Applied Computer Science, vol. 18, no. 4, pp. 16–27
doi: 10.35784/acs-2022-26

Submitted: 2022-06-24 | Revised: 2022-09-25 | Accepted: 2022-10-01

Keywords: Meteorological data, M5 model tree, Linear model functions, Gradient boosting, Logistic Model trees

Sheikh Amir FAYAZ [0000-0001-6606-0864]*, Majid ZAMAN [0000-0003-1070-8195]**
Muheet Ahmed BUTT [0000-0002-8059-0180]*, Sameer KAUL [0000-0003-0911-0073]*

HOW MACHINE LEARNING ALGORITHMS ARE USED IN METEOROLOGICAL DATA CLASSIFICATION: A COMPARATIVE APPROACH BETWEEN DT, LMT, M5-MT, GRADIENT BOOSTING AND GWLM-NARX MODELS

Abstract

Rainfall prediction is one of the most challenging task faced by researchers over the years. Many machine learning and AI based algorithms have been implemented on different datasets for better prediction purposes, but there is not a single solution which perfectly predicts the rainfall. Accurate prediction still remains a question to researchers. We offer a machine learning-based comparison evaluation of rainfall models for Kashmir province. Both local geographic features and the time horizon has influence on weather forecasting. Decision trees, Logistic Model Trees (LMT), and M5 model trees are examples of predictive models based on algorithms. GWLM-NARX, Gradient Boosting, and other techniques were investigated. Weather predictors measured from three major meteorological stations in the Kashmir area of the UT of J&K, India, were utilized in the models. We compared the proposed models based on their accuracy, kappa, interpretability, and other statistics, as well as the significance of the predictors utilized. On the original dataset, the DT model delivers an accuracy of 80.12 percent, followed by the LMT and Gradient boosting models, which produce accuracy of 87.23 percent and 87.51 percent, respectively. Furthermore, when continuous data was used in the M5-MT and GWLM-NARX models, the NARX model performed better, with mean squared error (MSE) and regression value (R) predictions of 3.12 percent and 0.9899 percent in training, 0.144 percent and 0.9936 percent in validation, and 0.311 percent and 0.9988 percent in testing.

* Department of Computer Sciences, University of Kashmir, J&K, India, skh.amir88@gmail.com, ermuheet@gmail.com, kaulsameer78@gmail.com
** Directorate of IT & SS, University of Kashmir, J&K, India, zamanmajid@gmail.com

1. INTRODUCTION

Rainfall forecasting is useful in preventing floods, which saves lives and property. In reality, it aids in the management of the water supply. Rainfall data from the past has aided farmers in better managing their crops, resulting in increased economic growth in the country. Weather forecasting is difficult for climatological scientists because of the variability in rainfall timing and volume. Precision rainfall modeling is becoming one of the most pressing topics in hydrology, as early warning of extreme weather can help prevent natural disasters and losses if forecasts are made quickly and accurately. Forecasting is one of the most difficult problems for experts in a variety of domains, including meteorological data mining (Yang et al., 2007), and statistical forecasting (Pucheta et al., 2009). A typical question in such situations is how to analyze the past and make future forecasts. Even over a short period, the characteristics needed to predict rainfall are extremely complicated and delicate.

Traditional approaches that apply statistical techniques to examine the link between rainfall, geographic coordinates, and other atmospheric parameters have been used to forecast rainfall for years (like pressure, temperature, wind speed, and humidity). Rainfall, on the other hand, is difficult to anticipate due to its complexity, such as its nonlinearity (Wu & Chau, 2013). As a result, Singular Spectrum Analysis, Empirical Mode Decomposition, and Wavelet analysis, among other techniques, have been used to minimize non-linearity (Xiang et al., 2018). However, the mathematical and statistical models used need a lot of computational power (Singh & Borah, 2013; Singh et al., 2015) and can be time-consuming with little impact.

Weather prediction has gotten more successful in recent years in resolving a topic that has perplexed mankind for centuries, yet precise and timely weather forecasting remains a problem for scientists. Weather and weather forecasting are two subjects that practically everyone is interested in. Weather forecasting may therefore be anticipated utilizing many applicable machine learning methodologies in current era where machine-learning techniques are applied in every industry.

