency for ANN is better but can't be said satisfactory (Figs. 5, 6 and 7).

**Table 2** Prediction performance

Sr. No.	Algorithm (model)	MAE	MSE	RMSE	R2_Score
1	LR	96.324	21,919.821	148.05	0.405
2	SVR	127.160	41,470.85	203.64	-0.1243
3	ANN	76.324	19,490.114	139.44	0.6726



**Fig. 5** Prediction of Year 2005, 2010, 2015 for Maharashtra and Konkan using Linear Regression

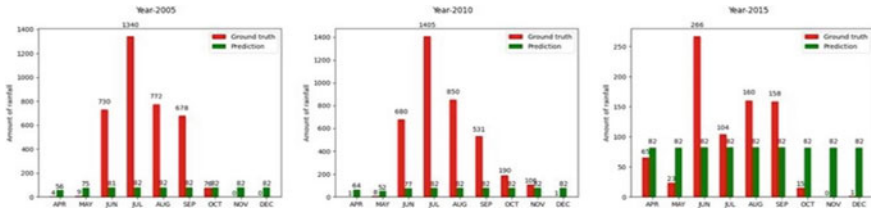


Fig. 6 Prediction of Year 2005, 2010, 2015 for Maharashtra and Konkan using SVR

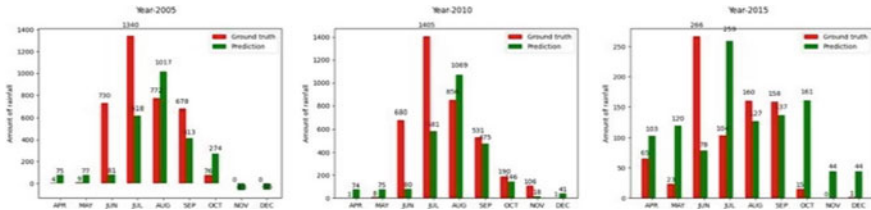


Fig. 7 Prediction of Year 2005, 2010, 2015 for Maharashtra and Konkan using ANN

## 9 Conclusion

Forecasting rainfall often can be helpful to make decisions regarding agro-based fields, having prior knowledge about the rainfall can help mitigate and foresee problems. In this paper a long term forecasting of rainfall in Maharashtra is done with the help of predictive modelling and deep neural networks, this study also provides a practical proof which supplements the fact that for annual rainfall in Maharashtra and Konkan is variable and unevenly scattered over the monsoon and non-monsoon period. The visualization of Indian rainfall made it clear that there is a lot of variance and discreteness in the rainfall pattern. To produce desired load forecasts, three forecasting techniques i.e. LR, SVR and ANNs were considered for evaluation by using multiple performance metrics. Significant weather profiles from eight different cities were selected to develop a synthetic weather station.. The results yielded by the two machine learning models (linear regression and SVR) depict the incompetency of the machine learning models for prediction of rainfall due to fluctuations in rainfall. It is also observed that the neural network model used for experimentation has performed better on the nonlinear time series data even though the number parameters in the dataset were limited on the basis of the evaluation metrics such as MAE, MSE, R2\_score. In future, multiple prospects of this research can be explored for further development. In future experimentation more powerful neural network model such as LSTM (Long Short-Term memory) and RNN can be implemented on the non- linear data in order to get more accurate results for prediction and analysis. Even the dataset can be made more modular by adding some more parameters to obtain better results.

