https://doi.org/10.31436/iiumej.v23i1.1822

DEVELOPMENT OF RAINFALL FORECASTING MODEL USING MACHINE LEARNING WITH SINGULAR SPECTRUM ANALYSIS

PUNDRU CHANDRA SHAKER REDDY[1*], YADALA SUCHARITHA[2] AND GODDUMARRI SURYA NARAYANA[3]

[1]*Department of Computer Science and Engineering, CMR College of Engineering & Technology, Hyderabad, India*
[2]*Department of Computer Science and Engineering, CMR Institute of Technology, Hyderabad, India*
[3]*Department of Computer Science and Engineering, Vardhaman College of Engineering, Hyderabad, India*

[*]*Corresponding author: chandu.pundru@gmail.com*

(Received: 4th February 2021; Accepted: 14th April 2021; Published online: 4th January 2022)

ABSTRACT: Agriculture is the key point for survival for developing nations like India. For farming, rainfall is generally significant. Rainfall updates are help for evaluate water assets, farming, ecosystems and hydrology. Nowadays, rainfall anticipation has become a foremost issue. Forecast of rainfall offers attention to individuals and knows in advance about rainfall to avoid potential risk to shield their crop yields from severe rainfall. 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 rainfall 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 a monthly climate dataset which is separated into 80:20 ratios for training and testing respectively. Performance of the model was assessed using Root Mean Square Error (RMSE) and Nash–Sutcliffe Efficiency (NSE) and the proposed model produces the values as 71.6 %, 90.2 % respectively. Experimental outcomes illustrate that the proposed model can productively predict the rainfall.

ABSTRAK: Pertanian adalah titik utama kelangsungan hidup negara-negara membangun seperti India. Untuk pertanian, curah hujan pada amnya ketara. Kemas kini hujan adalah bantuan untuk menilai aset air, pertanian, ekosistem dan hidrologi. Kini, jangkaan hujan telah menjadi isu utama. Ramalan hujan memberikan perhatian kepada individu dan mengetahui terlebih dahulu mengenai hujan untuk menghindari potensi risiko untuk melindungi hasil tanaman mereka dari hujan lebat. Kajian ini bertujuan untuk menyelidiki kebolehpercayaan mengintegrasikan teknik pra-pemprosesan data yang disebut analisis-spektrum tunggal (SSA) dengan model pembelajaran yang diawasi yang disebut regresi vektor sokongan paling rendah (LS-SVR), dan Random-Forest (RF), ramalan hujan. Menggabungkan SSA dengan LS-SVR dan RF, kerangka gabungan dirancang dan dibeza-bezakan dengan pendekatan biasa (LS-SVR dan RF). Kerangka kerja yang disajikan dilatih dan diuji dengan menggunakan set data iklim bulanan yang masing-masing dipisahkan menjadi nisbah 80:20 untuk latihan dan ujian. Prestasi model dinilai menggunakan Root Mean Square Error (RMSE) dan Nash – Sutcliffe Efficiency (NSE) dan model yang dicadangkan menghasilkan nilai masing-masing sebanyak 71.6%,

https://doi.org/10.31436/iiumej.v23i1.1822

90.2%. Hasil eksperimen menggambarkan bahawa model yang dicadangkan dapat meramalkan hujan secara produktif.

KEYWORDS:*singular-spectrum-analysis; machine learning;rainfall; SVR; RF*

1. INTRODUCTION

India's welfare is farming, where most the agribusiness is subject to rainfall as its standard wellspring of water, the time and proportion of rainfall 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 the latest couple of centuries in the field of science and innovation. As India's economy notably relies upon cultivation, rainfall has a significant impact [1]. Variation in the timing of rainfall and its quantity makes estimating rainfall 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 rainfall. Climate anticipation stands apart for all nations around the world in all the advantages and administrations gave by the meteorological department [2].

