Nanotechnology Perceptions
ISSN 1660-6795
www.nano-ntp.com

Predicting Rainfall And Humidity Using Machine Learning Models

Uday Kumar Sinha[1,*], K. Bandyopadhayay[2] and S.C.Dutta[3]

[1.] *University Department of Physics, BBMK University, Dhanbad, Jharkhand, 826001*
&
Department of Vocational Studies, Guru Nanak College, Dhanbad, Jharkhand, 826001
Email ID: Udaydhn@gmail.com, ORCID ID: 0000-0003-2858-2463

[2.] *University Department of Physics, BBMK University, Dhanbad, Jharkhand, 826001*
Email ID: dr.krishnanath123@gmail.com, ORCID ID: 0009-0001-3356-1811

[3.] *Department of CSE and IT, BIT Sindri, Dhanbad, Jharkhand,* 828123
Email ID: dutta_shubhash@yahoo.com, ORCID ID: 0009-0002-9947-3619
** Corresponding Author: Uday Kumar Sinha, Department of Vocational Studies, Guru Nanak College, Dhanbad, Jharkhand – 826001, Email Id: udaydhn@gmail.com*

In the present paper nine machine learning tools have been considered for rainfall as well as humidity prediction in Dhanbad, Jamshedpur and Ranchi districts of Jharkhand. Our study reveals that the Random Forest and XGBoost (RF_XG) hybrid model gives most accurate results for rainfall (~87%) and humidity (~86%) parameters with minimum absolute error(~10%) for both rainfall and humidity in case of these three districts. It is to be mentioned that Random Forest and XGBoost exhibit better performance (~82% for rainfall and ~81% for humidity) next to this hybrid model. The linear regression and decision tree are found to be good in terms of minimum absolute error but these models yield poor accuracy for prediction of both rainfall and humidity. Interestingly, RF_XG hybrid model demonstrates best average accuracy (~84%) with average mean absolute error (~3%) for both rainfall and humidity predictions during July and August of 2024 for above three districts. It is therefore, unhesitatingly be stated that the hybrid model RF_XG is to be considered as best among all nine machine learning techniques for prediction of rainfall and humidity for aforesaid three districts of Jharkhand.

Keywords: Machine Learning, Rainfall, Humidity, Hybrid models, Python.

1. Introduction

Rainfall is a key factor for agricultural development of any country. It is because of the fact that except a few countries, most of the countries are either developing or underdeveloped and hence agricultural activities depends on natural rainfall. Before the development of Machine learning techniques in the field of computer science, people in general were dependent on natural phenomenon for prediction of rainfall. However, after the fast development of computer science, in particular data analysis based on Artificial Intelligence (AI) techniques

[1,2], it has become comparatively easier to make prediction about rainfall locally as well as globally. At the same time, rainfall also affects production of crops a lot. Due to excessive rainfall, floods are also seen at some places of different countries every year. It is therefore necessary to get accurate information about rainfall earlier and then we can definitely reduce all these losses, if can not eliminate them totally.

Forecasting of rainfall has been a challenging task for scientists from the very beginning. However, in the last several years, new technologies have emerged, and as a result, our predictions are now more accurate than they were in the past. In this context, we have considered models of machine learning [3, 4] that plays a critical role for the present purpose. In the present work, we have considered five basic models of machine learning such as Linear Regression (LR)[5], Support Vector Machine (SVM)[6,7], Decision Tree (DT)[8], Random Forest (RF)[9,10] and XGBoost[11,12] and developed four hybrid models [13] like Random Forest and XGBoost (RF_XG), Random Forest and Linear Regression (RF_LR), Random Forest and Support Vector Machine (RF_SVM) and XGBoost and Linear Regression (XG_LR) for prediction of rainfall and humidity of three major districts of Jharkhand i.e. Dhanbad, Jamshedpur and Ranchi.

Machine learning techniques are presented in section 2. Section 3 contains methodology of the present work. Results and Summary of the work are elucidated in section 4. Conclusions are given in section 5.

2. Machine Learning Techniques

Software technologies can be properly trained with complex data to classify, conjecture, and assess physical events and their qualities using Machine learning techniques. The aim of the present work is to predict effective pattern recognition and model self-learning by using different algorithms. The fundamental concepts of five basic and four hybrid algorithms are presented in the following subsections.

2.1 Linear Regression

One of the simplest and most widely used machine learning methods is linear regression. It is a statistical technique for forecasting analysis. The concept of "linear regression" refers to a process that displays a linear relationship between one or more independent variables and a dependent variable. Since linear regression displays a linear relationship, it can be used to determine how the value of the independent variable affects the value of the dependent variable.

2.2 Support Vector Machine

Support Vector Machines are supervised learning methods for problems with regression and classification. It produces a hyperplane that divides different data points into intervals from which the output of the model can be inferred. SVM is available in two separate versions. One variant is utilized for classification problems, while the other is applied to regression problems.

The kernel approach is used to transform data of lower dimensions to higher dimensional feature space due to nonlinear nature of problems. Kernel functions, such as linear, polynomial, radial basis function (RBF), sigmoid, hyperbolic tangent, etc., are utilized to convert the inseparable input data into separable ones. For current problem Radial Basis Function Kernel is considered because of greater accuracy.

2.3. Random Forest (RF)

Another well-known machine learning approach is random forest. This method practices supervised ensemble learning to complete tasks related to regression or classification. Using a training set of data, it constructs several decision trees. The mean value of the decision tree set is then used to envisage the value for new input data.

2.4. eXtreme gradient boosting (XGBoost)

One more significant machine learning technique is the XGBoost algorithm. XGBoost combines a set of decision trees with gradient boosting to make predictions. The XGBoost algorithm solves regression and classification problems using the community-based weak learning method. This algorithm produces useful results since it relies on parallel tree structures considering hardware as well as software parameters.

2.5. Decision Tree (DT)

A Decision Tree technique also involves regression and classification problems. In this tree-like model, a class label value is represented by leaf node and an attribute is characterised by interior node while a decision rule is signified by branch.

2.6 Stacking Ensemble Learning

The methodology's fundamental component is a stacking based ensemble learning model [14]. To increase the prediction model's accuracy, this model integrates the best features of several machine learning models. To handle the complexity of rainfall prediction, the ensemble method is especially useful.

3. Methodology

The proposed system deals with predictions of rainfall and humidity with nine machine learning techniques i.e., Linear Regression, Support Vector Machine, Decision Tree, Random Forest, XGBoost, RF_XG, RF_LR, RF_SVM and XG_LR for achieving more accurate solutions. The most effective algorithm(s) will then be provided to the output by the system after comparing the models. The stages associated with the proposed approach include data entry, data preprocessing, data division, training of an algorithm, dataset verification, comparison among the algorithms, prediction(s) of the most accurate algorithm(s), and final findings.

Collection, pre-processing, forecasting, and estimation of data are the four phases of the prediction system. The data for rainfall have taken from the website of Indian Metrological Department[15] and other weather prediction websites[16,17] containing several atmospheric parameters like minimum temperature, maximum temperature, humidity, pressure, wind speed, dew point and precipitation. The absence of information have been noticed and proper measures have considered in order to achieve optimal data analysis performance [18, 19].

3.1 Data Collection

	Date	Temp_Max	Temp_Min	Humidity	Pressure	Wind_Speed	Dew_Point	Percipitation_in_mm	City
0	2019-01-01	24	19	34	1200	8	5.00	0.0	Dhanbad
1	2019-01-02	25	20	21	1016	7	5.00	0.0	Dhanbad
2	2019-01-03	25	21	33	1017	7	4.00	0.2	Dhanbad
3	2019-01-04	26	20	39	1018	8	4.00	0.0	Dhanbad
4	2019-01-05	24	21	46	1016	9	4.00	0.0	Dhanbad
...	...	...	...	...	...	...	...	...	...
5473	2023-12-27	27	16	43	1017	2	17.78	0.0	Jamshedpur
5474	2023-12-28	27	16	44	1017	3	19.44	0.0	Jamshedpur
5475	2023-12-29	27	16	41	1016	3	15.56	0.0	Jamshedpur
5476	2023-12-30	27	15	38	1015	2	17.22	0.0	Jamshedpur
5477	2023-12-31	25	12	40	1012	4	15.30	0.0	Jamshedpur

5478 rows × 9 columns

Table 1: Rainfall data of last five consecutive years

The data set Comma Separated Values (CSV) file is shown in Table 1. In this data set, data for daily rainfall record of Dhanbad, Jamshedpur and Ranchi have been collected. It contains daily rainfall and humidity data from January 01, 2019 to December 31, 2023 and the data set consists of 5478 rows and it includes features such as maximum temperature, minimum temperature, Humidity, Pressure, Wind Speed, Dew point and precipitation.

3.2 Data Pre-processing

Data pre-processing [20] has four different stages as mentioned below.

(i) Removal of Null Values: When data are not available then there are two ways to eliminate the null values. Either the row containing the null value should be removed or the mean value of that specific column should be considered.