On various raw datasets in various places, numerous machine-learning algorithms have been constructed to forecast the amount of rainfall. In one of our studies (Kaul et al., 2023), comparisons between DT, DDT, and RF were done. We intended to apply more algorithms in order to select the best one in terms of accuracy measurement based on this comparison. This encourages us to work with the same dataset and assess the efficacy of different methods. This research compares the performance of models based on Logistic Model Tree, Gradient Boosting, GWLM-NARX and M5 Model Tree Networks to that of the original decision tree model and a projected model that will be the outcome of using an Automated Machine Learning tool (Mohd, Butt & Baba, 2022).

This paper is organised as follows: Section 1 provides a brief explanation of rainfall and its influence on agriculture, the environment, and the different machine learning techniques used to predict rainfall in JK. Section 2 is a quick overview of the literature. In section 3, numerous geographical setting and the satellite map photos have been evaluated which determines the climate of JK. Section 4 contains a description of the content and dataset, while sections 5 and 6 outline the algorithmic framework and technique. Section 7 outlines the experimental setup and assessment of several machine-learning methods. Finally, section 8 concludes the report with recommendations for the future.

2. REVIEW OF LITERATURE

The majority of weather prediction research is done using numerical approaches. The primary focus of this study will be on prior research based on various classical and ensemble machine learning algorithms utilized in rainfall prediction. Decision trees (DT), logistic model trees (LMT), M5 model trees (M5-MT), Gradient Boosting, and GWLM-NARX models are among these techniques.

Authors (Adnan, 2021), provided a comparison analysis of four machine learning techniques for rainfall modelling. The capabilities of OPELM, MARS, and the M5 model tree in daily precipitation modelling is the topic of this research. It was discovered that accuracy improves dramatically, with RMSE and MAE improvements of more than 90% in most cases. MARS-K also surpasses the other options tested in this study.

A research on groundwater level predictions using MARS and M5 Model Tree machine learning techniques was proposed by (Rezaie-balf et al., 2017). The data spans almost ten years, from August 1996 to July 2006. Validation of the models is done using the parameters utilized in this study. Validation was done using statistical performance assessment parameters such as RMSE, NNSE, and Coefficient of Determination.

Mohd et al. (Mohd, Butt & Baba 2020) developed a time series prediction model use the GWLM-NARX model. With rainfall data from the preceding period as input and results derived using the NARX model's GWLM algorithm, this model was employed as an adaptive forecast model.

(Fayaz, Zaman & Butt, 2021b) use a stepwise machine learning technique to estimate rainfall in India's Kashmir area. They used an LMT technique in their research, where the leaf node predicts model functions using logistic-regression approaches. The data for their analysis came from the Indian meteorological department in Pune, and it covered the years 2012 to 2017. Season, temperature at various intervals, humidity from 12 a.m. – 3 p.m., and rainfall were some of the variables studied. The study finishes with a comparison analysis in which the performance of several traditional and ensemble techniques is compared to that of the LMT, demonstrating that the LMT's accuracy measure is far superior to the other models utilized in the study.

(Fayaz, Kaul, Zaman & Butt, 2022) again used the same labelled data set as was used in (Zaman & Butt, 2012; Fayaz, Zaman & Butt, 2021a). The use of an ensemble distributed decision tree for rainfall prediction is defined in this work. The dataset was separated into three portions in this analysis depending on the station id. The performance of each decision tree was calculated after each decision tree was formed. A final accuracy was estimated based on the voting technique of the three smaller decision trees. The resulting accuracy was then compared to the accuracy of the original decision tree. The accuracy of the distributed decision tree was found to be much lower than the accuracy of the original decision tree.

Kaul et al. (2022) performs the comparative study on same set of data used in (Zaman & Butt, 2012; Fayaz, Zaman & Butt, 2022b, 2022c). In this research a comparsion has been made between Decision tree, Distributed decision tree (DDT) and Random forest (RF) on the geographical dataset. This research concluded that the decision tree performs better accuracy results in comparsion to the DDT and RF.