Rainfall anticipation is significant because severe and sporadic rainfall can have numerous effects like obliteration of yields and farm, harm of property so a superior anticipating model is crucial for an early notice that can limit dangers to life and property and also dealing with the farming a better way. This forecasting predominantly helps ranchers and furthermore, water assets can be used proficiently [3]. Rainfall forecast is a difficult task and the outcomes ought to be precise. There are numerous equipment tools for anticipating rainfall by utilizing the climate conditions like temperature, humidity, pressure and so on. These customary strategies can't work in a proficiently so by utilizing ML-based strategies we can design for exact outcomes. We can only do it by having the past data examination of rainfall and can foresee the rainfall for future seasons [4]. Taking into account that farming activities and crop yield depend on the rainfall distribution, monthly rainfall estimating is significant for farming planning and flood control. Monthly rainfall anticipation with a suitable technique is a fundamental prerequisite to help water management. Accordingly, monthly rainfall 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 because stochastic approaches consider that the time-series are fixed and have a restricted capacity to catch profoundly nonlinear qualities of rainfall series. Climatology time-series in tangible utilize are normally non-fixed and non-linear, so foreseeing the greatest values is very complex [6]. The hypothesis of the modular approach and the incorporation of various models have presently increased more enthusiasm for rainfall 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 is generally applied, direct elucidation of the guidelines is complex. Then again, classification utilizing decision tree methods is helps display complex associations among attributes with the additional focal points of recognizing the significance of every attribute

https://doi.org/10.31436/iiumej.v23i1.1822

[7]. RF and SVR are computationally quick, effectively reasonable and don't need earlier knowledge of 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 to turn into general inductive practice in rainfall 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 rainfall [8]. The most commonly utilized ML-based approaches for rainfall 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 because 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.

The Singular Spectrum Analysis (SSA) is a proficient device that can divide the original time-series into a number 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. 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 rainfall estimation.

We have seen that the majority of the works claiming higher accuracy have labeled rainfall into three or under three parts or have predicted rainfall utilizing ML methods but have not done rainfall anticipating utilizing ML strategies, very few of them have utilized barely any meteorological parameters for the anticipation of the rainfall [10]. Most of the existing works are utilized the regression methods for forecasting rainfall and outcomes are not up to the mark due to selection of correct parameters for modeling. We have proposed a model to foresee the rainfall utilizing a combination ML-based techniques. The expectation of rainfall relies upon different climate attributes. Categorizing the rainfall gives us great classification precision but our definitive objective is to anticipate the rainfall utilizing the other climate attributes [11]. In this investigation, objective isn't just to accurately classify rainfall but additionally effectively foresee the rainfall utilizing different climate attributes.

There is persuading proof that the hybrid design dependent on LS-SVR and RF is solid and that the SSA generates excellent results. Subsequently, for enhancing the prediction efficiency of the state-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 and RF) to build hybrid models (SSA-LSSVR and SSA-RF) for rainfall prediction.

https://doi.org/10.31436/iiumej.v23i1.1822

2. Comparison among the crossbreed models and the conventional models to assess the effectiveness of the data pre-processing strategy.

The Proposed work is concentrated around understanding the impacts of various meteorological attributes in rainfall estimation alongside an investigation of approaches that were utilized for anticipating rainfall, ML, and their restriction. The proposed model predicts the rainfall 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.

2. LITERATURE REVIEW

In this section, the existing works on rainfall forecasting proposed by various researchers are presented. Abbot et al. [12] designed a rainfall forecasting model by artificial neural networks which produce more precise results compared to conventional statistical and numerical techniques. Fahimi et al. [13] developed a hybrid framework called an Adaptive-Neuro-Fuzzy-Inference-System (ANFIS) for accurate long-term rainfall prediction. The outcome shows that the ANFIS is fit for catching the rainfall data-dynamic behavior and produces agreeable results. 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 rainfall anticipating. Pandhiani et al. [15] described a rainfall anticipation model by SSA and SVR for seasonal rainfall anticipation. The outcomes illustrate a noteworthy growth in model effectiveness contrasted to the standard SVR approach. 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. Karthikeyan et 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 conveys the finest output. Finally, there is no investigation of RF-dependent models with a data pre-processing strategy for rainfall prediction, which timely this recent research to present a crossbreed model pairing RF with a data pre-processing approach. Ji et al. [18] designed a rainfall prediction model by a 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 Minet al. [19] explored and designed an algorithm called quantitative rainfall estimates (QPEs) based on random forest (RF), machine learning (ML) techniques for summer-time rainfall forecasting utilizing the 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. Navidet 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. Rodrigueset al. [21] designed the Multiple Linear Regression (MLR) technique

https://doi.org/10.31436/iiumej.v23i1.1822

and time-series ARIMA techniques for monthly rainfall anticipation. The experimental results demonstrate that MLR and ARIMA produce accurate forecast results over the traditional forecasting methods. In the end, the proposed performance of the two models is evaluated in terms of MAPE and the outcomes and the exactness of predictions made for rainfall by MLR (1.14) is found to be greater over ARIMA (19.61). Swainet 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 R^2 (0.974) and adjusted R^2 (0.963) when compared to the existing models.Razeef Mohdet 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 rainfall data were gathered from Kashmir and Jammu, India. The experiments were conducted and model performance was evaluated with RMSE (1.721%) and MSE (1.721%) values.