(ii) Removal of Outliers: A dataset may contain extreme values that are out of the expected range and inconsistent with the rest of the data. Deletion of these outliers improves model skill in general.

(iii) Feature Selection: A portion of the extensive feature space is selected to minimize the dimensionality of the dataset. Pearson's Correlation coefficient method [21] could be used at this stage.

(iv) Splitting of data: The function train_test_split () divides the database into two distinct sets: the test data and the train data. In present work, we use 80% of the data for algorithm training and reserve the remaining 20% as test data.

3.3 Designing of Model

Python 3.12.0[22] with packages like Matplotlib [23], Seaborn [24], and Sklearn [25] package are used for learning and prediction of different models with the help of Panda's [26] data frame. Algorithms discussed above are trained using training data whereas test data is used to validate them.

3.4. Evaluation Criteria of the Models

Statistical matrices such as R-Squared (R^2), Mean Absolute Error (MAE), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE) are used to evaluate the effectiveness of prediction models [27, 28].

4. Results and Summary

In the present work we have studied the physical entities i.e. rainfall and humidity with five basic important machine learning theories (Linear Regression, Support Vector Machine, Decision Tree, Random Forest and XGBoost) and four proposed hybrid models (RF_XG, RF_LR, RF_SVM and XG_LR) for last five years (January 01, 2019 to December 31, 2023) considering three important districts (Dhanbad, Jamshedpur and Ranchi) of Jharkhand state. Different values for accuracy and errors are presented below.

R^2 values of three districts for different models:

Models	Dhanbad	Jamshedpur	Ranchi
Linear Regression	0.656482	0.593168	0.405429
Decision Tree	0.782377	0.763157	0.813421
Random Forest	0.852237	0.824588	0.841287
Support Vector Machine	0.493322	0.446913	0.544164
XGBoost	0.821722	0.795641	0.823813
RF_XG	0.891601	0.884157	0.854192
RF_LR	0.607106	0.623189	0.682613
RF_SVM	0.580912	0.519535	0.565355
XG_LR	0.569145	0.457013	0.510633

Table 2: R^2 values of three districts for different models

MAE values of three districts for different models:

Model	Dhanbad	Jamshedpur	Ranchi
Linear Regression	0.0565894	0.0499442	0.0326717
Decision Tree	0.0445015	0.0453581	0.0329815
Random Forest	0.0410369	0.0397594	0.0300370
SVM	0.0397408	0.0697305	0.0497779
XGBoost	0.0398807	0.0367226	0.0338003
RF_XG	0.0419752	0.0413553	0.0484206
RF_LR	0.0449781	0.0436633	0.1176670
RF_SVM	0.0408863	0.0345013	0.0766739
XG_LR	0.0555289	0.0436814	0.0922800

Table 3: MAE values of three districts for different models

MSE values of three districts for different models:

Model	Dhanbad	Jamshedpur	Ranchi
Linear Regression	0.00683172	0.00774970	0.00166546
Decision Tree	0.00827764	0.00887415	0.00556598
Random Forest	0.00466395	0.00686898	0.00278523
SVM	0.00438731	0.01006610	0.00411757
XGBoost	0.00481740	0.00721084	0.00408571
RF_XG	0.00629540	0.00799940	0.01052930
RF_LR	0.00578520	0.00849293	0.01763850
RF_SVM	0.00584625	0.00753913	0.00897340
XG_LR	0.00987275	0.00892047	0.02497580

Table 4: MSE values of three districts for different models

RMSE values of three districts for different models:

Model	Dhanbad	Jamshedpur	Ranchi
Linear Regression	0.0826542	0.0880324	0.0408100
Decision Tree	0.0909816	0.0942027	0.0746055

Random Forest	0.0682931	0.0828793	0.0527752
SVM	0.0662367	0.1003300	0.0641683
XGBoost	0.0694075	0.0849167	0.0639196
RF_XG	0.0793436	0.0894394	0.1026130
RF_LR	0.0760605	0.0921571	0.1328100
RF_SVM	0.0764608	0.0868281	0.0947280
XG_LR	0.0993617	0.0944482	0.1580370

Table 5: RMSE values of three districts for different models

On the basis of the above data, bar charts and line charts have been designed for above mentioned three districts of Jharkhand.

(I A)For Dhanbad district the charts are given below

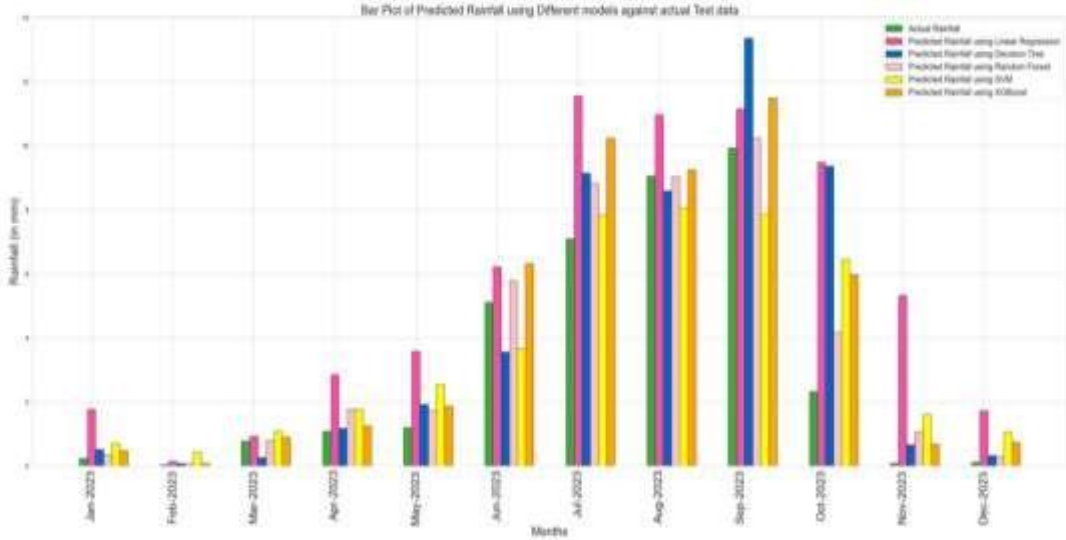

Fig. 1: Bar Plot of Predicted Rainfall of Dhanbad using Different Base ML Models

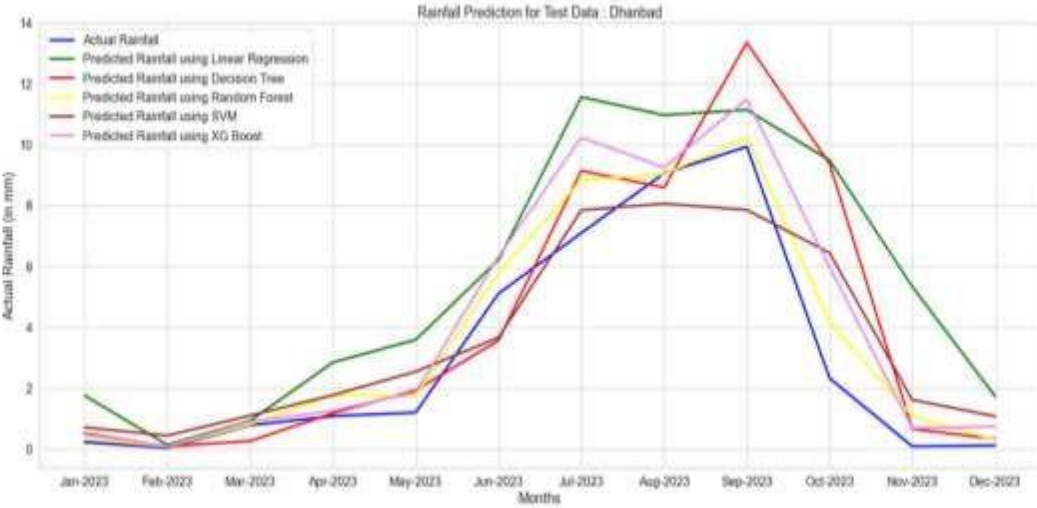

Fig. 2: Line Chart of Predicted Rainfall of Dhanbad using Different Base ML Models