## References

1. Zhang X, Mohanty SN, Parida AK, Pani SK, Dong B, Cheng X (2020) Annual and non-monsoon rainfall prediction modelling using SVR-MLP: an empirical study from Odisha. In: IEEE access, vol 8, pp 30223–30233. <https://doi.org/10.1109/ACCESS.2020.2972435>
2. Khan MI, Maity R (2020) Hybrid deep learning approach for multi- step-ahead daily rainfall prediction using GCM simulations. IEEE Access 8:52774–52784. <https://doi.org/10.1109/ACCESS.2020.2980977>
3. Sarma C (2022) Estimation and validation study of soil moisture using GPS-IR technique over a tropical region: variability of SM with rainfall and energy fluxes. IEEE J Sel Top Appl Earth Obs Remote Sens 15:42–49. <https://doi.org/10.1109/JSTARS.2021.3127469>
4. Bhomia S, Jaiswal N, Kishitawal CM, Kumar R (2016) Multimodel prediction of monsoon rain using dynamical model selection. IEEE Trans Geosci Remote Sens 54(5):2911–2917. <https://doi.org/10.1109/TGRS.2015.2507779>
5. Ratna SB (2012) Summer monsoon rainfall variability over Maharashtra, India. Pure Appl Geophys 169:259–273. <https://doi.org/10.1007/s00024-011-0276-4>
6. Gupta R et al (2021) WB-CPI: weather based crop prediction in India using big data analytics. In: IEEE access, vol 9, pp 137869–137885. <https://doi.org/10.1109/ACCESS.2021.3117247>
7. Tealab A et al (2017) Forecasting of nonlinear time series using ANN. Future Comput Inf J. <https://doi.org/10.1016/j.fcij.2017.05.001>
8. Sherpa SF, Shirzaei M, Ojha C, Werth S, Hostache R (2020) Probabilistic mapping of August 2018 flood of Kerala, India, using space-borne synthetic aperture radar. IEEE J Select Top Appl Earth Observ Remote Sens 13:896–913. <https://doi.org/10.1109/JSTARS.2020.2970337>
9. Maitra A, De A, Adhikari A (2019) Rain and rain-induced degradations of satellite links over a tropical location. IEEE Trans Antennas Propag 67(8):5507–5518. <https://doi.org/10.1109/TAP.2019.2911376>
10. Gopalakrishnan D, Chandrasekar A (2018) On the improved predictive skill of WRF model with regional 4DVar initialization: a study with north Indian ocean tropical cyclones. IEEE Trans Geosci Remote Sens 56(6):3350–3357. <https://doi.org/10.1109/TGRS.2018.2798623>
11. Samsudin R, Shabri A, Saad P (2013) A comparison of time series forecasting using support vector machine and artificial neural network model. J Appl Sci 10(11):950–958; Trivedi SK, Dey S (2013) Effect of various kernel and feature selection methods on SVM performance for detecting email spams. Int J Comput Appl 66(21):18–23
12. Üstün B, Melssen W, Buydens L (2006) Facilitating the application of support vector regression by using a universal Pearson VII function based kernel. Chemometr Intell Lab Syst 81(1):29–40; Paras SM, Kumar A, Chandra M (2009) A feature based neural network model for weather forecasting. Int J Comput Intell 4(3):209–216
13. Mishra AK (2013) A new technique to estimate precipitation at fine scale using multifrequency satellite observations over indian land and oceanic regions. IEEE Trans Geosci Remote Sens 51(7):4349–4358. <https://doi.org/10.1109/TGRS.2012.2226733>
14. Kumarasiri AD, Sonnadara DUJ (2006) Rainfall forecasting: an artificial neural network approach. Proc Tech Sess 1–16
15. Chattopadhyay S, Chattopadhyay M (2007) A soft computing technique in rainfall forecasting. [arXiv:nlin/0703042](https://arxiv.org/abs/nlin/0703042) [online]. <https://arxiv.org/abs/nlin/0703042>; Deshpande RR (2012) On the rainfall time series prediction using multilayer perceptron artificial neural network. Int J Emerg Technol Adv Eng 2(1):2250–2459
16. Ko M, Tiwari A, Mehnen J (2010) A review of soft computing applications in supply chain management. Appl Soft Comput 10:661–674
17. Abraham ER, Mendes dos Reis JG, Vendrametto O, de Oliveira Costa Neto PL, Carlo Tolo R, de Souza AE, de Oliveira Morais M (2020) Time series prediction with artificial neural networks: an analysis using brazilian soybean production. Agriculture 10(10):475
18. Salvi L, Patil H Data mining techniques and IoT used for building a futuristic water conservation system. IOSR J Eng (IOSRJEN). ISSN (e) 2250–3021. ISSN (p): 2278–8719

19. Galit S, Koppius OR (2011) Predictive analytics in information systems research. *Mis Quar* 35(3):553–572
20. Seber G, Lee AJ (2012) Linear regression analysis. Wiley Series in Probability and Statistics
21. Montgomery DC, Peck EA, Vining GG (2015) Introduction to linear regression analysis. Wiley Series in Probability and Statistics
22. Kavitha S, Varuna S, Ramya R (2016) A comparative analysis on linear regression and support vector regression. In: 2016 online international conference on green engineering and technologies (IC-GET), pp 1–5. <https://doi.org/10.1109/GET.2016.7916627>
23. Wang Y, Du T, Liu T, Zhang L (2019) Dynamic multi-objective squirrel search algorithm based on decomposition with evolutionary direction prediction and bidirectional memory populations. *IEEE Access* 7:115997–116013. <https://doi.org/10.1109/ACCESS.2019.2932883>
24. Zhang X (2017) Prediction modeling of the population of preschool children based on logistic model. In: 2017 9th international conference on measuring technology and mechatronics automation (ICMTMA), pp 300–305. <https://doi.org/10.1109/ICMTMA.2017.0079>
25. Ma W (2020) Prediction and analysis of population aging based on computer Leslie model. In: 2020 management science informatization and economic innovation development conference (MSIEID), pp 102–105. <https://doi.org/10.1109/MSIEID52046.2020.00027>
26. Wang J, Ma X, Song R, Yan C, Luo H (2021) A novel disease prediction method based on inductive representation learning. In: 2021 IEEE 6th international conference on big data analytics (ICBDA), pp 239–243. <https://doi.org/10.1109/ICBDA51983.2021.9403168>
27. Liu J, Zhang Y, Xiao Z, Qiao T, Tan H (2015) Fast, finite, accurate and optimal WASD neuronet versus slow, infinite, inaccurate and rough BP neuronet illustrated via russia population prediction. In: Sixth international conference on intelligent control and information processing (ICICIP), pp 140–145. <https://doi.org/10.1109/ICICIP.2015.7388158>
28. Firmanuddin G, Supangkat SH (2016) City analytic development for modeling population using data analysis prediction. In: 2016 international conference on ICT for smart society (ICISS)
29. Chakraborty S, Verma P, Paudel B, Shukla A, Das S (2022) Validation of synthetic storm technique for rain attenuation prediction over high-rainfall tropical region. In: *IEEE geoscience and remote sensing letters*, vol 19, pp 1–4. Art no 1002104. <https://doi.org/10.1109/LGRS.2021.3068334>