Faulinaet 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. Crameret 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. Riveroet 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 rainfall sequence and chosen sequence from standards.

Mehr et al. [27] presented a new model called hybrid regression for advance month precipitation forecasting in northwest Iran. It is designed using SVR-FFA and it is trained and tested using monthly rainfall data. Johny et al. [28] designed a framework named AEEMD-ANN for rainfall forecasting and the experimental outcomes show that it is flourishing in predicting SWM precipitation of year 2002. Samantaray et al. [29] used RNN, ANFIS and SVM combinations for precipitation investigation. They conclude that SVM is gives good performance compared to others. Zhao et al. [30] developed an enhanced precipitation prediction model using five predictors. The experimental outcomes display that the IRFM gives good accuracy with five predictors inside of one predictor.

3. STUDY AREA AND DATA COLLECTION

Weather data is in extremely enormous amount and it includes forty-seven (47) general parameters such as rainfall, temperature, humidity, sunshine and cloudy hours, wind speed, etc. The major problem of this section is to investigate and process theclimate

https://doi.org/10.31436/iiumej.v23i1.1822

data of the study region, Nellore district, Andhra Pradesh, 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

Nellore zone of Andhra Pradesh State in India is the investigation region for the evaluation of rainfall anticipation. Nellore station (latitude 14°26' N/longitude 79°92' E) is located in the coastal Andhra Pradesh region and it is surrounded on the South and North by Chittoor and Prakasam districts, on the west by Veligonda hills neighboring Kadapa district and on the east by the Bay of Bengal as shown in Fig. 1. The total area of the Nellore region covers 1307600 hectares and 39 % of the region is bound to farming. The major crops around there are rice, sugarcane, cotton, sunflower, groundnut, and tobacco. Rainfall got during the southwest monsoon (June-September) is one of the main factors for the groundwater and the average annual rainfall of the Nellore is around 835 mm [31].

The study area is one of the most significant agricultural regions in the nation. Rapid and independent industrialized development projects have caused environmental changes in the earlier period, consequently raised the need to evaluate elements impacting the present climate anticipation. The daily rainfall information, estimated in millimeters (mm), was acquired from the IMD, Hyderabad, TS, and India. Severe rainfall is seen in many regions during June to September 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 Nellore district encounters substantial rainfall 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 [32].