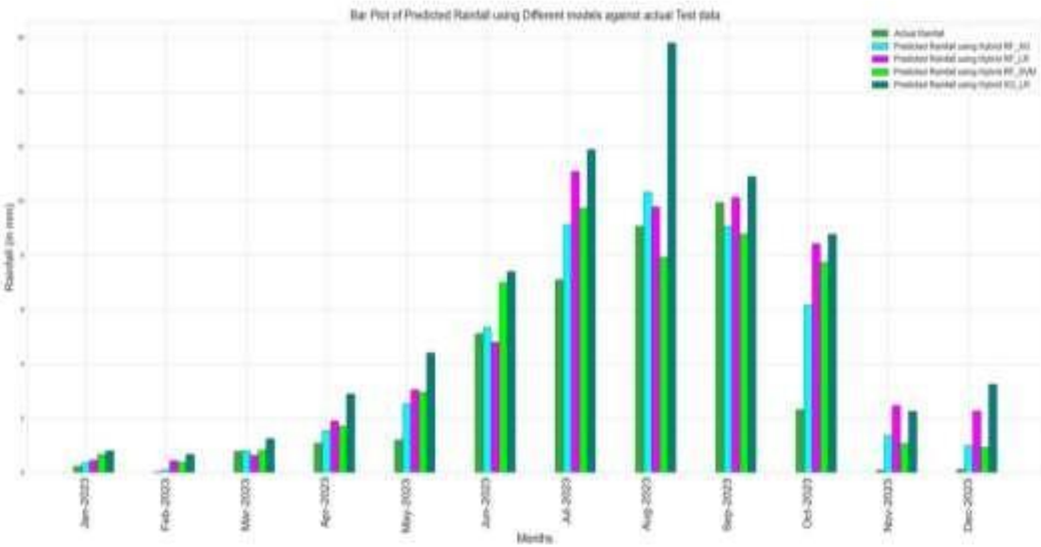

Fig. 3: Bar Plot of Predicted Rainfall of Dhanbad using Different Hybrid ML Models

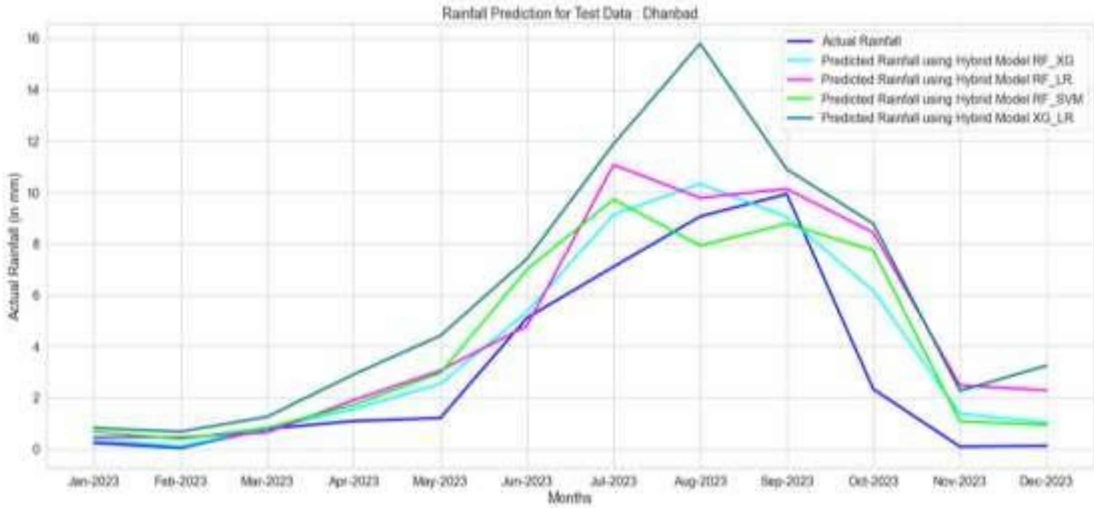

Fig. 4: Liner Chart of Predicted Rainfall of Dhanbad using Different Hybrid ML Models

(II A)Following charts are drawn for Jamshedpur district

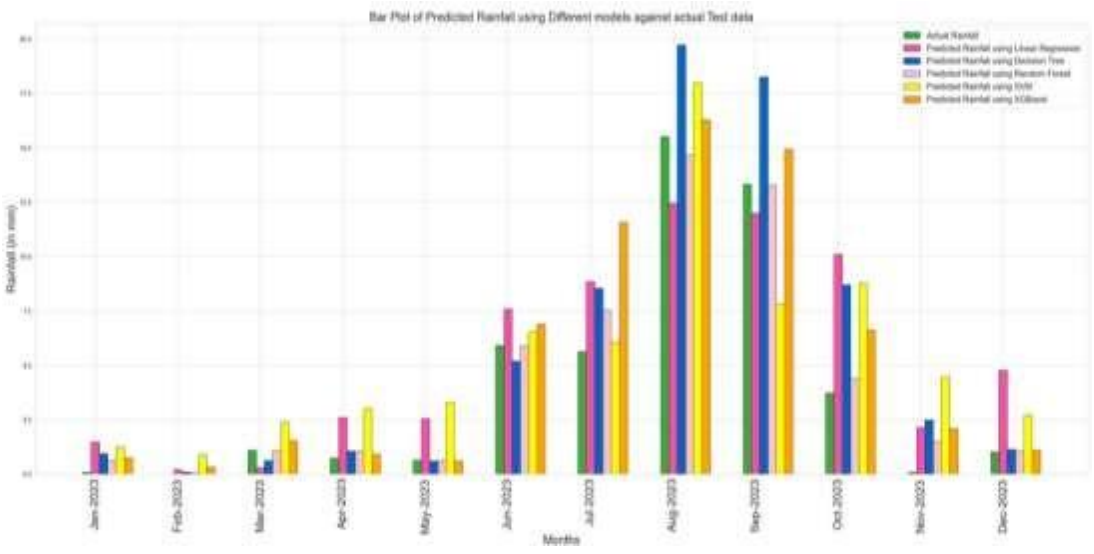

Fig. 5: Bar Plot of Predicted Rainfall of Jamshedpur using Different Base ML Models

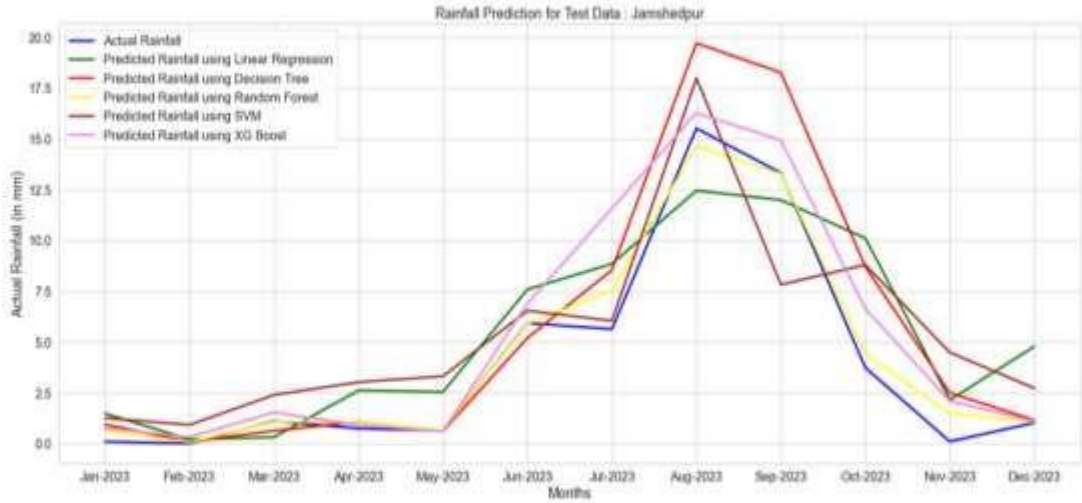

Fig. 6: Line Chart of Predicted Rainfall of Jamshedpur using Different Base ML Models

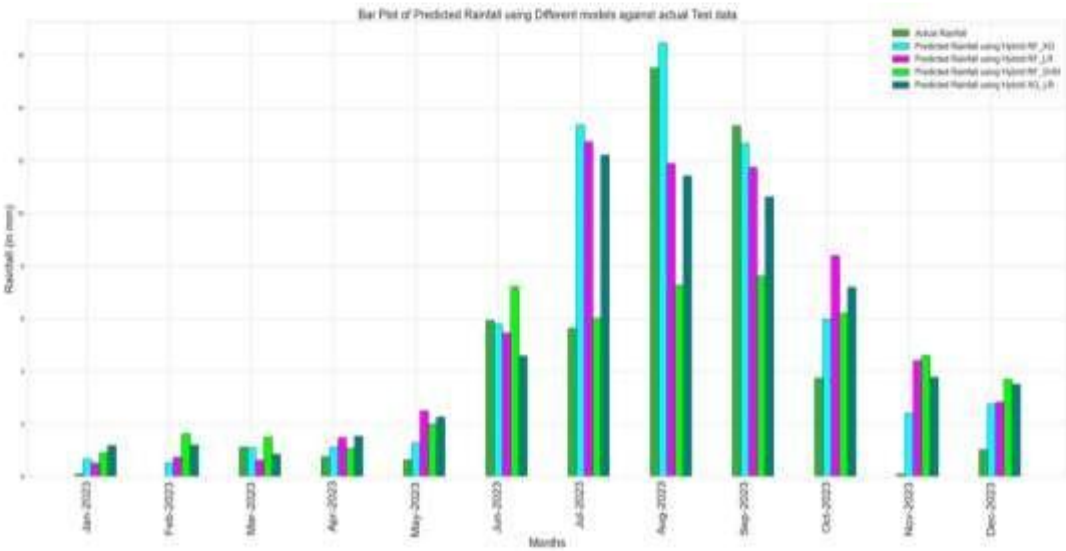

Fig. 7: Bar Plot of Predicted Rainfall of Jamshedpur using Different Hybrid ML Models

