JOURNAL OF NORTHEASTERN UNIVERSITY

Volume 25 Issue 04, 2022 ISSN: 1005-3026 https://dbdxxb.cn/ Original Research Paper

FORECASTING MODEL FOR RAINFALL USING SUPERVISED LEARNING

T. Bhaskar
Asst. Prof., Dept. Of CSE, CMR College of Engineering & Technology, Hyderabad, Telangana.

MN Narsaiah
Assoc. Prof. Dept. Of ECE, KG Reddy College of Engineering & Technology, Hyderabad, Telangana

ABSTRACT: Agriculture is the key point for survival for developing nations like India. For farming, precipitation is generally significant. Precipitation updates are helpful for evaluating water assets, farming, ecosystems and hydrology. Nowadays precipitation anticipation has become a foremost issue. Forecast of precipitation offers attention to individuals and know in advance about precipitation to avoid potential risk to shield their crop yields from severe precipitation. This study intends to investigate the dependability of integrating a data pre-processing technique called singular-spectrum-analysis (SSA) with supervised learning models called least-squares support vector regression (LS-SVR), and Random-Forest (RF), for precipitation prediction. Integrating SSA with LS-SVR and RF, the combined framework is designed and contrasted with the customary approaches (LS-SVR and RF). The presented frameworks were trained and tested utilizing monthly climate dataset which is separated into 80:20 ratios for training and testing respectively. Performance of the model was assessed using RMSE and NSE. Experimental outcomes illustrate that the proposed model can productively predict the rainfall.

KEYWORDS: Singular-Spectrum-Analysis, Supervised Learning, Rainfall, SVR, RF, Prediction

INTRODUCTION

India's welfare is farming, where most of agribusiness is subject to precipitation as its standard wellspring of water, the time and proportion of precipitation hold high significance and can affect the whole economy of the country. Weather plays a major part in our regular day to day life. Climate estimating is one of the most testing issues seen by the world, in a latest couple of century in the field of science and innovation. As India's economy notably relies upon cultivation, precipitation has a significant impact [1]. Variation in the timing of precipitation and its quantity makes estimating of precipitation is a challenge for meteorological researchers. Predicting is one the greatest difficulties for researchers from an assortment of fields, for example, climate data mining, ecological machine learning, functional-hydrology, and numerical prediction, to make a forecast model for precise precipitation. Climate anticipation stands apart for all nations around the world in all the advantages and administrations gave by the meteorological department [2].

Precipitation anticipation is significant because severe and sporadic precipitation can have numerous effects like obliteration of yields and farms, harm of property so a superior anticipating model is crucial for an early notice that can limit dangers to life and property and furthermore

.

Submitted: 21/10/2022 **Accepted**: 24/11/2022

dealing with the farming in better way. This forecasting predominantly helps ranchers and furthermore water assets can be used proficiently [3]. Precipitation forecast is a difficult task and the outcomes ought to be precise. There are numerous equipment tools for anticipating precipitation by utilizing the climate conditions like temperature, humidity, pressure and so on. These customary strategies can't work in a proficient manner so by utilizing ML based strategies we can design for exact outcomes. We can only do it by having the past data examination of precipitation and can foresee the precipitation for future seasons [4]. Taking into account that farming activities and crop yield depend on the precipitation distribution, monthly precipitation estimating is significant for farming planning and flood control. Monthly precipitation anticipation with a suitable technique is a fundamental prerequisite to help water management. Accordingly, monthly precipitation estimating is broadly appropriate in the field of hydrology. A few procedures for predicting time series have been designed on a worldwide scale [5].

Machine Learning (ML) or Artificial Intelligence (AI) and stochastic techniques dependent on information extraction procedures are the mainly utilized time-series modeling for hydrological estimating. Nonetheless, ML has been given more consideration in weather anticipating, predominantly in light of the fact that stochastic approaches consider that the time-series is fixed and have a restricted capacity to catch profoundly nonlinear qualities of precipitation series. Climatology time-series in tangible utilize are normally non-fixed and non-linear, so foreseeing greatest values is very complex [6]. The hypothesis of the modular approach and the incorporation of various models have presently increased more enthusiasm for precipitation estimating to address this issue. Regression, Artificial Neural Network (ANN), Decision Tree, Random Forest, Fuzzy logic and group cycle of data handling methods are the majority utilized computational strategies utilized for climate forecasting. Even though ANN or SVM in such non-linear issues are generally applied, direct elucidation of the guidelines are complex. Then again, classification utilizing decision tree methods is helpful in displaying of complex associations among attributes with the additional focal points of recognizing the significance of every attribute [7]. RF & SVR are computationally quick, effectively reasonable and don't need an earlier knowledge on the data. Because of these reasons, the utilization of RF in the case of prediction is picking up popularity. Further, because of its ability to recognize the persuasive part of various features, it's rather beneficial to utilize RF in prediction applications. Techniques dependent on data driven, ML frameworks are broadly and effectively practical in numerous fields. ML has turn into a general inductive practice in precipitation estimation outstanding to its incredibly non-linear, adaptable, and data-driven method training without first receiving catchment and flow methods. Notwithstanding the fame of ML techniques for time-series anticipation, they are not an effective device to foresee long-standing precipitation [8]. The most commonly utilized ML-based approaches for precipitation prediction include SVM, Genetic-Programming, ANN, and Fuzzy-Logic. Cross-breed methodologies for tending to various climatology issues have been supported by meteorologists. Least-square SVR has newly got extensive consideration in different forecasting issues. In this research, LS-SVR was picked on the grounds that it is computationally more engaging than the conventional SVR. In light of the utilization of quadratic programming by non-linear conditions, the conventional SVR has computational challenges to choose the best possible solution. The hybrid reproduction approach dependent on LS-SVR as an estimating model conveys incredible results.

