International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 7s
DOI: https://doi.org/10.17762/ijritcc.v11i7s.7010
***Article Received:** 02 April 2023 **Revised:** 20 May 2023 **Accepted:** 02 June 2023*

Leveraging Machine Learning based Ensemble Time Series Prediction Model for Rainfall Using SVM, KNN and Advanced ARIMA+ E-GARCH

[1]C.Nagesh, [2] Koushik Reddy Chaganti, [3] Sathvik Chaganti, [4] S.K.Khaleelullah, [5] P.Naresh, [6] M.I.Thariq Hussan
[1]Assistant Professor,Dept of CSE, Srinivasa Ramanujan Institute of Technology, Anantapur
[2] Masters in Computer Science, Dept of Computer and Information Systems, University of the Cumberlands, KY USA
[3] Masters in Computer Science, Dept of Computer Science, Georgia State University,Atlanta
[4,5] Assistant.Professor, Dept of IT, Vignan Institute of Technology and Science, Hyderabad
[6] Prof. & Head, Department of IT, Guru Nanak Institutions Technical Campus, Hyderabad, India

Abstract— Today's precipitation is growing increasingly variable, making forecasting increasingly difficult. The Indian Meteorological Department (IMD) currently employs Composite and Stochastic approaches to forecast spring storm precipitation in Asia. As a corollary, planners are unlikely to predict the macroeconomic effects of disasters (due to excessive precipitation) or famine (less precipitation). The amount of precipitation that drops dependent on a variety of factors, including the temperature of the atmosphere, humidity, velocity, mobility, and weather conditions. This paper would then employ the Hybrid time-series predictive ARIMA+ E-GARCH (Exponential Generalized Auto-Regressive Conditional Heteroskedasticity) to predict precise runoff by taking into account different climatic considerations such as maritime tension, water content, relative dampness, min-max heat, heavy ice, geostrophic tallness, breeze patterns, soil dampness, and barometric force. In perspective of RMSE, MAE, and MSE, the proposed hybrid ARIMA+E-GARCH paradigm outperformed single simulations and latest hybrid techniques.

Keywords— Yearly precipitation, rainfall frequency, Hybrid Time series prediction models.

I. INTRODUCTION

Researchers in the discipline of hydrology are always seeking for innovative ways to get a good awareness of the Earth 's environment and construct reliable weather forecasting simulations. A number of methodologies have been used in weather forecasting [1]. In recent years, Machine Learning (ML) perspectives such as Decision Tree (DT), Seasonal Auto Regressive Integrated Moving Average (SARIMA), Multiple Linear Regression model (MLR), Support Vector Machine (SVM), Linear Regression (LR), Logistic Regression (LGR), and Random Decision Forest (RDF) have indeed been progressively used to replace traditional climatologically forecasting methodologies [2]. One of the greatest prominent environmental occurrences is precipitation severity, which has a huge influence on agriculture and ecosystems.

The precipitation severity in India varies depending on the environment. Agricultural laborers reportedly faced huge financial losses and destruction as a result of the sudden rains [14]. As a consequence of all this, agricultural workers are trying to commit murder on purpose. A comparative research and forecast of precipitation using cutting-edge methodologies can reduce revenue damage and agricultural labor death. Severe and inadequate precipitation has remained a source of worry for remote zones like Andhra Pradesh. If precipitation amounts can be predicted advance of time, precautions can be taken to protect the crop. The goal of this study, which incorporates machine learning approaches, is to predict rainfall amounts. Because weather patterns are uncertain, it is critical to create trustworthy weather forecasting systems that can save lives by successfully warning people to an oncoming disaster.

The amount of precipitation expected is a major worry for the IMD in terms of individual survival and the business. Drought and flooding are two of the greatest prevalent natural disasters on the planet, and the major cause is abundant rains. Environmental forecasting is being used to aid any business in making judgments on what to do in the event of a disaster [16]. Estimating the quantity of precipitation to fall is one of the greatest significant and challenging tasks in climate modeling. Furthermore, those techniques employ ML and deep learning (DL) techniques, both of which are capable of being used [3-4]. Furthermore, those techniques employ ML and DL techniques, both of which are capable of being used [3-4].