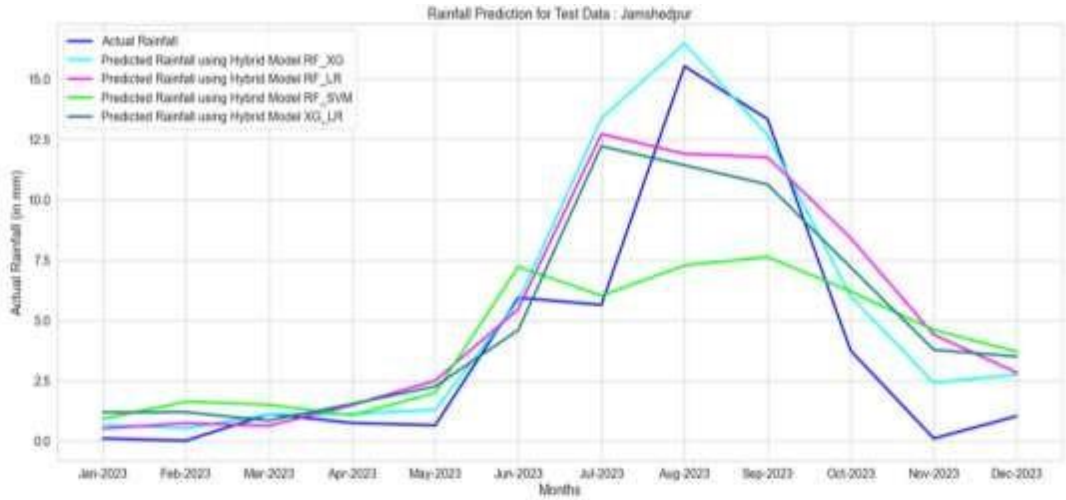

Fig. 8: Liner Chart of Predicted Rainfall of Jamshedpur using Different Hybrid ML Models

(III A) Charts for Ranchi district are plotted below

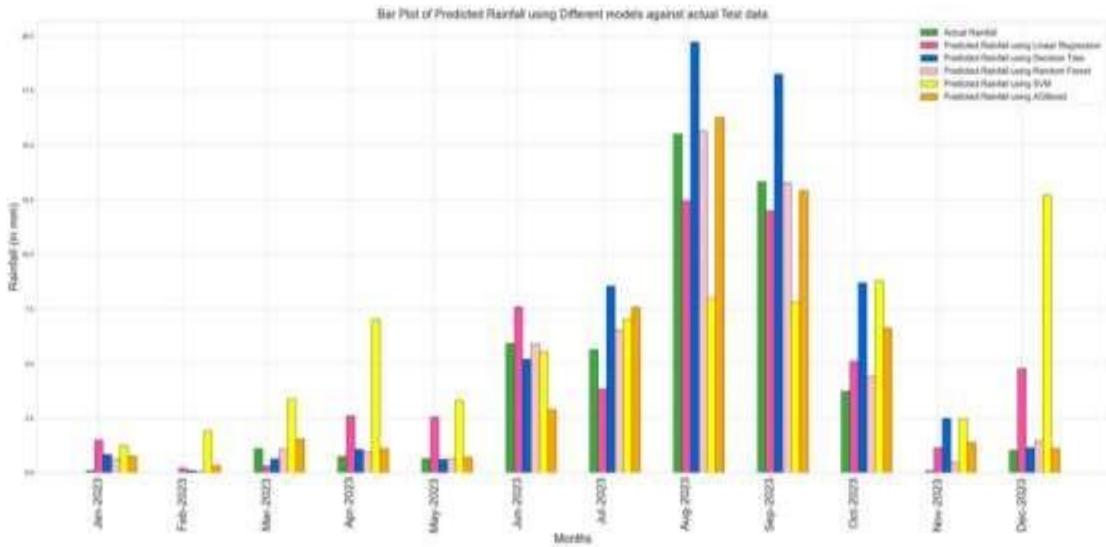

Fig. 9: Bar Plot of Predicted Rainfall of Ranchi using Different Base ML Models

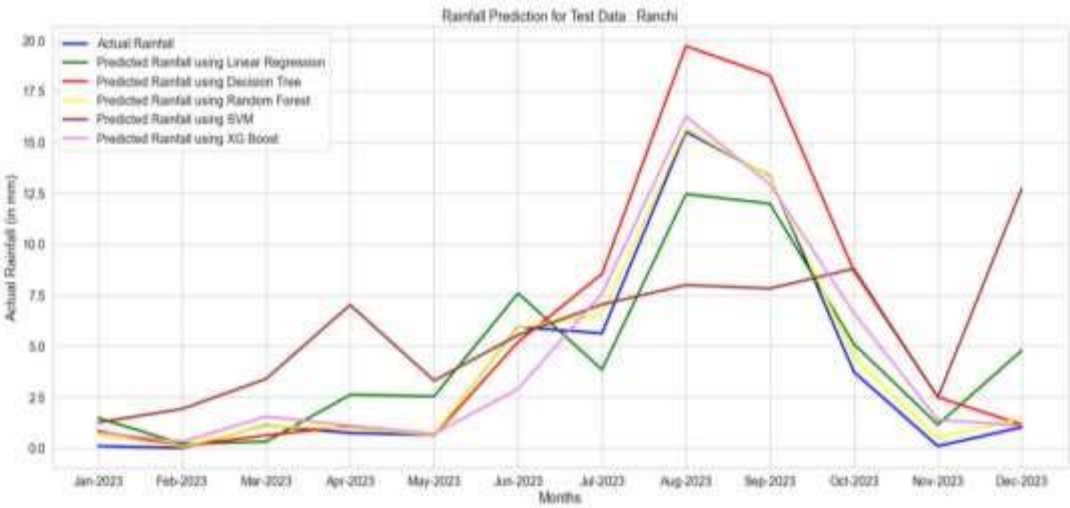

Fig. 10: Line Chart of Predicted Rainfall of Ranchi using Different Base ML Models

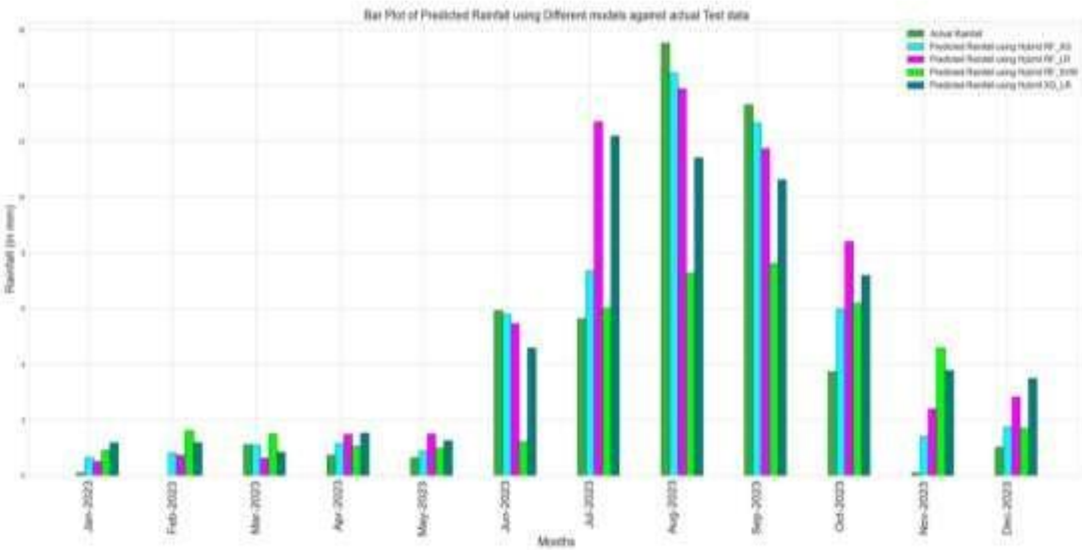

Fig. 11: Bar Plot of Predicted Rainfall of Ranchi using Different Hybrid ML Models