Copyright © 2022. Journal of Northeastern University. Licensed under the Creative Commons Attribution Non-commercial No Derivatives (by-nc-nd). Available at https://dbdxxb.cn/

The Singular Spectrum Analysis (SSA) is a proficient device that can divide the original time-series into an amount of discrete segments, together with pattern designs, oscillating segments, and noise. SSA is applied to yearly, monthly, and hourly water temperature time-series to assess its ability and prediction capability to recognize major data from those series. The study infers that the SSA can acquire and give magnificent forecasts to noteworthy hydrological time series segments with unique uneven practices, for example, precipitation and runoff series. SSA is used to extract the trend and it is a striking trend-extraction procedure, since it needs no approach portrayal of time-series and pattern, mines noisy time-series patterns through unsure motions in time-series, and susceptible to outliers. The noteworthiness of utilizing an appropriate SSA to translate unrefined input data to gracefully superior quality data earlier than being actualized as a design input [9]. Existing methods have been led to investigate the benefit of united data pre-processing and ML; evidently, this is the unique that SSA combined among LSSVR and RF has been utilized for precipitation estimation.

There is persuading proof that the hybrid design dependent on LS-SVR& RF is solid and that the SSA generates excellent results. Subsequently, for enhancing the prediction efficiency of the sate-of-art models, the data pre-processing procedure is implemented in the present investigation.

The main objectives of the study are:

1. Connecting SSA with ML-based methods (i.e., LS-SVR& RF) to build crossbreed models (SSA-LSSVR &SSA-RF) for precipitation prediction.
2. Comparison among the crossbreed models and the conventional models to assess the effectiveness of the data pre-processing strategy.

We have seen that the majority of the works claiming higher accuracy have labeled precipitation into three or under three parts or have predicted precipitation utilizing ML methods but have not done precipitation anticipating utilizing ML strategies, very few of them have utilized barely any meteorological parameters for the anticipation of the precipitation [10]. Most of the existing works are utilized the regression methods for forecasting rainfall and outcomes are not up to the mark due selection of correct parameters for modeling. We have proposed a model to foresee the precipitation utilizing a combination ML based techniques. Expectation of precipitation relies upon different climate attributes. Categorizing the precipitation gives us the great classification precision but our definitive objective is to anticipate the precipitation utilizing the other climate attributes [11]. In this investigation objective isn't just to accurately classify precipitation but additionally effectively foresee the precipitation utilizing different climate attributes. Our work is concentrated around understanding the impacts of various meteorological attributes in precipitation estimation alongside an investigation of approaches which were utilized for anticipating precipitation, ML, and their restriction. The proposed model predicts the precipitation for the following season utilizing ML and forecasting approaches. Our contribution to this problem is to predict the monthly rainfall using ML-based techniques and compare the performance of the model with state of the art approaches.

The remaining of the article is organized as follows. Study location and Dataset description is presented in Section 2, Related works are discussed in Section 3, The proposed methodology is discussed in Section 4, Results and Discussions of the proposed model is presented in Section 5 and, finally Conclusion and future work is stated in Section 6.

Copyright © 2022. Journal of Northeastern University. Licensed under the Creative Commons Attribution Non-commercial No Derivatives (by-nc-nd). Available at https://dbdxxb.cn/

LITERATURE REVIEW

In this section, the existing works on rainfall forecasting proposed by various researchers are presented. J Abbot et al. [12] designed a rainfall forecasting model by artificial neural networks which produce more precise results compare to conventional statistical and numerical techniques. F Fahimi et al. [13] developed a hybrid framework called an Adaptive-Neuro-Fuzzy-Inference-System (ANFIS) for accurate long-term precipitation prediction. The outcome shows that the ANFIS is fit for catching the precipitation data-dynamic behavior and produces agreeable results. O Kisi et al. [14] presented a framework, which combines ANNs with wavelet examination (WA) to estimate rainfalls and the model performance was compared with standard ANFIS. The outcomes exhibit that the proposed model is more productive than the ANFIS and is reasonable for precipitation anticipating. SM Pandhiani et al. [15] described a rainfall anticipation model by SSA and SVR for seasonal precipitation anticipation. The outcomes illustrate a noteworthy growth in model effectiveness contrasted to the standard SVR approach. JC Chan et al. [16] presented a monthly rainfall forecasting model by LS-SVR. The investigation demonstrates that the SVR design is better than the ARIMA approach. The investigation infers that the clarification for SVR acceptable execution lies in the non-direct quality of the caught and utilized SVR space. L Karthikeyanet al. [17] described a comparative study of ML-based techniques on rainfall forecasting. They concluded has ANN, SVR, and RF exhibitions all in all good and RF conveyed the finest output. Finally, there is no investigation of RF-dependent models with a data pre-processing strategy for precipitation prediction, which timely this recent research to present a crossbreed model pairing RF with a data pre-processing approach. S.Y. Ji et al. [18] designed a rainfall prediction model by decision tree with CART and C4.5 techniques. The proposed strategy predicts rainfall and it is characterized into three classes in hourly rainfall 0.0 to 0.5 mm as level 1, 0.5 to 2.0 mm as level 2, > 2.0mm as level 3.