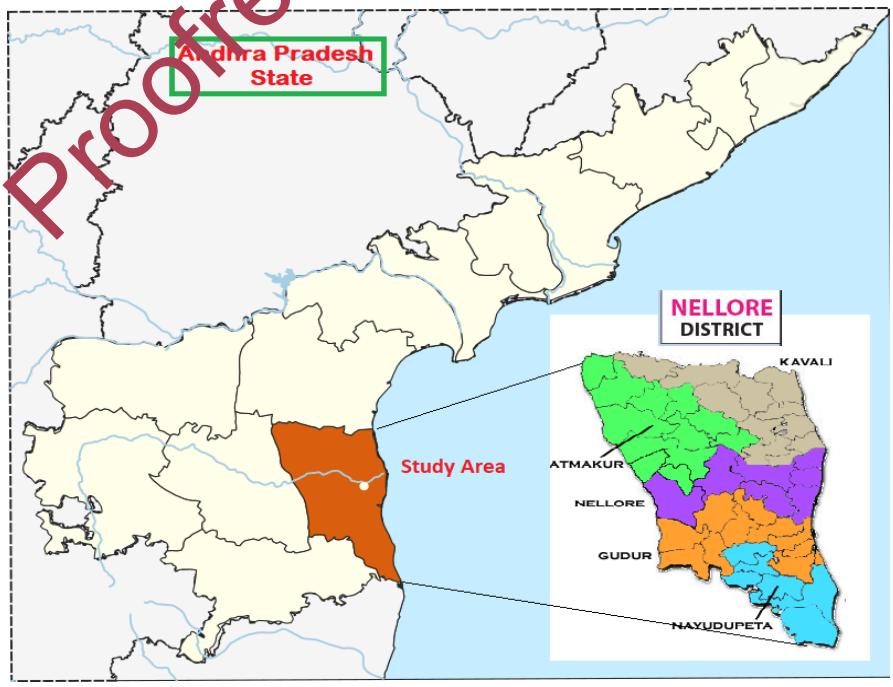

Fig. 1: Geographical area of Nellore district (Source: en.wikepedia.org).

https://doi.org/10.31436/iiumej.v23i1.1822

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 (°C), vapor pressure (hPa), wind speed (km/h), humidity (%) and cloud cover (%). The dataset details are shown in Table 1.

Table 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 Nellore region, 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, the 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. To figure out the attributes of the monthly rainfall series in Nellore Dist., the evocative statistics were utilized for investigating their rainfall properties.

4. PROPOSED METHODOLOGY

The problem examined in this section is that of anticipation of precise monthly rainfall 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 rainfall value to the Nellore region 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. Given a training set as input x_i and output y_i, the regression equation that interfaces the input vector to the output can be defined 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 LS-SVR is utilized to anticipate monthly rainfall. In the accompanying

https://doi.org/10.31436/iiumej.v23i1.1822

sector, we discuss the essential hypothesis on LS-SVR in time-series anticipating. Because of a trainingset as input x_i and output y_i, the regression equation is defined as follows in Eq. (1).

$$F(y) = w^T \mathbb{Q}(x) + b \quad (1)$$

Where, $\mathbb{Q}(x)$ 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 Eqs. (2) and (3):

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

Subject to the following limitations:

$$y_i = w^T \mathbb{Q}(x_i) + b + e_i \ (i = 1, 2, : : : , m) \quad (3)$$

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

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

Where $\alpha_i$represents the Lagrange values.

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

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

4.2 Random Forest

Random-Forest is the most remarkable ML technique for analytics, including a collection of basic trees. RF is an improved version of the decision-tree dependent on the strategy for bagging. In bagging, numerous stochastic-error situations are performed by picking training-sets independently and arbitrarily from the training set; totaling the anticipation of every individual trained model is cultivated by taking its average value. RF can likewise evaluate the noteworthiness of disclosing factors to the estimate as per the straightforward guideline "the more pertinent the illustrative variable, the more significant the impact on the prediction" to utilize the RF model for the selection of factors. In bootstrap-sampling, the annotations are isolated into two parts for every approach: in-bag-subset and out-of-bag subset [33]. The out-of-bag subsets might be utilized to decide the noteworthiness for 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.

4.2 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 refuses the requirement for an earlier method [34]. 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

https://doi.org/10.31436/iiumej.v23i1.1822

(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 the 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.

4.3 Linking SSA with LS-SVR and RF

The monthly rainfall 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 rainfall 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 approach through the cross-validation system. RF has two elements, the number of factors $\sqrt{M}$and the number of trees (ntree), which should be estimated. $\sqrt{M}$ would for the most part produce close ideal outcomes, so the estimation of $\sqrt{M}$ was chosen by experimentation utilizing the incentive around $\sqrt{M}$ (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 incentive for ntree of 750. Subsequently, the estimation of 750 for ntree was embraced in this research.

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

The set of suitable inputs is a significant worry for LS-SVR and RF modelling. Diverse mix of forerunner values of the rainfall 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 rainfall 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.