The following is the continuation of this manuscript: The related work explained in section II, section III focused on implementation, section IV results and finally section V conclusion and ended with references.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 7s
DOI: https://doi.org/10.17762/ijritcc.v11i7s.7010
***Article Received:** 02 April 2023 **Revised:** 20 May 2023 **Accepted:** 02 June 2023*

II. RELATED WORK

Numerous investigators, intellectuals, statisticians, and professionals have used varied methods and methodologies to anticipate amount of rain, such as DT [14], DESN [16], BPNN [9], and Support Vector Regression (SVR) by analyzing multiple quality assessment specifications such as RMSE.

A study on precipitation estimation techniques is shown in Table 1.

Table 1. A Survey of existing models

Study	Method	Environmental factors considered	Case Study Location	Evaluation parameter	Interpretation of Research
C. Z. Basha et al., (2020). [5]	DL	heat, moisture	India	Epoch, RMSE	The 20th period produced the best results, with the best Epoch number for the testing database coming during the 26th stage.
R. K. Grace et al., (2020). [6]	MLR	Wind speed	India	RMSE	RMSE = 3.449
S. Kaushik et al.,(2020). [7]	ML	Barometric pressure	Punjab	MAE, RMSE	Training and testing accuracy of 95% and 92%.
N. Tiwari et al., (2020). [8]	NN	Thermal gradient	India	MAE	MAE=2.20
S. Srivastava et al., (2020). [9]	ML	humidity	Uttarakhand	MSE, MAE, and RMSE	BPNN has the capability of dominating and providing the least MSE, MAE, and RMSE numbers.
U. Harita et al., (2020). [10]	. ML	Dew point	India	Accuracy	Their research focused on forecasting precipitation and agricultural homicide rates in India. In terms of extreme precipitation prediction, RDF outperforms LR, LGR, and SVM.
H. A. Y. Ahmed et al., (2020). [11]	MLR	temperature	Khartoum	RMSE	RMSE reduced by 85%.
D. S. Rani et al., (2020). [12]	ML	Pressure	Hyderabad	MAE	MAE=21.8
U. Ashwini et al., (2021). [13]	ML	humidity	Tamilnadu	RMSE, MSE	The ARIMA framework accurately forecasts Monsoon precipitation with substantially less volatility.
A. Samad et al., (2020). [15]	NN	Wind direction	Australia	MSE	Good results were achieved by LSTM when compared to ANN
M. I. Khan et al., (2020). [17]	DL	moisture	Maharashtra	RMSE	RMSE = 6.6 to 24.19
I. Prakaisak et al., (2021). [18]	ML	warmth	Thailand	Accuracy	Accuracy can be enhanced
P. Zhang et al., (2020). [19]	MLP	hotness	China	RMSE	RMSE=1.61

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 7s
DOI: https://doi.org/10.17762/ijritcc.v11i7s.7010
***Article Received:** 02 April 2023 **Revised:** 20 May 2023 **Accepted:** 02 June 2023*

III. METHODOLOGY

A. ARIMA

The AM (u) approach forecasts probable tendencies by looking at the sequenced content of prior occurrences for a characteristic. By merging already anticipated deficits consecutively in the history, the MM (v) prototype estimates the projected variance in precipitation information. When combined with the MM (v) and AM (u) algorithms, ARMA (Autoregressive Moving Average) may be used for asymmetric and dynamic time information. It investigates the relationship between previous precipitation information and preceding precipitation information strategically mistakes. As a consequence, when dealing with non-static processes like precipitation information, the ARIMA framework joins the forecasting theory. Because the ARIMA (u, e, v) paradigm mixes AM(u) and MM(v) approaches using little deconvolution to make supply stable, it may be used using non-stationary information. ARIMA is a combination of the Autoregressive Model AM(u) and the Moving Average Model MM (v).