Min Min et al. [19] explored and designed an algorithm called quantitative precipitation estimates (QPEs) based on random forest (RF), machine learning (ML) techniques for summer-time rainfall forecasting utilizing Himawari study dataset, cloud substantial properties products, and GFS-NWP data. In this study, they used a hybrid forecasting model that incorporates regression techniques with RFs classification for rainfall forecasting. The proposed method works tremendously and is different from the traditional and existing models because it utilizes the RFs ML approach for nowcasting. MAI Navid et al. [20]proposed amultiple linear regressions (MLR) technique for forecasting the rainfall in Bangladesh. In MLR, first, apply correlation investigation and then regression analysis. MLR is very useful in future rain prediction and the results are very helpful to the agriculture sector for crop management. Finally, they concluded has rainfall is influenced by many climate factors and utilized those factors in the forecasting of rainfall to increase accuracy. Jesleena Rodrigues et al. [21] designed the Multiple Linear Regression (MLR) technique and time-series ARIMA techniques for monthly precipitation anticipation. The experimental results demonstrate that, MLR & ARIMA produce accurate forecast results over the traditional forecasting methods. In the end, the proposed performance of the two models are evaluated in terms of MAPE and the outcomes and the exactness of predictions made for precipitation by MLR (1.14) is found to be greater over ARIMA (19.61). S. Swain et al. [22] developed a multiple linear regression model (MLRM) to anticipate the yearly rainfall over the Cuttack region, Odisha, India, utilizing the annual average rainfall data of the previous three years. The proposed model results describe that it is capable to generate precise accuracy matching with the actual data values and the proposed model acquired high R2 (0.974) and adjusted R2 (0.963) when compared to the

Copyright © 2022. Journal of Northeastern University. Licensed under the Creative Commons Attribution Non-commercial No Derivatives (by-nc-nd). Available at https://dbdxxb.cn/

existing models.Razeef Mohd et al. [23]presented a rainfall forecast model utilizing Nonlinear Auto- Regressive with External Input (NARX), which was trained by the proposed Self Adaptive Levenberg-Marquardt (Self Adaptive LM) framework and to make it more adaptive, the LM approach was customized with the learning rate to make it more precise for predicting rainfall. The precipitation data was gathered from Kashmir and Jammu, India. The experiments were conducted and model performance was evaluated with RMSE (1.721 %) and MSE (1.721 %) values.

Faulina et al. [24] designed a hybrid monthly rainfall prediction framework by using ARIMA and ANFIS at particular locations in Indonesia, namely Pujon and Wagir. The proposed model was executed and the performance of the model was compared in terms of accuracy against the existing works. The experimental outcomes display that the ANFIS model is more precise in forecasting the monthly rain-data of Pujon, whereas the ARIMA technique outcomes were superior in forecasting the monthly rain-data of Wagir. Sam Cramer et al. [25] describe an intelligent machine learning system for rainfall forecasting where the forecasts of a few target parameters are decisive to a particular purpose. The proposed model was applied and compared with the forecast performance of the base-lines and six other well-liked ML prediction models called M5 Model trees, K-Nearest Neighbours (KNN), Radial Basis Neural Networks (RBNN), SVR, Genetic Programming, and M5 Rules. The investigation results state that the ML prediction models are capable to do better than the conventional models. C. R. Rivero et al. [26] implemented a short –term rain forecast model using Bayesian Enhanced Modified Approach (BEMA) with relative entropy. The experimental outcomes state the accuracy of the proposed methodology through diverse forecasting models utilizing the SMAPE index for short term precipitation sequence and chosen sequence from standards.

STUDY AREA AND DATA COLLECTION

Weather data is in extremely enormous amount and it includes of forty-seven (47) general parameters such as rainfall, temperature, humidity, sun-shine and cloudy hours, wind speed, etc. The major problem of this section is to investigate and process theclimate data of the study region, Nalgondadistrict, TS, India. The attributes have to be selected in such a way that it includes the effect of seasonal rainfall. We have to select different input and output variables and find a correlation between those variables using various strategies such as probability distribution method, Karl Pearson's coefficient, etc. and preprocess the data using appropriate methods.

3. 1 Case Study Region

Nalgonda district is one of the 23 districts of Andhra Pradesh, with a totalgeographical area of 14240 Sq.km. It has a total population of 34,83,648 as per 2011census. The district has 1178 Gram Panchayats, 1161 revenue villages and 59Mandals. For Administrative convenience the district is divided into 4 revenuedivisions located at Bhongir, Nalgonda, Miryalguda and Suryapet.The district lies between North latitudes 16° 25' and 17° 50' and between eastlongitudes 78° 40' and 80° 05' forms a part of major basin of Krishna river and iscovered by Survey of India toposheet Nos. 56K, 56L, 56O and 56P. As for the agriculture is concerned,the main source of irrigation is groundwater being 72.56% of total gross areairrigated, whereas surface water irrigation accounts for 27.33% of gross area.jowar,bajra, grams, are mostly rainfed crops. The average annual rainfall of the Nalgondais around 751 mm [27].