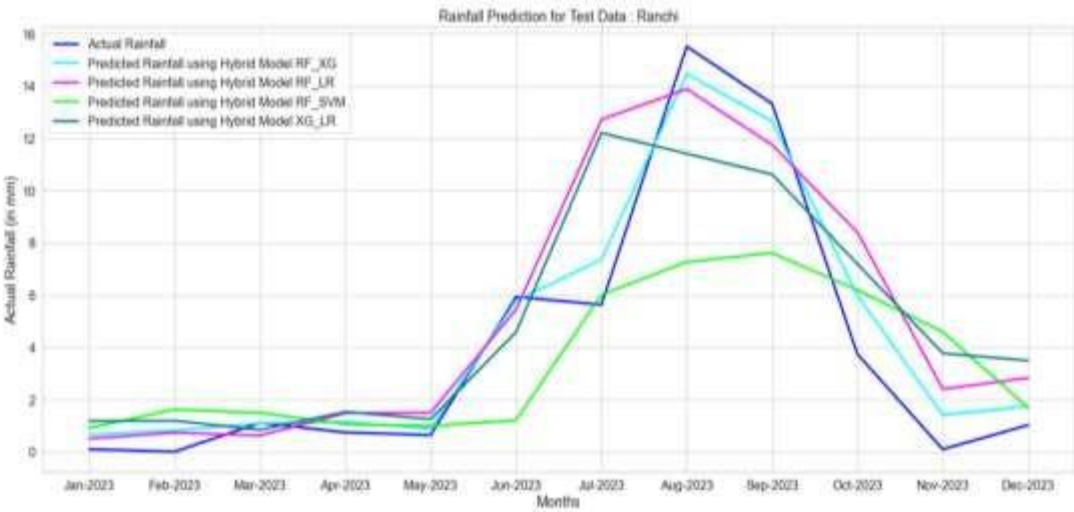

Fig. 12: Liner Chart of Predicted Rainfall of Ranchi using Different Hybrid ML Models

Another physical parameter i.e. Humidity has also been considered for discriminating above mentioned nine machine learning techniques. Parameters (accuracy and errors) calculated on the basis of data collected for this purpose have been tabled below. Different line and bar charts are also shown for analysis.

R^2 values of three districts for different models:

Models	Dhanbad	Jamshedpur	Ranchi
Linear Regression	0.704598	0.695421	0.729514
Decision Tree	0.776495	0.712498	0.796130
Random Forest	0.802346	0.819621	0.811302
Support Vector Machine	0.658943	0.713594	0.654792
XGBoost	0.806512	0.776403	0.797751
RF_XG	0.882315	0.879862	0.823624
RF_LR	0.761497	0.721492	0.682492
RF_SVM	0.732168	0.654211	0.556548
XG_LR	0.712956	0.735671	0.667845

Table 6: R^2 values of Humidity Prediction of three districts for different models

MAE values of three districts for different models

Model	Dhanbad	Jamshedpur	Ranchi
Linear Regression	0.0443742	0.0401299	0.250465
Decision Tree	0.0325529	0.0441993	0.236188
Random Forest	0.0384402	0.030738	0.17775
SVM	0.0477602	0.0365607	0.249205
XGBoost	0.0373696	0.0348547	0.190596
RF_XG	0.0329901	0.0312269	0.413514
RF_LR	0.040115	0.0343405	0.394807
RF_SVM	0.0398404	0.0361487	0.403523
XG_LR	0.0416032	0.0419395	0.372631

Table 7: MAE values of Humidity Prediction of three districts for different models

MSE values of three districts for different models

Model	Dhanbad	Jamshedpur	Ranchi
Linear Regression	0.00301095	0.00212551	0.0958637
Decision Tree	0.00201201	0.00280993	0.0989256
Random Forest	0.00338423	0.00137727	0.0546235
SVM	0.00361334	0.00173581	0.105507
XGBoost	0.00257217	0.00174045	0.0569895
RF_XG	0.00225441	0.00145957	0.228136
RF_LR	0.00278052	0.00192452	0.210015
RF_SVM	0.00279912	0.00193489	0.218799
XG_LR	0.00317932	0.00296414	0.203617

Table 8: MSE values of Humidity Prediction of three districts for different models

RMSE values of three districts for different models

Model	Dhanbad	Jamshedpur	Ranchi
Linear Regression	0.0548721	0.0461032	0.309619

Decision Tree	0.0448554	0.0530087	0.314524
Random Forest	0.0581741	0.0371115	0.233717
SVM	0.060111	0.0416631	0.324818
XGBoost	0.0507165	0.0417187	0.238725
RF_XG	0.0474806	0.0382043	0.477636
RF_LR	0.0527306	0.0438694	0.458274
RF_SVM	0.0529067	0.0439874	0.46776
XG_LR	0.0563855	0.0544439	0.451239

Table 9: RMSE values of Humidity Prediction of three districts for different models

(I B) Plots for humidity of Dhanbad district:

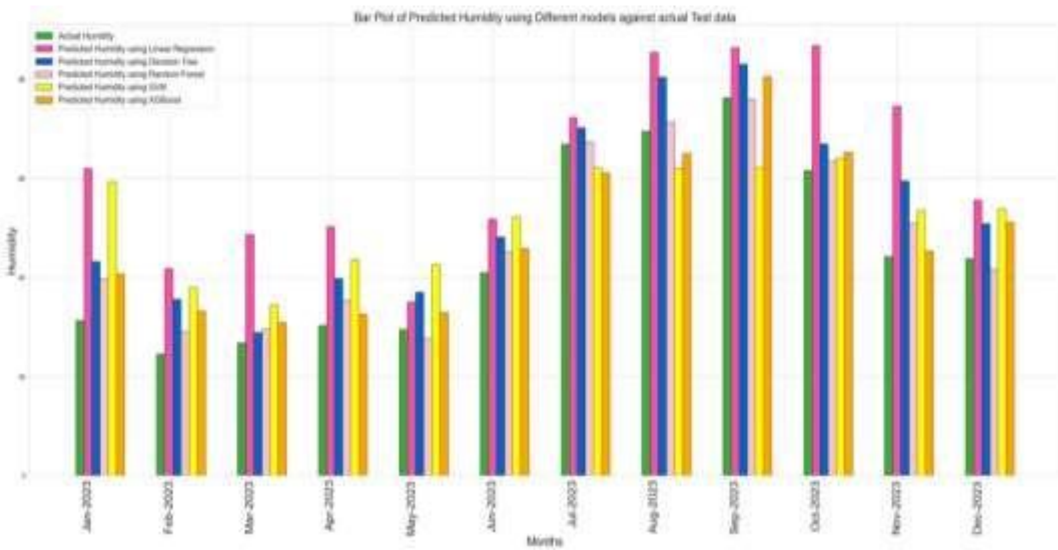

Fig. 13: Bar Plot of Predicted Humidity of Dhanbad using Different Base ML Models

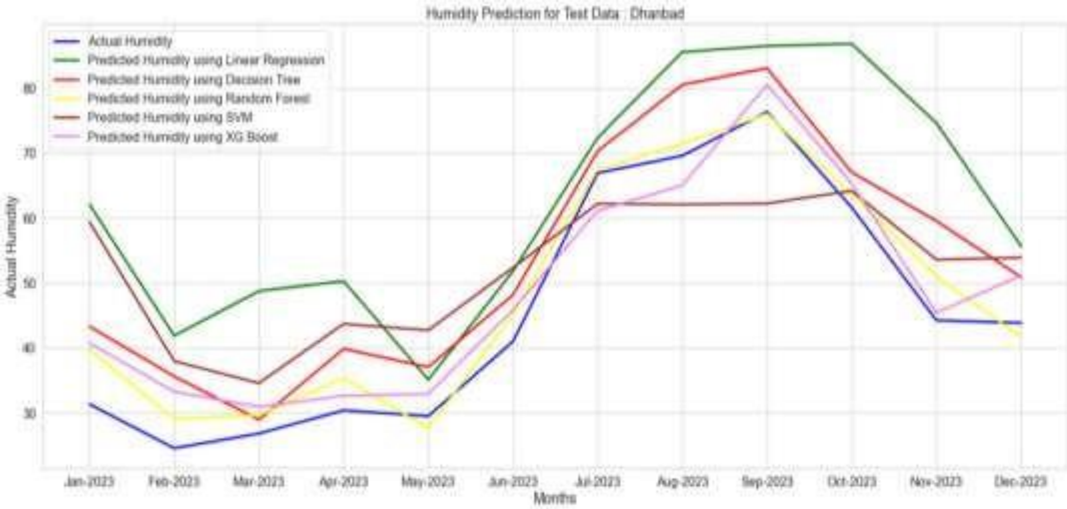

Fig. 14: Line Chart of Predicted Humidity of Dhanbad using Different Base ML Models