B. GARCH

For response variable with quasi elements, including such precipitation events, the GARCH paradigm is utilised. The GARCH paradigm evaluates the constant distribution of the information. It works well on information that has a big SD. The GARCH paradigm is a modification of the ARCH paradigm that incorporates both the MM and AM terms. It merely displays the constant polynomial depending on prior residual errors. In GARCH, unpredictability is based on the prior frequency. The GARCH version incorporates the constant functional as a component of prior mistake cubed and its magnitude to forecast and quantify volatility variations.

C. E-GARCH

The EGARCH paradigm is a variation of the GARCH model. In the year 1991 Nelson created the E-GARCH paradigm to address the shortcomings of GARCH's management of monetary temporal periods. Allowing for asymmetrical impacts between favorable and unfavorable property gains, in general. An E-GARCH (u,v) is defined as follows:

$$X_t = \mu + A_t$$

The temporal sequence frequency at period *t* is x_t. μ is the GARCH paradigm's average. A_t is the residue of the simulation at period *t*. At period interval, σ_t is the contingent point difference (i.e. unpredictability). The ARCH constituent paradigm's sequence is u. $\alpha_o, \alpha_1, \alpha_2,...,\alpha_u$ are indeed the ARCH element paradigm's characteristics. The GARCH constituent paradigm's sequence is v.$\beta_1, \beta_2,..., \beta_v$ are indeed the E-GARCH element paradigm's characteristics. The standardized samples are represented by [ϵ_t].

In various areas, the E-GARCH paradigm varies from the GARCH paradigm. The recorded contingent deviations, for example, were utilized to reduce the potential restriction of prototype parameters. EGARCH is an unitary EGARCH framework. The EGARCH module returns a co-integration item that holds the field settings and specifies the operational shape of an EGARCH (U,V) paradigm. The following are crucial elements of an EGARCH framework: The GARCH equation is made up of postponed and recorded contingent deviations. The level is represented by the U. ARCH quadratic formula, which is made up of delayed standardized creativity amplitudes. Stagnated standardised advances make up the leveraged coefficient. V is the greatest of the ARCH and leveraging quadratic grades.

In the GARCH exponential, U is the greatest nonnegative delay, while in the ARCH and leveraged coefficients, V is the greatest positive integer discrepancy. An invention median modeling offsetting, a variation decomposition modelling perpetual, and the discoveries dispersion are other modelling elements. Until you use the moniker combination parameter style, all equations are uncertain (NaN numbers) and exemplary. Employ estimates to guess models with full or substantially undetermined component rating provided information. Modeling or anticipate reactions utilizing simulation or forecasting, correspondingly, for thoroughly described structures (features whereby all variable values are computed). By addressing the leveraged implications of a market shift on the dependent variation, the EGARCH paradigm gives an additional unbalanced framework. As a result, a significant price drop might had a greater influence on unpredictability than just a significant pricing gain. EGARCH is a standard statistical framework with explanatory variables.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 7s
DOI: https://doi.org/10.17762/ijritcc.v11i7s.7010
***Article Received:** 02 April 2023 **Revised:** 20 May 2023 **Accepted:** 02 June 2023*