The daily precipitation information, estimated in millimeter (mm), was acquired from the IMD, Hyderabad, TS, and India. Severe rainfall is seen in many regions during June to September

Copyright © 2022. Journal of Northeastern University. Licensed under the Creative Commons Attribution Non-commercial No Derivatives (by-nc-nd). Available at https://dbdxxb.cn/

period because of the Southwest Monsoon (SWM). During the SWN monthly rainfall at certain regions goes up to 700 mm however during the lean time it stays under 50 mm. The Nalgondadistrict encounters substantial precipitation and floods each year which causes river-bank erosions and landslides in several parts of the district. Farming plays a significant job in the economy of the State. Farming which is the principal livelihood of the individuals who establish almost 90% of the all-out populace is additionally influenced by the rainfalls and floods [28].

3.2 Dataset Description

To predict seasonal monthly rainfall, we have to portray various elements which indirectly or directly influence the rainfall. The input variables of rainfall estimation are Maximum, Minimum and Average Temperature (oC), Vapor Pressure (hPa), Wind Speed (Km/h), Humidity (%) and Cloud Cover (%). The dataset details are shown in Table 3.1.

Table 3.1: List of Input and Output parameters

SNO	Variable	Name of the attribute	Short name	Type
1	X_1	Minimum Temperature (°C)	MinTemp	Input
2	X_2	Maximum Temperature (°C)	MaxTemp	Input
3	X_3	Average Temperature (°C)	AvgTemp	Input
4	X_4	Vapor Pressure (hpa)	VP	Input
5	X_5	Wind Speed (Km/h)	WS	Input
6	X_6	Relative Humidity (%)	RH	Input
7	X_7	Cloud Cover (%)	CC	Input
8	Y	Rainfall (mm)	RF	Output

3.3 Primary source of data

In this research, the 110 years of climate data (1901-2012) of Nalgondaregion, AP, India is taken from the Indian Meteorological Department (IMD), Hyderabad, TS, India at http://www.imdhyderabad.gov.in, and National Climatic Data Center, Asheville, USA at ncdc.noaa.gov in the form of monthly means. Parameters like MinTemp, MaxTemp, AvgTemp, VP, WS, RH and CC are considered as input values and rainfall as output parameter which relies on input attributes. We considered just the relevant, required elements for design the forecasting model. Table 3.1 portrays the climate data depiction utilized in the rainfall prediction model. In our proposed model, we considered extended Southwest monsoon season sets in by June and lasts till October while the Northeast monsoon begins in October and ends in December. The evocative statistics were utilized for investigating their precipitation properties.

PROPOSED METHODOLOGY

The problem examined in this section is that of anticipation of precise monthly precipitation for the accompanying season dependent on the past climate dataset, by using relating prior data. The combined SSA with LSSVM and RF framework is applied over this data so as to build up a model to anticipate the monthly precipitation value to the Nalgondaregion which is situated on the southeast coast of India, with the water of the Bay of Bengal.

4.1 Least-Squares Support Vector Machine

In this segment, we quickly discuss the basic hypothesis on LS-SVR in time-series anticipation. In view of a training set as input xi and output yi, the regression equation that interfaces the input vector to the output can be define as:

The LS-SVR, a novel kind of SVR, involves a series of same supervised strategies that investigate data and recognize trends. It works well in solving convex quadratic issues with more intensity and produces good results for linear equations. Further, it has the advantages of shortening the issue and solution finding without losing exactness. In this investigation, the

Copyright © 2022. Journal of Northeastern University. Licensed under the Creative Commons Attribution Non-commercial No Derivatives (by-nc-nd). Available at https://dbdxxb.cn/

LS-SVR is utilized to anticipate monthly precipitation. In the accompanying sector, we discuss the essential hypothesis on LS-SVR in time-series anticipating. In view of a trainingset as input xi and output yi, the regression equation defined as follows in Equation (1).

$F(y) = w^T \phi(x) + b$ Eq. (1)

Where, $\phi(x)$ is a nonlinear mapping function. Converting the regression issue in Eq. (1) into a constrained-quadratic optimization issue, by diminishing the cost element, w and b can be determined. In the structural minimization rule, the regression issue can be originated in Equation (2) & (3):

$\text{Min } J(w, e) = \frac{1}{2} w^T w + \frac{\acute{r}}{2} \sum_{i=1}^{m} e_i^2$ Eq. (2)

Subject to the following limitations:

$y_i = w^T \phi(x_i) + b + e_i \ (i = 1, 2, : : : , m)$ Eq. (3)

Where ŕ represents the consequence term and ei is the training-error for xi. To crack the optimization problem, the answer for optimizing the LS-SVM is to generate a Lagrangian equation expressed in Equation (4):

$L(w, b, e, \alpha) = J(w, e) - \sum_{i=1}^{m} \alpha_i \{ w^T \phi(x_i) + b + e_i - y_i \}$ Eq. (4)

Where α_irepresents the Lagrange values.

By constructing the kernel function K(x, xi); Mercer's theorem can be fulfilled. Next, the LS-SVR method is executed in Equation (5):

$f(x)=\sum_{i=1}^{m} \alpha_i K(x, xi) b$Eq. (5)

Random Forest

Random-Forest is most remarkable ML technique for analytics, including a collection of basic trees. RF is an improved version of decision-tree dependent on the strategy for bagging. The more significant the impact on the prediction" so as to utilize the RF model for the selection of factors. In bootstrap-sampling, the annotations are isolated into two parts for every each approach: in-bag-subset and out-of-bag subset. The out-of-bag subsets might be utilized to decide the noteworthiness foe every testing attribute: (1) randomizing the qualities for one picked logical attribute in the out-of-sack subset; (2) utilizing the randomized out-of-sack-subset and the first example to make new expectations; (3) testing the importance of the picked illustrative attribute by expanding the MSE of the new conjecture.