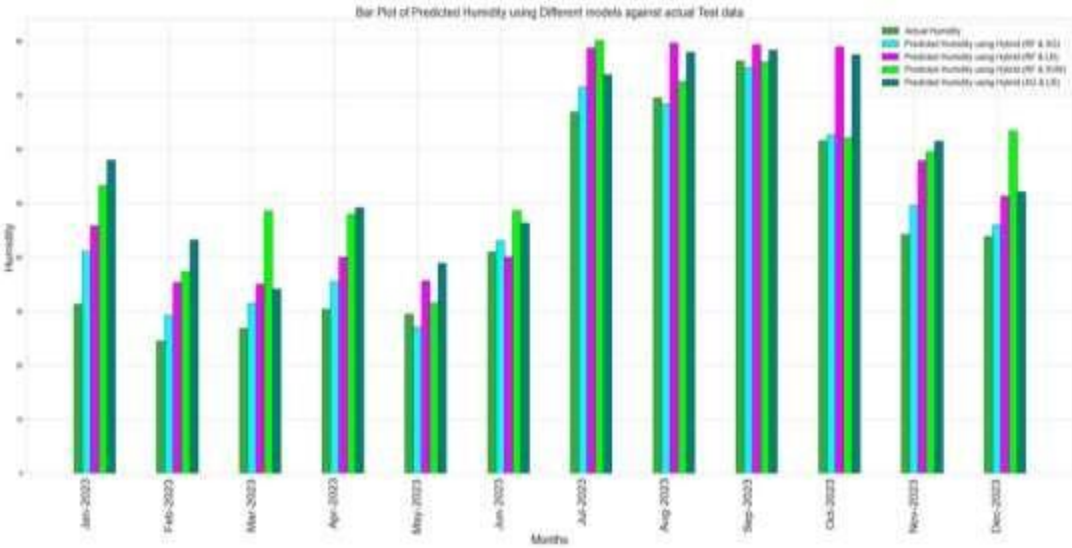

Fig. 15: Bar Plot of Predicted Humidity of Dhanbad using Different Hybrid ML Models

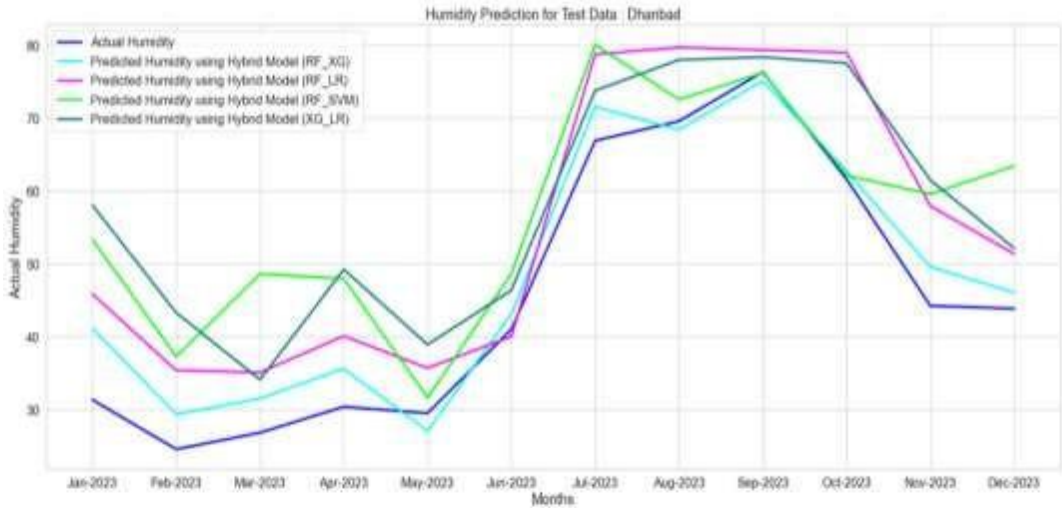

Fig. 16: Line Chart of Predicted Humidity of Dhanbad using Different Hybrid ML Models

(II B) Bar and Line charts for Humidity of Jamshedpur district

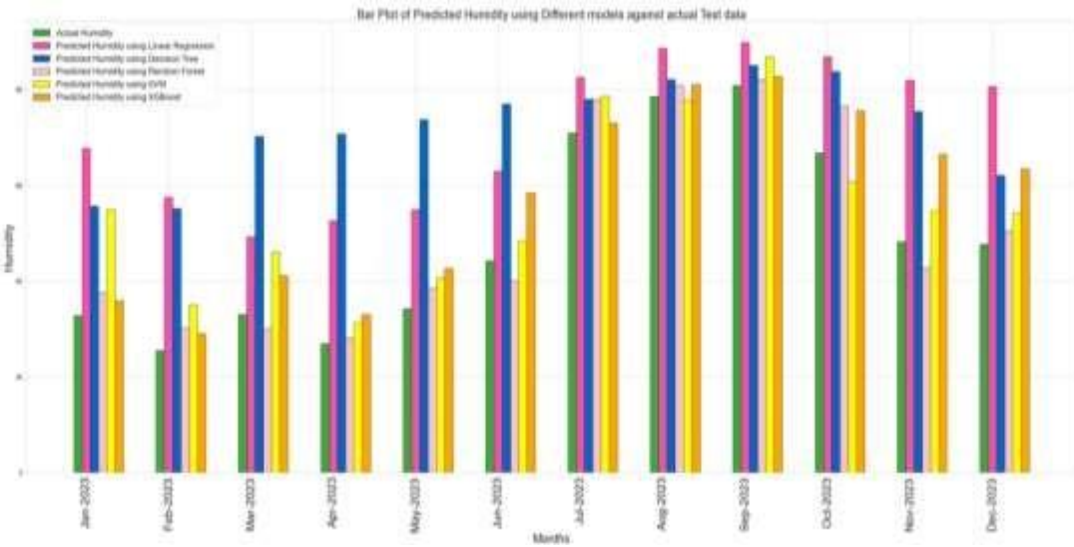

Fig. 17: Bar Plot of Predicted Humidity of Jamshedpur using Different Base ML Models

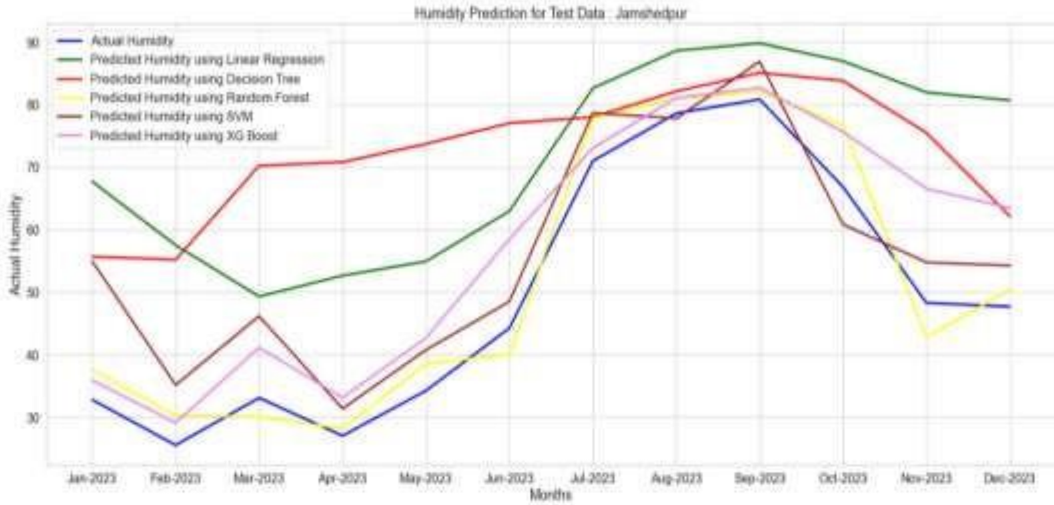

Fig. 18: Line Chart of Predicted Humidity of Jamshedpur using Different Base ML Models

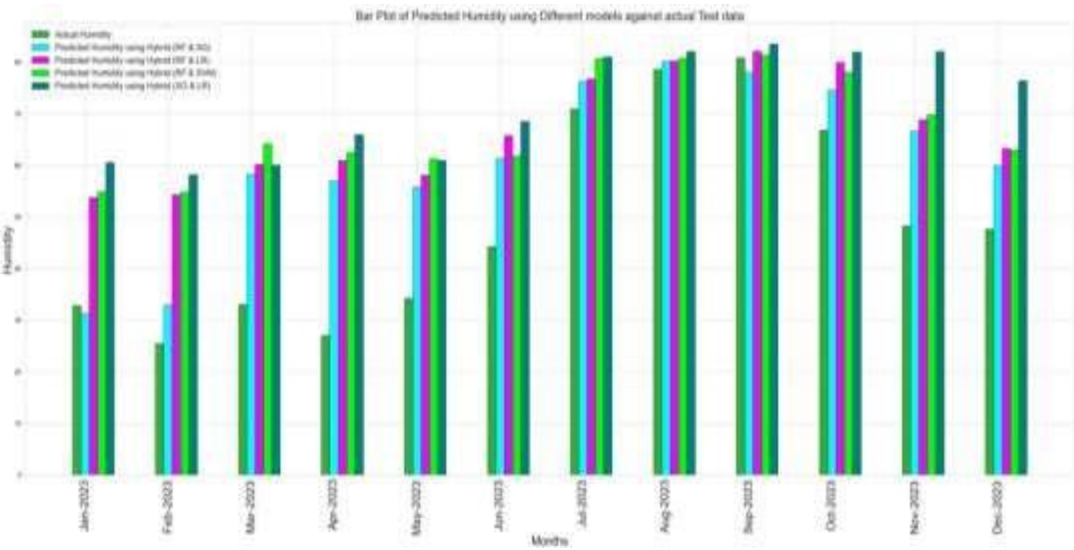

Fig. 19: Bar Plot of Predicted Humidity of Jamshedpur using Different Hybrid ML Models