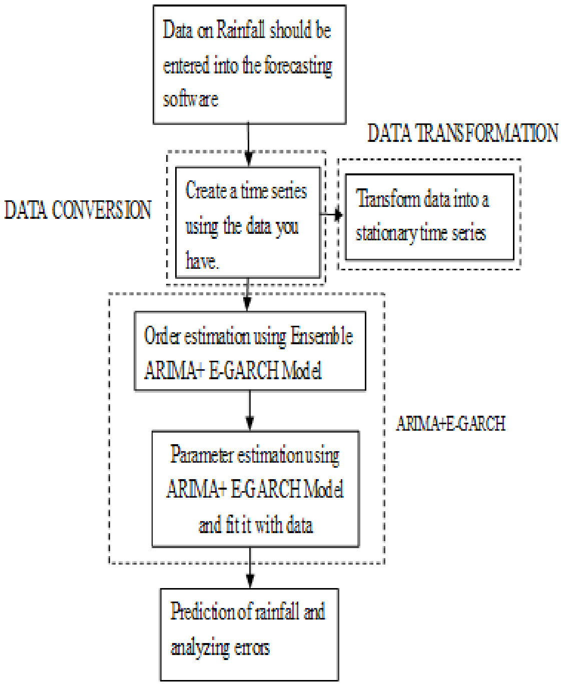

Fig. 1 Architecture of suggested ensemble ARIMA+GARCH Model

D. ARIMA+ E-GARCH Paradigm

A simple introduction to the ARIMA and E-GARCH systems demonstrates the importance of using the constant function towards precipitation forecasting. A covariance is another name for the constant distribution. The first step is to input precipitation information with the forecasting programme. Make a time period utilising the information you have and convert it to a fixed timed period in phase 2. In phase 3, the recommended composite ARIMA+ E-GARCH framework is used to estimate sequence and parameters. Finally, in step 4, precipitation predictions are made and inaccuracies are examined. Figure 1 depicts the entire procedure of the proposed technique. The hybrid ARIMA+ E-GARCH paradigm is indeed covered in this chapter.

IV. RESULTS AND DISCUSSION

The failure comparison between ARIMA+E-GARCH and latest composite techniques is shown in Fig 2. In respect of RMSE, MSE, and MAE, the recommended approach has a lower failure ratio of 1.115, 2.129, and 2.102, which is indicated in green in table 2.

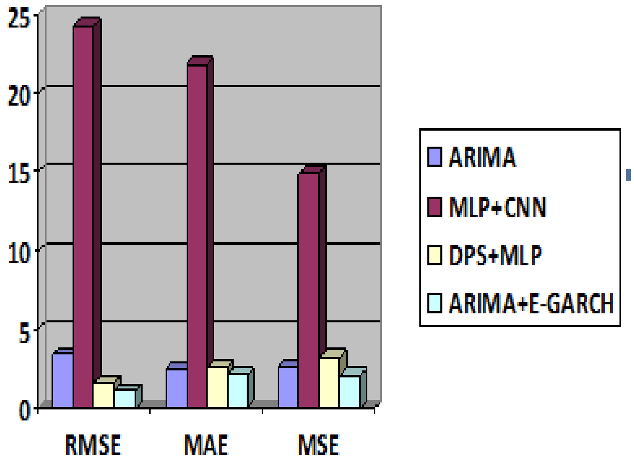

Fig.2 Performance comparison of all models

Table 2. Failure comparison of the suggested composite technique with current composite methods

Method	RMSE	MAE	MSE
MLP+CNN	6.7 to 23.28	22.9	13.8
DPS+MLP	1.52	2.489	3.129
ARIMA+E-GARCH	1.267	2.325	3.842

Table.3 KNN,SVM and proposed model performances

	Precission	Recall	F1 score
KNN	94.13	93.57	93.75
SVM	92.97	94.35	94.66
Proposed	95.36	96.35	95.22