Singular Spectrum Analysis

SSA illustrates a powerful way to deal with time-series investigation in numerous fields of research. It is especially significant when time-series are disintegrated into significant parts like trends, motions, and noise. A significant advantage of the SSA method is that it is non-parametric, which means it tends to be custom fitted to the basic informational index and refuse the requirement for an earlier method. Consequently, the SSA is viewed as a model-free technique. As indicated by, two valuable stages are engaged with the SSA strategy: decomposition and reconstruction. The method of decomposition consists of two stages: combining and singular-value-decomposition (SVD). This disintegration is the fundamental outcome of the SSA technique and it is significant when every restored sub-section can be sorted as either a sample design, or as a part of noise. Embedding is the first stage in the SSA technique. This strategy changes the fitted time-series to a multi-dimensional vector series. Singular Value Decomposition (SVD) is the trajectory-matrix is the principal module of the decomposition procedure.

Copyright © 2022. Journal of Northeastern University. Licensed under the Creative Commons Attribution Non-commercial No Derivatives (by-nc-nd). Available at https://dbdxxb.cn/

Linking SSA with LS-SVR and RF

The monthly precipitation values were normalized by their particular means and standard-deviations earlier to training the standard approaches (LS-SVR and RF). To train these approaches, the normalized precipitation values are utilized. Two elements, the penalty-term and the kernel-width, should be chosen in the calibration procedure of the LS-SVR. The matrix-search strategy is used for improving elements for the period of the calibrating time of the LS-SVR. It is fit for delivering optimum element set and can conquer the issues of over-fitting of the appraoch by means of the cross-validation system. RF has two elements, the quantity of factors √Mand the quantity of trees (ntree), which should be estimated. √M would for the most part produce close ideal outcomes, so the estimation of √M was chosen by experimentation utilizing the incentive around √M (estimation of M is 4). The scope of ntree from 0 to 3000 was utilized to look through the best worth. Be that as it may, no significant change was accomplished contrasted with the default an incentive for ntree of 750. Subsequently, the estimation of 750 for ntree was embraced in this research.

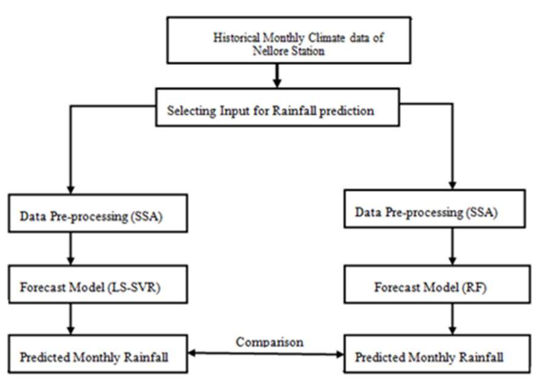
The monthly rainfall of Nalgonda District was taken from 1925–2000 and the initial 60-years of precipitation data were applied for training and the left over 15 years of precipitation data were utilized for validation. The areal precipitation prediction for Nalgonda District was executed utilizing the LS-SVR and RF.

The set of suitable inputs is a significant worry for LS-SVR &RF modelling. Diverse mix of forerunner values of the precipitation data were considered as inputs (i.e., (1) P(t); (2) P(t), P(t-1); (3) P(t), P(t-1), P(t-2)). The output is precipitation time-series data to be predicted with 1-, 2-, and 3-month lead-time (i.e., P(t+1), P(t+2), and P(t+3)).

The crossbreed models were acquired by consolidating two distinct strategies. Taking into account the SSA's strength, these approaches are intended to progress determining execution and dependability. The outcomes of the normal and crossbreed models are contrasted to evaluate the accuracy of the model in rainfall anticipation.

The proposed framework has following steps and depicted in Figure 2.

1. At first, the time-series of precipitation data was rotten into various principle components (PCs) utilizing SSA.
2. The appropriate PCs are determined based on the pattern or time of every series and new series of each parameter is established by including the essential parts to be characterized.
3. LS-SVR &RF approaches are designed to each part of the reconstruction so the design of LS-SVR &RF is distinctive for every segment of the restoration.
4. At last, LS-SVR &RF approaches are fed with the new series to anticipate the future rainfalls for 1-, 2-, and 3, 4-month lead-time. This is the main thought of pairing SSA with ML strategies.



Copyright © 2022. Journal of Northeastern University. Licensed under the Creative Commons Attribution Non-commercial No Derivatives (by-nc-nd). Available at https://dbdxxb.cn/

Fig. 2. Framework of the Proposed Rainfall Prediction Model

Forecast Verification

The proposed model performance is assessed in terms of RMSE and NSE, for the calibration and testing duration. The formulas are represented in Equations (6) & (7) respectively.

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(X-Y)^2} \quad \text{Eq. (6)}$$

Where, X and Yare the fitted and predicted values correspondingly. Lesser values of RMSE recommend higher precision.

$$\text{NSE} = 1- -\frac{\sum_{i=1}^{n}(X-Y)^2}{\sum_{i=1}^{n}(X-M)^2} \quad \text{Eq. (7)}$$

Where, M is mean of the fitted values. An NSE of 0.75–1.0 relates to an "excellent" accuracy, 0.65–0.75 to a "good" precision, and 0.5–0.65 to a "sensible precision", while values below 0.5 imitate unacceptable precision. In spirit, the nearer NSE is to 1, the more precise is predicted.