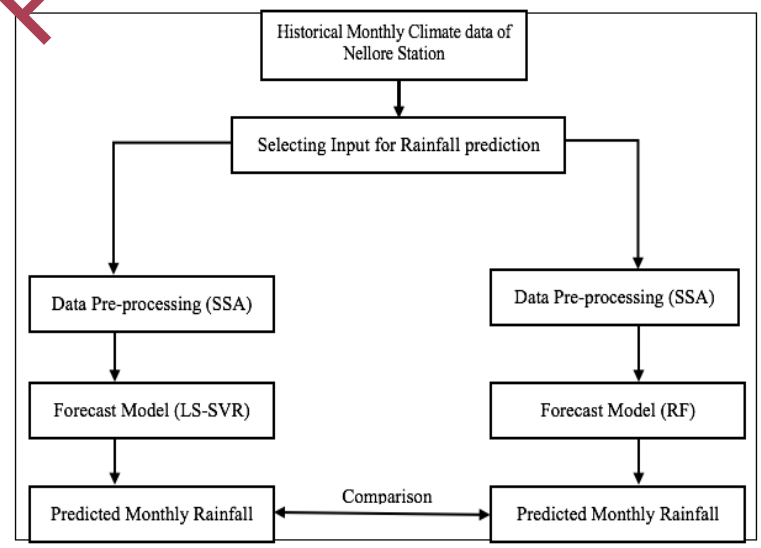

Fig. 2: The famework of the proposed rainfall prediction model.

https://doi.org/10.31436/iiumej.v23i1.1822

The proposed framework has the following steps and is depicted in Fig. 2.

1. At first, the time-series of rainfall 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 and RF approaches are designed for each part of the reconstruction so the design of LS-SVR and RF is distinctive for every segment of the restoration.
4. At last, LS-SVR and 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.

4.4 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 Eqs. (6) and (7) respectively.

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

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

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

Where M is the mean of the fitted values. An NSE of 0.75–1.0 relates to"excellent" accuracy, 0.65–0.75 to "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.

5. RESULTS AND DISCUSSION

In this article we designed the monthly rainfall prediction approaches for 1, 2, 3, 4-month lead time for the Nellore station. 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 hybrid models in terms of RMSE and NSE.

Table 2:Prediction performances of the proposed models

Model	RMSE				NSE			
	1-month	2-month	3-month	4-month	1-month	2-month	3-month	4-month
LSSVR	194.58	193.31	199.56	199.85	0.04	0.06	0.05	0.07
SSA-LSSVR	**72.29**	**69.48**	**71.09**	**73.54**	**0.96**	**0.95**	**0.86**	**0.84**
RF	197.87	199.24	194.18	200.54	0.04	0.05	0.04	0.03
SSA-RF	**111.76**	**126.77**	**132.83**	**142.56**	**0.73**	**0.59**	**0.49**	**0.65**

https://doi.org/10.31436/iiumej.v23i1.1822

In Table 3, it is seen that RMSE and 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 additional 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 rainfall data will diminish the noise impacts and its helpful in enhancing the model performance. Figures 3(a,b) and 4(a,b) outline the time-series charts of the fitted and one-month lead-time anticipated rainfall by the customary and hybrid approaches of Nellore station. From the figures, we uncover that the predicted values from the 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 rainfall determined by the customary models is not near the line of balance.

5.1 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 Nellore station. Here we had taken the average values of 1-, 2-, 3-, 4- month percentages RMSE, and NSE values for comparison of the proposed method. These 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
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)**	**0.716**	**0.902**

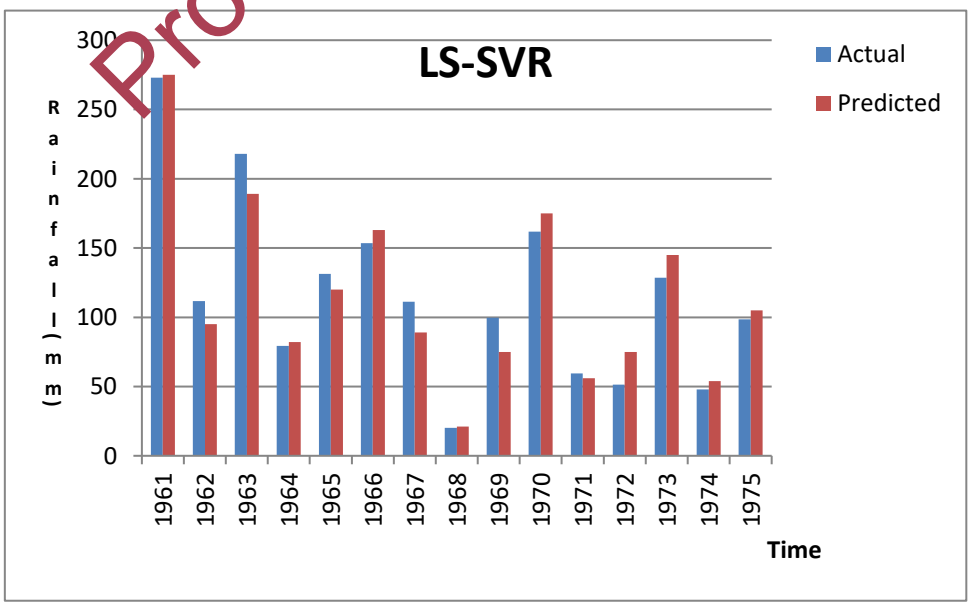

Fig. 3a: Fitted and predicted values for LSSVR model.