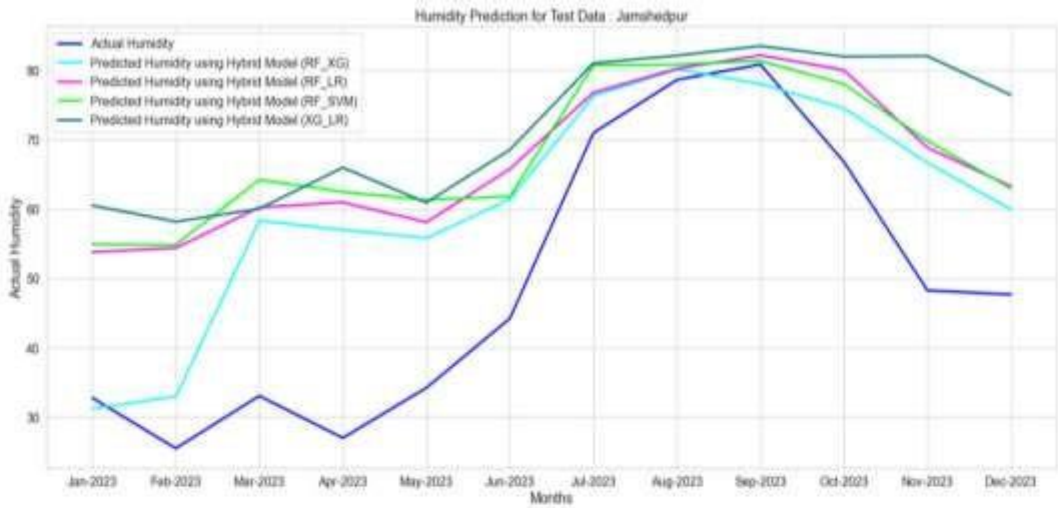

Fig. 20: Line Chart of Predicted Humidity of Jamshedpur using Different Hybrid ML Models

(III B) Following charts are shown for Ranchi district

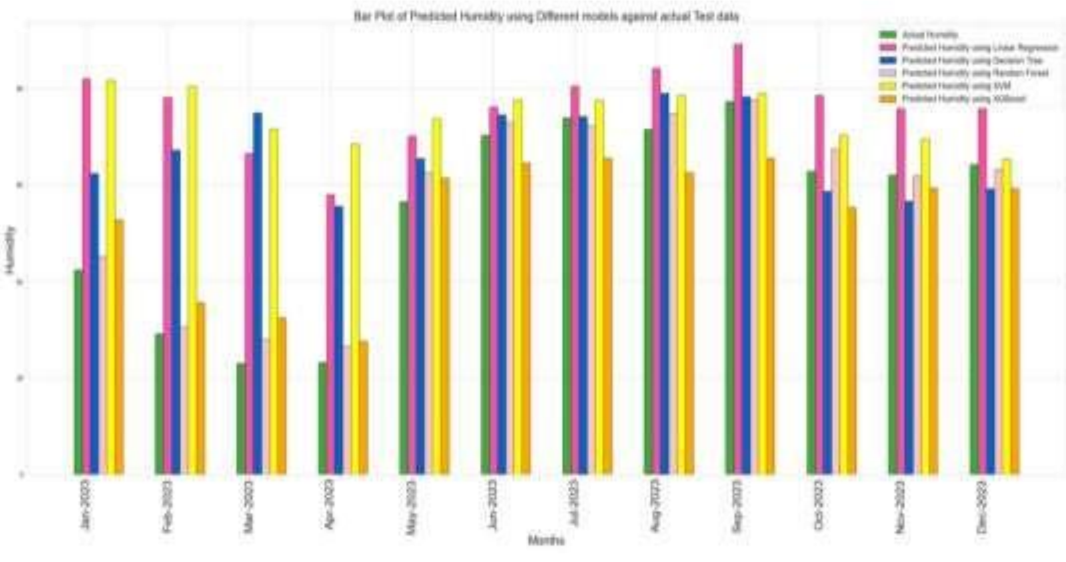

Fig. 21: Bar Plot of Predicted Humidity of Ranchi using Different Base ML Models

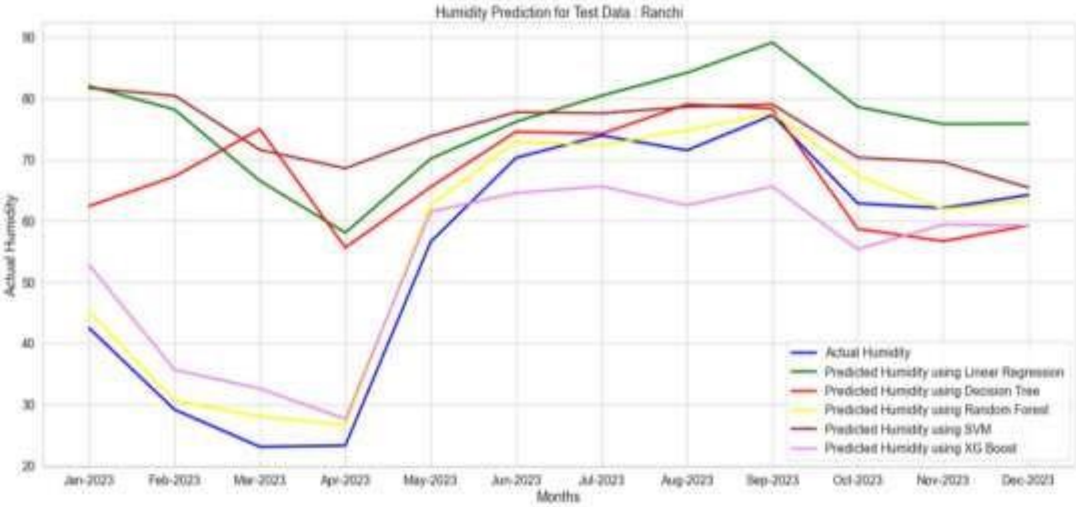

Fig. 22: Line Chart of Predicted Humidity of Ranchi using Different Base ML Models

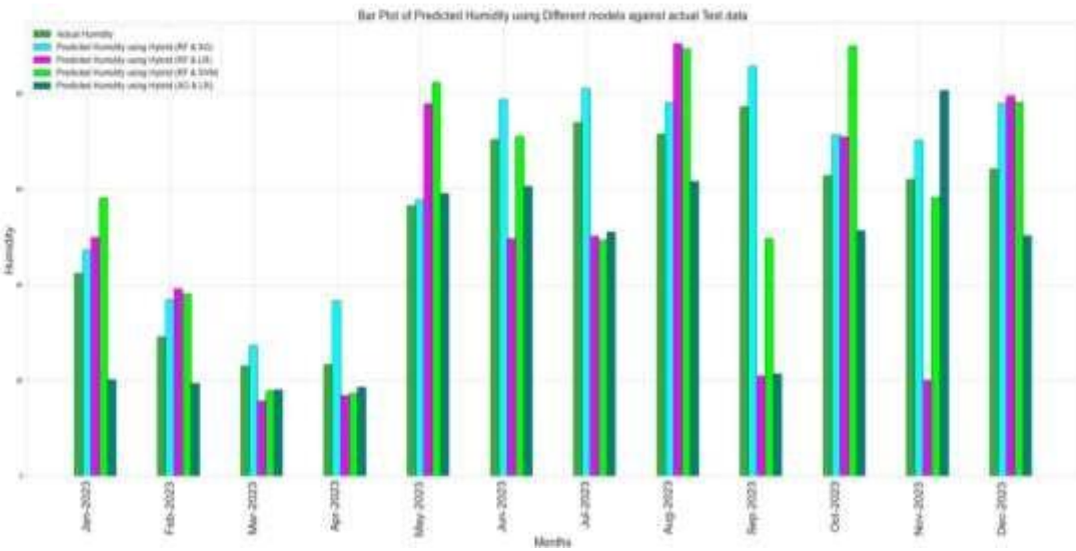

Fig. 23: Bar Plot of Predicted Humidity of Ranchi using Different Hybrid ML Models