RESULTS AND DISCUSSION

In this article we designed the monthly precipitation prediction approaches for 1, 2, 3, 4-month lead time for Nalgonda District. The performance of the standard and hybrid models is compared. The variation is that the customary approaches utilized the direct noisy data as model input, while the crossbreed models utilized the deteriorated input data produced by SSA rather than unrefined data. Table 2 displays the prediction performances for 1, 2, and 3, 4-month intervals for the customary and crossbreed models in terms of RMSE and NSE.

Table 2. Prediction performances of the proposed models

Model	**RMSE**				**NSE**			
	M1	M2	M3	M4	M1	M2	M3	M4
LSSVR	94.58	93.31	99.56	99.85	0.05	0.07	0.06	0.08
SSA-LSSVR	**82.29**	**89.48**	**81.09**	**83.54**	**0.98**	**0.96**	**0.88**	**0.86**
RF	97.87	99.24	94.18	100.54	0.05	0.06	0.05	0.03
SSA-RF	**91.76**	**96.77**	**92.83**	**92.56**	**0.73**	**0.59**	**0.56**	**0.75**

In Table 3, it is seen that RMSE &NSE display extremely deprived values for the customary methods utilizing raw data when contrasted with the crossbreed models utilizing data created by SSA. There is additionally proof that the customary approaches have been inadequately approved for each interval. The LS-SVR outcome is noise-sensitive and might not be successful when the degree of clamor is lofty. Thusly, the LS-SVR pairing with the SSA separating the crude precipitation data will diminish the noise impacts and its helpful in enhance the model performance. Figures 3a, 3b and 4a, 4b outline the time-series charts of the fitted and one-month lead-time anticipated precipitation by the customary and hybrid approaches of Nalgonda. From the figures, we uncover that the predicted values from the

Copyright © 2022. Journal of Northeastern University. Licensed under the Creative Commons Attribution Non-commercial No Derivatives (by-nc-nd). Available at https://dbdxxb.cn/

crossbreed models are nearer to fitted values than the values estimated by the standard models. The rainfalls anticipated by the crossbreed models were discovered to be firmly restricted to the line of uniformity, though the precipitation determined by the customary models is not near the line of balance.

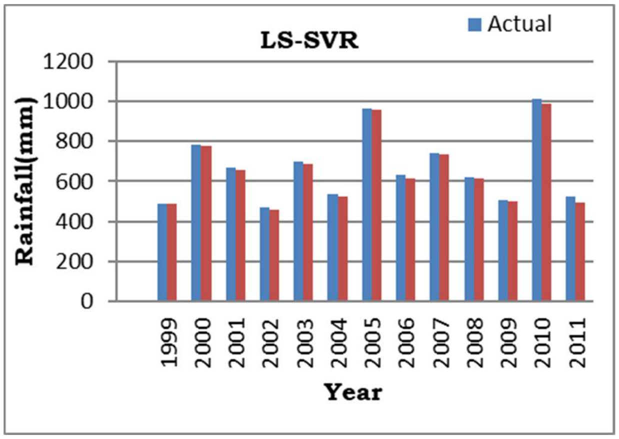


Fig. 3a. Fitted and Predicted values for LSSVR Model

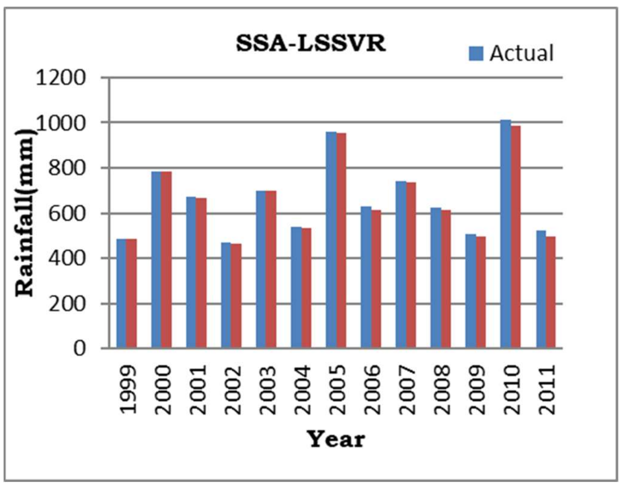


Fig. 3b. Fitted and Predicted values for SSA-LSSVR Model

Copyright © 2022. Journal of Northeastern University. Licensed under the Creative Commons Attribution Non-commercial No Derivatives (by-nc-nd). Available at https://dbdxxb.cn/