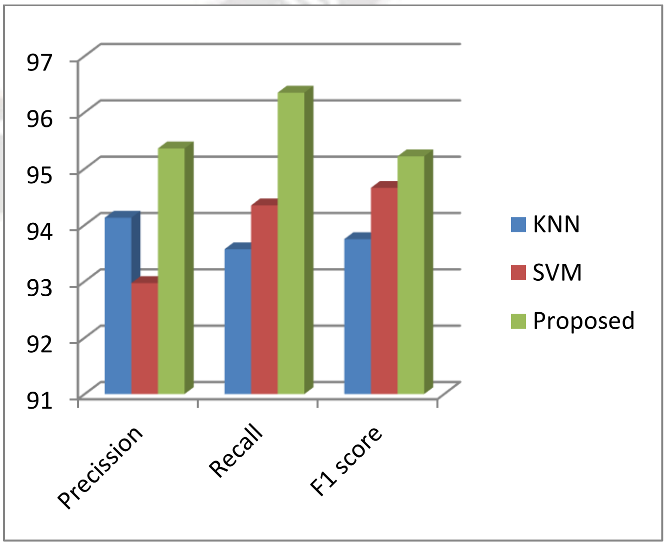

Fig.3 Graph for Accurracy levels of various algorithms over proposed

V. CONCLUSION AND FUTURE SCOPE

Understanding the strength of precipitation during the spring period is critical to agricultural output. The quantity of rainfall determines the amount of agriculture items produced. To aid farmed labourers in agrarian productivity, a year's rainfall should be forecast. The proposed method uses cointegration hybrid forecasting algorithms to examine the quantity of precipitation for the Indian database and provides much superior results in respect of RMSE, MAE, and MSE. In the upcoming, we want to investigate precipitation forecasting using GARCH modeling. Comparisons were made with existing works and the proposed one given best results by means of accuracy and efficiency.

REFERENCES

[1] J. Niu and W. Zhang, "Comparative analysis of statistical models in rainfall prediction," 2015 IEEE International Conference on Information and Automation, 2015, pp. 2187-2190, doi: 10.1109/ICInfA.2015.7279650.

[2] A. Kala and S. G. Vaidyanathan, "Prediction of Rainfall Using Artificial Neural Network," 2018 International Conference on Inventive Research in Computing Applications (ICIRCA), 2018, pp. 339-342, doi: 10.1109/ICIRCA.2018.8597421.

[3] Mondal , D. . (2021). Remote Sensing Based Classification with Feature Fusion Using Machine Learning Techniques. Research Journal of Computer Systems and Engineering, 2(1), 28:32. Retrieved from https://technicaljournals.org/RJCSE/index.php/journal/article/view/16

[4] C. Thirumalai, K. S. Harsha, M. L. Deepak and K. C. Krishna, "Heuristic prediction of rainfall using machine learning techniques," 2017 International Conference on Trends in Electronics and Informatics (ICEI), 2017, pp. 1114-1117, doi: 10.1109/ICOEI.2017.8300884.

[5] P.Naresh,et.al., “Implementation of Map Reduce Based Clustering for Large Database in Cloud”, Journal For Innovative Development in Pharmaceutical and Technical Science,vol.1,pp 1-4,2018.

[6] C. Z. Basha, N. Bhavana, P. Bhavya and S. V, "Rainfall Prediction using Machine Learning & Deep Learning Techniques," 2020 International Conference on Electronics and Sustainable Communication Systems (ICESC), 2020, pp. 92-97, doi: 10.1109/ICESC48915.2020.9155896.

[7] R. K. Grace and B. Suganya, "Machine Learning based Rainfall Prediction," 2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS), 2020, pp. 227-229, doi: 10.1109/ICACCS48705.2020.9074233.

[8] S. Kaushik, A. Bhardwaj and L. Sapra, "Predicting Annual Rainfall for the Indian State of Punjab Using Machine Learning Techniques," 2020 2nd International Conference on Advances in Computing, Communication Control and Networking (ICACCCN), 2020, pp. 151-156, doi: 10.1109/ICACCCN51052.2020.9362742.

[9] N. Tiwari and A. Singh, "A Novel Study of Rainfall in the Indian States and Predictive Analysis using Machine Learning Algorithms," 2020 International Conference on Computational Performance Evaluation (ComPE), 2020, pp. 199-204, doi: 10.1109/ComPE49325.2020.9200091.

[10] S. Srivastava, N. Anand, S. Sharma, S. Dhar and L. K. Sinha, "Monthly Rainfall Prediction Using Various Machine Learning Algorithms for Early Warning of Landslide Occurrence," 2020 International Conference for Emerging Technology (INCET), 2020, pp. 1-7, doi: 10.1109/INCET49848.2020.9154184.

[11] U. Harita, V. U. Kumar, D. Sudarsa, G. R. Krishna, C. Z. Basha and B. S. S. P. Kumar, "A Fundamental Study on Suicides and Rainfall Datasets Using basic Machine Learning Algorithms," 2020 4th International Conference on Electronics, Communication and Aerospace Technology (ICECA), 2020, pp. 1239-1243, doi: 10.1109/ICECA49313.2020.9297440.