Early Access Date: 13 October 2021

https://doi.org/10.31436/iiumej.v23i1.1822

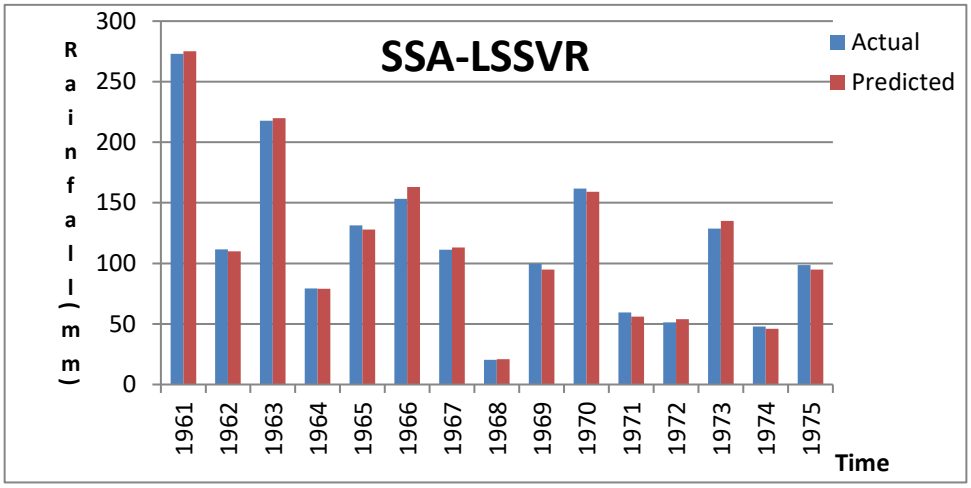

Fig. 3b: Fitted and Predicted values for SSA-LSSVR model.

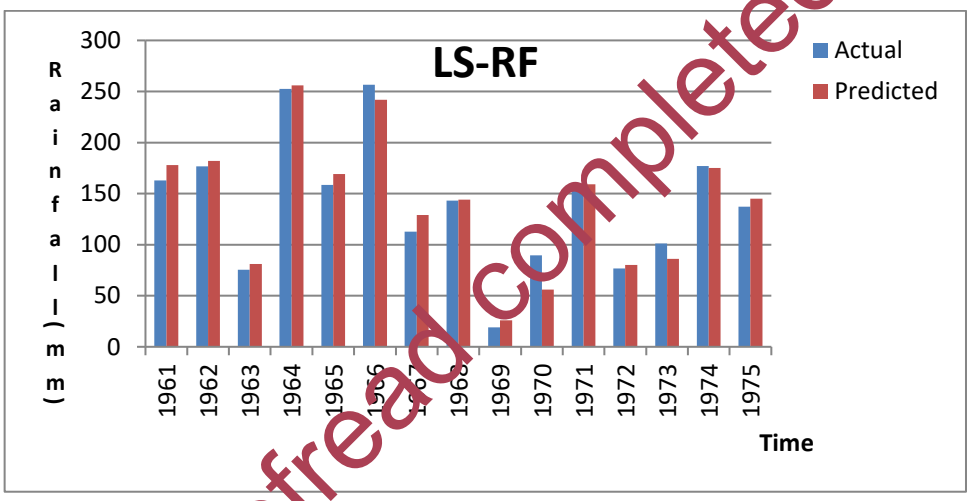

Fig. 4a: Fitted and predicted values for LSRF model.