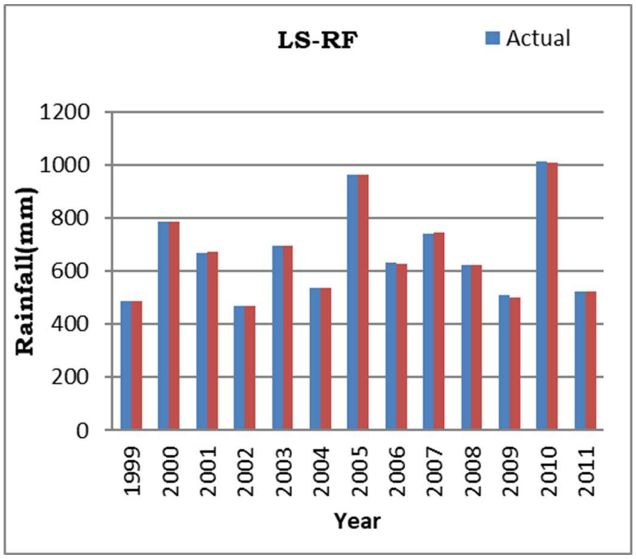


Fig. 4a. Fitted and Predicted values for LSRF Model

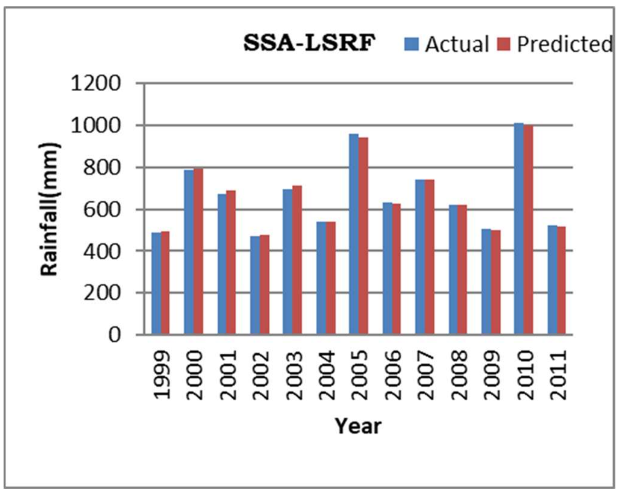


Fig. 4b. Fitted & Predicted values for SSA-LSRF Model

Comparative Analysis

The proposed hybrid models are predicted the monthly rainfall reasonably and produced significant results. The proposed model accuracy is assessed by RMSE and NSE and the results display that the SSA-LSSVR and SSA-LSRF are good compare to LSSVR and LSRF respectively. In this section the proposed model performance is compared with existing models in terms of RMSE, NSE. The results show that the proposed hybrid framework looking good compared to existing techniques in monthly rainfall prediction of NalgondaDistrict. Here we had taken average value of 1-, 2-, 3-, 4- month percentages RMSE, and NSE values for comparison of the proposed method. This comparison values are displayed in Table 3.

Table 3. The proposed model performance comparison with existing techniques

SNo	Model Name	RMSE (%)	NSE
1	Deep ESN Model (Echo state Network) [9] (2019)	1.51	0.02

Copyright © 2022. Journal of Northeastern University. Licensed under the Creative Commons Attribution Non-commercial No Derivatives (by-nc-nd). Available at https://dbdxxb.cn/

2	Multiple Linear Regression [8] (2020)	26.5	0.837
3	Artificial Neural Network [11] (2018)	68.49	0.69
4	The proposed model (SSA-LSSVAR, SSA-LSRF)	0.816	0.942

CONCLUSIONS AND FUTURE WORK

The current research investigated the dependability of integrating a data preprocessing approach (SSA) with ML-based strategies, LSSVR and RF, for monthly precipitation estimating in Nalgonda, India. One of the significant discoveries is that the crossbreed approaches (SSA-LSSVR &SSA-RF) have superior accuracy over the customary approaches (LS-SVR &RF). It tends to be inferred that the crossbreed models are a potential modelling procedure that can be applied to anticipate the monthly precipitation in the current examination area. Connecting the precipitation properties of study area to the estimating design exhibitions is recommended to be additionally examined in future research. Further, only one data pre-processing strategy has been accepted; consequently, future work may consider different pre-processing methods and contrast their accuracies with accomplish added exact prescient results. Additionally, just the areal monthly precipitation data from one District was utilized in the present research. More study regions ought to be incorporated for testing the findings.

REFERENCES

[1] Bojang PO, Yang TC, Pham QB, Yu PS. Linking Singular Spectrum Analysis and Machine Learning for Monthly Rainfall Forecasting. Applied Sciences. 2020 Jan;10(9):3224.

[2] Kashiwao T, Nakayama K, Ando S, Ikeda K, Lee M, Bahadori A. A neural network-based local rainfall prediction system using meteorological data on the Internet: A case study using data from the Japan Meteorological Agency. Applied Soft Computing. 2017 Jul 1;56:317-30.

[3] Reddy PC, Babu AS. Survey on weather prediction using big data analytics. In2017 Second International Conference on Electrical, Computer and Communication Technologies (ICECCT) 2017 Feb 22 (pp. 1-6). IEEE.

[4] Basha CZ, Bhavana N, Bhavya P, Sowmya V. Rainfall Prediction using Machine Learning & Deep Learning Techniques. In2020 International Conference on Electronics and Sustainable Communication Systems (ICESC) 2020 Jul 2 (pp. 92-97). IEEE.

[5] Choi C, Kim J, Kim J, Kim D, Bae Y, Kim HS. Development of heavy rain damage prediction model using machine learning based on big data. Advances in Meteorology. 2018 Jun 13;2018.

[6] P Chandrashaker Reddy and A Suresh Babu, “An Enhanced Multiple Linear Regression Model for Seasonal Rainfall Prediction”, International Journal of Sensors, Wireless Communications and Control (2020) 10: 1.

[7] Das S, Chakraborty R, Maitra A. A random forest algorithm for nowcasting of intense precipitation events. Advances in Space Research. 2017 Sep 15;60(6):1271-82.