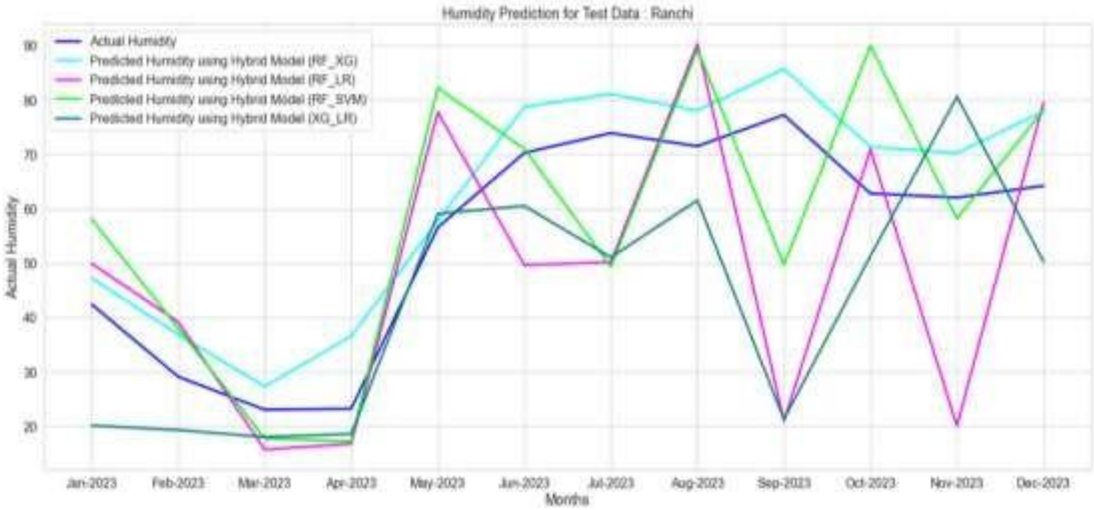

Fig. 24: Line Chart of Predicted Humidity of Ranchi using Different Hybrid ML Models

Analyzing the above data and charts we have found that Random Forest model is the most accurate basic model for rainfall prediction (~83%) whereas RF_XG hybrid model yields most accurate results (~87%) considering all the three districts. Interestingly, overall minimum error (considering MAE, MSE and RMSE) occurs in Random forest model (~4%) and RF_XG model (~4%). However, RF (~8%) and XGBoost (~9%) are found to be best suited for MAE only for the above three cities for humidity predictions. It is important to mention that the Random Forest model produces most accurate results (~81%) for humidity factor for the above three districts while RF_XG hybrid model shows the maximum accuracy (~86%) for humidity again. Further, overall error is minimum for RF (~6%), XGBoost (~6.5%) and RF_XG (~10%) models for all the three districts among the nine ML models. Moreover, XGBoost model comes next to Random Forest model in case of accuracy (~80% in case of humidity and ~ 81% for rainfall prediction) for aforesaid cities. It is to be noted that MAE is minimum for LR and DT models but these are not considered as preferred models because of less accuracy compared to RF and RF_XG models. It was also seen that for the months of July and August of 2024, rainfall and humidity had shown maximum accuracy with minimum error for the above mentioned districts for RF_XG model only.

5. Conclusions

It is clear from our study that Random Forest and XGBoost models are the most suited basic models for rainfall and humidity predictions for Dhanbad, Jamshedpur and Ranchi of Jharkhand state in the context of accuracy measurement. The same is true for RF_XG hybrid model. For overall minimum error Random forest model is best preferred for rainfall as well as humidity prediction whereas RF_XG hybrid model gives best results for both cases. From the above study it can be stated that Random Forest, XGBoost and RF_XG hybrid models should be treated as most important machine learning techniques to address the environmental issues like rainfall and humidity predictions in the context of three districts of Jharkhand.

Interestingly, RF_XG hybrid model is found to be most appropriate for accurate rainfall (~87%) and humidity (~ 86%) predictions among the all nine above mentioned machine learning models. The same is even true for July and August of 2024 also in the context of maximum accuracy and minimum absolute error.

References

1. I. Salehin, I. M. Talha, M. Mehedi Hasan, S. T. Dip, M. Saifuzzaman and N. N. Moon, "An Artificial Intelligence Based Rainfall Prediction Using LSTM and Neural Network," 2020 IEEE International Women in Engineering (WIE) Conference on Electrical and Computer Engineering (WIECON-ECE), Bhubaneswar, India, pp. 5-8, 2020.
2. Y. Dash, S. K. Mishra and B. K. Panigrahi. "Rainfall prediction for the Kerala state of India using artificial intelligence approaches." Computers & Electrical Engineering 70 (2018): 66-73.
3. C. Thirumalai, K. S. Harsha and M. L. Deepak. "Heuristic prediction of rainfall using machine learning techniques." International Conference on Trends in Electronics and Informatics (ICEI). IEEE, 2017.
4. A. Parmar, K. Mistree and M. Sompura. "Machine learning techniques for rainfall prediction: A review." International Conference on Innovations in information Embedded and Communication Systems. 2017.
5. H. A. Y. Ahmed and S. W. A. Mohamed, "Rainfall Prediction using Multiple Linear Regressions Model," International Conference on Computer, Control, Electrical, and Electronics Engineering (ICCCEEE), Khartoum, Sudan, 2021, pp. 1-5, 2020.
6. U. K. Sinha, K. Bandyopadhayay, and S. C. Dutta, “Prediction of PM10 AND CO Using Support Vector Regression To Analyses The Air Quality of Dhanbad,” Emerging Trends in Multidisciplinary Research and Innovations, p. 1, 2022.
7. Y. Radhika and M. Shashi, “Atmospheric Temperature Prediction using Support Vector Machines”. International Journal of Computer Theory and Engineering, 1(1), 55–58, 2009.
8. R. Kumar, “Decision tree for the weather forecasting”, International Journal of Computer Applications, 76(2), 31-34, 2013.
9. S. Dhamodaran, G. KipsonRoy, A. Kishor, and S. L. JanyShabu, “A Comparative Analysis of Rainfall Prediction Using Support Vector Machine and Random Forest”, Journal of Computational and Theoretical Nanoscience, 17(8), 3539-3542, 2020.
10. T. R. Prajwala, “A comparative study on decision tree and random forest using R tool”, International Journal of Advanced Research in Computer and Communication Engineering, 4(1), 196-199, 2015.
11. M. A. Sharma and J. B. Singh, “Comparative Study of rainfall forecasting models,” New York Science Journal, pp. 115-120, 2011.
12. M. T. Anwar, E. Winarno, W. Hadikurniawati and M. Novita, “Rainfall prediction using extreme gradient boosting”, Journal of Physics: Conference Series ,Vol. 1869, No. 1, p. 012078, 2021.
13. G. Singh and D. Kumar. "Hybrid Prediction Models for Rainfall Forecasting", 9th International Conference on Cloud Computing, Data Science & Engineering, IEEE, 2019.
14. P. G. Jaiswal, A. Dhote, A. Dubey and S. Kale, “A stacking ensemble learning model for rainfall prediction based on Indian climate”, 6th International Conference on Information Systems and Computer Networks (ISCON), pp. 1-6, IEEE, 2023.
15. https://dsp.imdpune.gov.in
16. https://www.wunderground.com/history/daily/in/ranchi/VERC/date/2020-4-30
17. https://www.worldweatheronline.com/ranchi-weather-history/jharkhand/in.aspx
18. Priya, Shilpi, Vashistha and V. Singh, “Time Series Analysis of Forecasting Indian Rainfall”, Intenational Journal of Innovations & Advancement in Computer Science, vol. 3, no. 1, pp. 66-69, 2014.

19. P. S. Dutta & H. Tahbilder, "Prediction of Rainfall Using Data mining Technique over Assam", Indian Journal of Computer Science and Engineering (IJCSE), vol. 5, no. 2, 2014.
20. I. Cohen, Y. Huang and J. Chen "Pearson correlation coefficient. Noise reduction in speech processing", pp 1-4 2009.
21. U. K. Sinha, K. Bandyopadhyay and S. C. Dutta, "Air Quality of Dhanbad During Pre-Lockdown, Lockdown And Post-Lockdown Periods" Journal of Emerging Technologies and Innovative Research, Volume 8, Issue 12, 2021.
22. D. Parbat and M. Chakraborty, "A python based support vector regression model for prediction of COVID19 cases in India", Chaos Solitons Fractals, vol. 138, p. 109942, 2020.
23. J. D. Hunter, "Matplotlib: A 2D graphics environment", Computer Science Engineering, vol. 9, no. 03, pp. 90–95, 2007.
24. M. Waskom, "Seaborn: statistical data visualization", The Journal of Open Source Software, vol. 6, no. 60, p. 3021, 2021.
25. F. Pedregosa, G. Varoquaux and A. Gramfort, "Scikit-learn: Machine Learning in Python", The Journal of Machine Learning Research, vol. 12, pp. 2825-2830, 2011.
26. A. Sapre and S. Vartak. "Scientific Computing and Data Analysis using NumPy and Pandas." International Research Journal of Engineering and Technology, vol. 7, pp. 1334-1346, 2020.
27. U. K. Sinha, K. Bandyopadhyay and S. C. Dutta, "Assessment of Air Pollutants of Dhanbad Using Machine Learning Techniques', International Journal of Intelligent Systems and Applications in Engineering, vol. 12(4), pp. 2649 –2659 (2024).
28. D. Chicco, M. J. Warrens, and G. Jurman, "The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation," PeerJ Computer Science, vol. 7, pp. e623, 2021.