[12] H. A. Y. Ahmed and S. W. A. Mohamed, "Rainfall Prediction using Multiple Linear Regressions Model," 2020 International Conference on Computer, Control, Electrical, and Electronics Engineering (ICCCEEE), 2021, pp. 1-5, doi: 10.1109/ICCCEEE49695.2021.9429650.

[13] S. Khaleelullah, P. Marry, P. Naresh, P. Srilatha, G. Sirisha and C. Nagesh, "A Framework for Design and Development of Message sharing using Open-Source Software," 2023 International Conference on Sustainable Computing and Data Communication Systems (ICSCDS), Erode, India, 2023, pp. 639-646, doi: 10.1109/ICSCDS56580.2023.10104679.

[14] Naresh, P., & Suguna, R. (2019). Association Rule Mining Algorithms on Large and Small Datasets: A Comparative Study. 2019 International Conference on Intelligent Computing and Control Systems (ICCS). DOI:10.1109/iccs45141.2019.9065836.

[15] S. J. Basha, G. L. V. Prasad, K. Vivek, E. S. Kumar and T. Ammannamma, "Leveraging Ensemble Time-series Forecasting Model to Predict the amount of Rainfall in Andhra Pradesh," 2022 2nd International Conference on Artificial Intelligence and Signal Processing (AISP), 2022, pp. 1-7, doi: 10.1109/AISP53593.2022.9760553.

[16] Prof. Barry Wiling. (2018). Identification of Mouth Cancer laceration Using Machine Learning Approach. International Journal of New Practices in Management and Engineering, 7(03), 01 - 07. https://doi.org/10.17762/ijnpme.v7i03.66

[17] A. Samad, Bhagyanidhi, V. Gautam, P. Jain, Sangeeta and K. Sarkar, "An Approach for Rainfall Prediction Using Long Short Term Memory Neural Network," 2020 IEEE 5th International Conference on Computing Communication and Automation (ICCCA), 2020, pp. 190-195, doi: 10.1109/ICCCA49541.2020.9250809.

[18] M. I. Thariq Hussan, D. Saidulu, P. T. Anitha, A. Manikandan and P. Naresh (2022), Object Detection and Recognition in Real Time Using Deep Learning for Visually Impaired People. IJEER 10(2), 80-86. DOI: 10.37391/IJEER.100205.

[19] B. Narsimha, Ch V Raghavendran, Pannangi Rajyalakshmi, G Kasi Reddy, M. Bhargavi and P. Naresh (2022), Cyber Defense in the Age of Artificial Intelligence and Machine Learning for

Financial Fraud Detection Application. IJEER 10(2), 87-92. DOI: 10.37391/IJEER.100206.

[20] Omondi, P., Rosenberg, D., Almeida, G., Soo-min, K., & Kato, Y. A Comparative Analysis of Deep Learning Models for Image Classification. Kuwait Journal of Machine Learning, 1(3). Retrieved from http://kuwaitjournals.com/index.php/kjml/article/view/128

[21] I. Prakaisak, E. Phaisangittisagul, M. Maleewong, K. Sarinnapakorn and C. Chantrapornchai, "Detecting Anomaly and Replacement Prediction for Rainfall Open Data in Thailand," 2021 18th International Joint Conference on Computer Science and Software Engineering (JCSSE), 2021, pp. 1-6, doi: 10.1109/JCSSE53117.2021.9493814.

[22] T. Aruna, P. Naresh, A. Rajeshwari, M. I. T. Hussan and K. G. Guptha, "Visualization and Prediction of Rainfall Using Deep Learning and Machine Learning Techniques," 2022 2nd International Conference on Technological Advancements in Computational Sciences (ICTACS), Tashkent, Uzbekistan, 2022, pp. 910-914, doi: 10.1109/ICTACS56270.2022.9988553.