[8] Mohammed M, Kolapalli R, Golla N, Maturi SS. Prediction Of Rainfall Using Machine Learning Techniques.

[9] Yen MH, Liu DW, Hsin YC, Lin CE, Chen CC. Application of the deep learning for the prediction of rainfall in Southern Taiwan. Scientific reports. 2019 Sep 4;9(1):1-9.

Copyright © 2022. Journal of Northeastern University. Licensed under the Creative Commons Attribution Non-commercial No Derivatives (by-nc-nd). Available at https://dbdxxb.cn/

[10] Reddy PC, Sureshbabu A. An applied time series forecasting model for yield prediction of agricultural crop. InInternational Conference on Soft Computing and Signal Processing 2019 Jun 21 (pp. 177-187). Springer, Singapore.
[11] Shah U, Garg S, Sisodiya N, Dube N, Sharma S. Rainfall Prediction: Accuracy Enhancement Using Machine Learning and Forecasting Techniques. In2018 Fifth International Conference on Parallel, Distributed and Grid Computing (PDGC) 2018 Dec 20 (pp. 776-782). IEEE.
[12] Abbot J, Marohasy J. The potential benefits of using artificial intelligence for monthly rainfall forecasting for the Bowen Basin, Queensland, Australia. Water Resources Management VII. 2013;171:287.
[13] Fahimi F, Yaseen ZM, El-shafie A. Application of soft computing based hybrid models in hydrological variables modeling: a comprehensive review. Theoretical and applied climatology. 2017 May 1;128(3-4):875-903.
[14] Kisi O, Shiri J. Precipitation forecasting using wavelet-genetic programming and wavelet-neuro-fuzzy conjunction models. Water resources management. 2011 Oct 1;25(13):3135-52.
[15] Pandhiani SM, Shabri AB. Time series forecasting using wavelet-least squares support vector machines and wavelet regression models for monthly stream flow data.
[16] Chan JC, Paelinckx D. Evaluation of Random Forest and Adaboost tree-based ensemble classification and spectral band selection for ecotope mapping using airborne hyperspectral imagery. Remote Sensing of Environment. 2008 Jun 16;112(6):2999-3011.
[17] Karthikeyan L, Kumar DN. Predictability of nonstationary time series using wavelet and EMD based ARMA models. Journal of hydrology. 2013 Oct 10;502:103-19.
[18] Ji SY, Sharma S, Yu B, Jeong DH. Designing a rule-based hourly rainfall prediction model. In2012 IEEE 13th International Conference on Information Reuse & Integration (IRI) 2012 Aug 8 (pp. 303-308). IEEE.
[19] Min, Min, Chen Bai, Jianping Guo, Fenglin Sun, Chao Liu, Fu Wang, Hui Xu et al. "Estimating summertime precipitation from Himawari-8 and global forecast system based on machine learning." IEEE Transactions on Geoscience and Remote Sensing 57, no. 5 (2018): 2557-2570.
[20] Navid, M. A. I., and N. H. Niloy. "Multiple Linear Regressions for Predicting Rainfall for Bangladesh." Communications 6, no. 1 (2018): 1-4.
[21] Rodrigues, Jesleena, and Arti Deshpande. "Prediction of Rainfall for all the States of India Using Auto-Regressive Integrated Moving Average Model and Multiple Linear Regression." In 2017 International Conference on Computing, Communication, Control and Automation (ICCUBEA), pp. 1-4. IEEE, 2017.
[22] Swain, S., P. Patel, and S. Nandi. "A multiple linear regression model for precipitation forecasting over Cuttack district, Odisha, India." In 2017 2nd International Conference for Convergence in Technology (I2CT), pp. 355-357. IEEE, 2017.
[23] Mohd, Razeef, Muheet Ahmed Butt, and Majid Zaman Baba. "SALM-NARX: Self Adaptive LM-based NARX model for the prediction of rainfall." In 2018 2nd International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud)(I-SMAC) I-SMAC (IoT in Social, Mobile, Analytics and Cloud)(I-SMAC), 2018 2nd International Conference on, pp. 580-585. IEEE, 2018.
[24] Faulina, Ria, Dwi Ayu Lusia, Bambang W. Otok, and Heri Kuswanto. "Ensemble method based on anfis-arima for rainfall prediction." In 2012 International Conference on Statistics in Science, Business and Engineering (ICSSBE), pp. 1-4. IEEE, 2012.

Copyright © 2022. Journal of Northeastern University. Licensed under the Creative Commons Attribution Non-commercial No Derivatives (by-nc-nd). Available at https://dbdxxb.cn/

[25] Cramer, Sam, Michael Kampouridis, Alex A. Freitas, and Antonis K. Alexandridis. "An extensive evaluation of seven machine learning methods for rainfall prediction in weather derivatives." Expert Systems with Applications 85 (2017): 169-181.
[26] Rivero, Cristian Rodríguez, Julian Pucheta, Josef Baumgartner, Sergio Laboret, and Victor Sauchelli. "Short-series Prediction with BEMA Approach: application to short rainfall series." IEEE Latin America Transactions 14, no. 8 (2016): 3892-3899.
[27]]http://cgwb.gov.in/District_Profile/Telangana/Nalgonda.pdf
[28] https://en.climate-data.org/asia/india/andhra-pradesh/nellore-6270/

Copyright © 2022. Journal of Northeastern University. Licensed under the Creative Commons Attribution Non-commercial No Derivatives (by-nc-nd). Available at https://dbdxxb.cn/