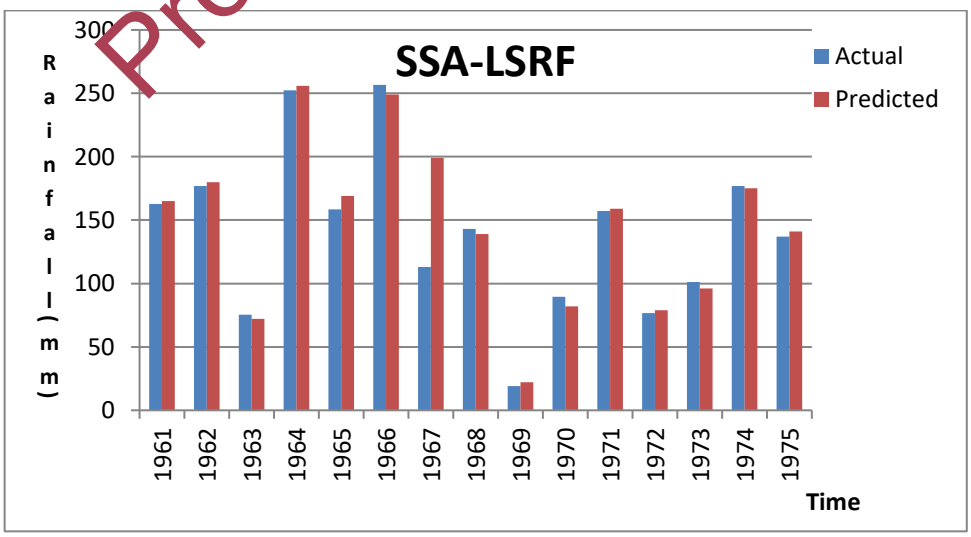

Fig. 4b: Fitted and predicted values for SSA-LSRF model.

https://doi.org/10.31436/iiumej.v23i1.1822

6. CONCLUSION 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 rainfall estimating in Nellore station, India. The study infers that the SSA can acquire and give magnificent forecasts to noteworthy hydrological time series segments with unique uneven practices, for example, rainfall and runoff series. One of the significant discoveries is that the crossbreed approaches (SSA-LSSVR and SSA-RF) have superior accuracy over the customary approaches (LS-SVR and RF). It tends to be inferred that the crossbreed models are a potential modelling procedure that can be applied to anticipate the monthly rainfall in the current examination area. Connecting the rainfall properties of the 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 accomplishing added exact prescient results. Additionally, just the areal monthly rainfall data from one station 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. (2020)Linking singular spectrum analysis and machine learning for monthly rainfall forecasting. Applied Sciences, 10(9):1-20.

[2] Kashiwao T, Nakayama K, Ando S, Ikeda K, Lee M, Bahadori A. (2017) 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, 56(1):317-330.

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

[4] Basha CZ, Bhavana N, Bhavya P, Sowmya V. (2020) Rainfall prediction using machine learning & deep learning techniques. InInternational Conference on Electronics and Sustainable Communication Systems (ICESC), IEEE: pp 92-97.

[5] Choi C, Kim J, Kim J, Kim D, Bae Y, Kim HS. (2018) Development of heavy rain damage prediction model using machine learning based on big data. Advances in Meteorology, 2018 (2):1-11.

[6] Reddy PC, Babu AS. (2020) An enhanced multiple linear regression model for seasonal rainfall prediction, International Journal of Sensors, Wireless Communications and Control, 10(1):473-483.

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

[8] Moulana M, Roshitha K, Niharika G, Sai MS. (2020) Prediction of rainfall using machine learning techniques. International Journal of Scientific & Technology Research, 9(3):236-240.

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

[10] Reddy PC, Sureshbabu A. (2019) An applied time series forecasting model for yield prediction of agricultural crop. InInternational Conference on Soft Computing and Signal Processing, Springer: pp 177-187.

[11] Shah U, Garg S, Sisodiya N, Dube N, Sharma S. (2018) Rainfall prediction: Accuracy enhancement using machine learning and forecasting techniques. InFifth International Conference on Parallel, Distributed and Grid Computing (PDGC), IEEE: pp 776-782.

https://doi.org/10.31436/iiumej.v23i1.1822

[12] Abbot J, Marohasy J. (2013) The potential benefits of using artificial intelligence for monthly rainfall forecasting for the Bowen Basin, Queensland, Australia. Water Resources Management VII, 171:287.

[13] Fahimi F, Yaseen ZM, El-shafie A. (2017) Application of soft computing based hybrid models in hydrological variables modeling: a comprehensive review. Theoretical and Applied Climatology, 128(3-4):875-903.

[14] Kisi O, Shiri J. (2011) Rainfall forecasting using wavelet-genetic programming and wavelet-neuro-fuzzy conjunction models. Water Resources Management, 25(13):3135-3152.

[15] Pandhiani SM, Shabri AB. (2013) Time series forecasting using wavelet-least squares support vector machines and wavelet regression models for monthly stream flow data. Open Journal of Statistics, 3: 183-194.

[16] Chan JC, Paelinckx D. (2008) 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, 112(6):2999-3011.

[17] Karthikeyan L, Kumar DN. (2013) Predictability of nonstationary time series using wavelet and EMD based ARMA models. Journal of Hydrology, 502:103-119.

[18] Ji SY, Sharma S, Yu B, Jeong DH. (2012) Designing a rule-based hourly rainfall prediction model. InIEEE 13th International Conference on Information Reuse & Integration (IRI), IEEE, pp 303-308.

[19] Min M, Bai C, Guo J, Sun F, Liu C, Wang F, Xu H, et al. (2018) Estimating summertime rainfall from Himawari-8 and global forecast system based on machine learning. IEEE Transactions on Geoscience and Remote Sensing, 57(5): 2557-2570.

[20] Navid MAI, Niloy NH. (2018) Multiple linear regressions for predicting rainfall for Bangladesh. Communications, 6(1): 1-4.

[21] Rodrigues J, Deshpande A. (2017) Prediction of rainfall for all the states of India using auto-regressive integrated moving average model and multiple linear regression. In International Conference on Computing, Communication, Control and Automation (ICCUBEA), IEEE: pp 1-4.

[22] Swain S, Patel P, Nandi S. (2017) A multiple linear regression model for rainfall forecasting over Cuttack district, Odisha, India. In 2nd International Conference for Convergence in Technology (I2CT), IEEE: pp 355-357.

[23] MohdRazeef, Butt MA, and Baba MZ. (2018) SALM-NARX: Self Adaptive LM-based NARX model for the prediction of rainfall. In 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), IEEE:pp 580-585.

[24] Ria F, Lucia DA, Otok BW, Kuswanto H. (2012) Ensemble method based on anfis-arima for rainfall prediction. In International Conference on Statistics in Science, Business and Engineering (ICSSBE), IEEE: pp 1-4.

[25] Cramer S, Kampouridis M, Freitas AA, Alexandridis AK. (2017) An extensive evaluation of seven machine learning methods for rainfall prediction in weather derivatives. Expert Systems with Applications, 85(2): 169-181.

[26] Rivero CR, Pucheta JA, Baumgartner JS, Laboret SO, Sauchelli VH, Patiño HD. (2016) Short-series Prediction with BEMA Approach: application to short rainfall series. IEEE Latin America Transactions,14(8): 3892-3899.

[27] Mehr AD, Nourani V, Khosrowshahi VK, Ghorbani MA. (2019) A hybrid support vector regression–firefly model for monthly rainfall forecasting. International Journal of Environmental Science and Technology, 16(1):335-346.

[28] Johny K, Pai ML, Adarsh S. (2020) Adaptive EEMD-ANN hybrid model for Indian summer monsoon rainfall forecasting. Theoretical and Applied Climatology, 18(1):1-7.

[29] Samantaray S, Tripathy O, Sahoo A, Ghose DK. (2020) Rainfall forecasting through ANN and SVM in Bolangir Watershed, India. InSmart Intelligent Computing and Applications, Springer: pp 767-774.

[30] Zhao Q, Liu Y, Ma X, Yao W, Yao Y, Li X. (2020) An improved rainfall forecasting model based on GNSS observations. IEEE Transactions on Geoscience and Remote Sensing, 58(7):4891-900.
[31] https://en.wikipedia.org/wiki/Nellore_district. 15.01.2021
[32] https://en.climate-data.org/asia/india/andhra-pradesh/nellore-6270/. 15.01.2021
[33] Abdel-Kader H, Abd-El Salam M, Mohamed M. (2021) Hybrid Machine Learning Model for Rainfall Forecasting. Journal of Intelligent Systems and Internet of Things, 1(1):5-12.
[34] Pham QB, Yang TC, Kuo CM, Tseng HW, Yu PS. (2021) Coupling singular spectrum analysis with least square support vector machine to improve accuracy of SPI drought forecasting. Water Resources Management, 35(3):847-868.