Furthermore, various network models for monthly rainfall rate forecasting and climate change were proposed in (Fayaz, Zaman & Butt, 2021c, 2022a) and the proposed models' performance was found to be extremely effective. The findings of the trial indicate that the accuracy rates will improve.

Since then, numerous writers have used the tabular dataset to estimate rainfall using a range of classical and ensembled methodologies, and some of them are described in this paper. From the literature, we may deduce that no one technique outperforms others on diverse types of datasets.

3. GEOGRAPHICAL SETTING AND CLIMATE OF J&K

In the state summary, the climatology of the UT of "Jammu and Kashmir" is described in terms of various meteorological parameters such as temperature, rainfall, rainfall variability, pressure, winds, relative humidity, clouds, weather hazards, and so on, followed by a detailed description of the climate of each district taking geographical and topographic characteristics into account as shown below (Figure 1).

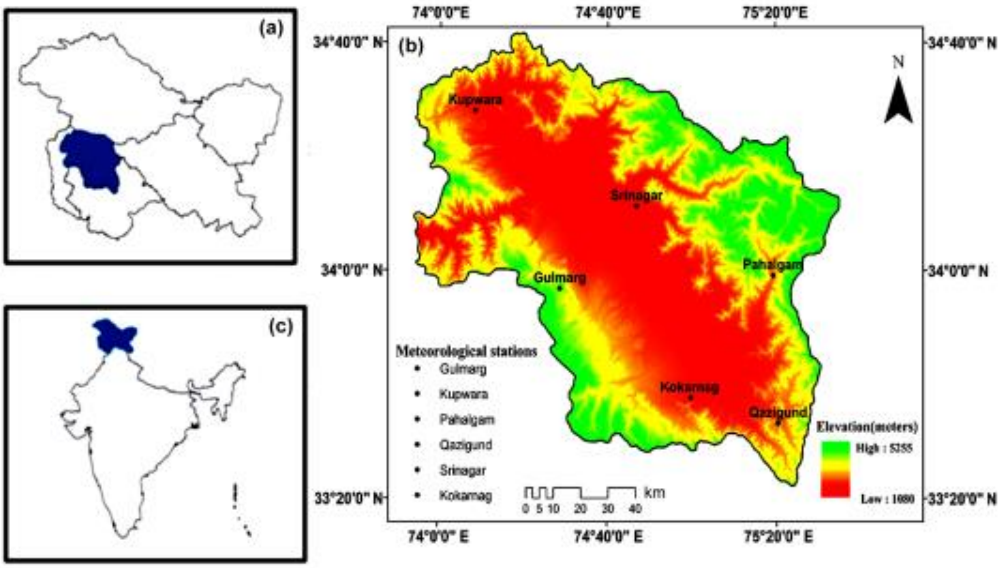

Fig. 1. Orographic and Geographical setting of the Kashmir valley (Zaz, 2019)

The Indian Union territory of Jammu and Kashmir, which is located in the Himalayas, may be divided into two climatic regions: Jammu and Kashmir. The winter capital is Jammu, whereas the summer capital is Srinagar. The greatest time to visit the 'Kashmir' region is between April and October, when the weather is nice in the summer and frigid in the winter. During the summer, the valley is blanketed in blooms, and the orchards are overflowing with fruit. Winter, from October to March, is the greatest season to explore the 'Jammu' region, with maximum temperatures about 18 degrees Celsius and minimum temperatures as low as 4 degrees Celsius on some days. The weather in Kashmir is nice, with temperatures ranging from 14 to 30 degrees Celsius. Although some days might be a little hot, the evenings are typically comfortable.

4. DATASET DESCRIPTION

Kashmir is classified as a temperate zone. As a result, the data were separated into four seasons to compute seasonal means: winter (December to February), spring (March to May), summer (June to August), and autumn (September to November). The seasonal rainfall in centimeters for each season is depicted in the graphs below figure (figure 2) (winter, spring, autumn, and summer) (Zaz et al., 2019).

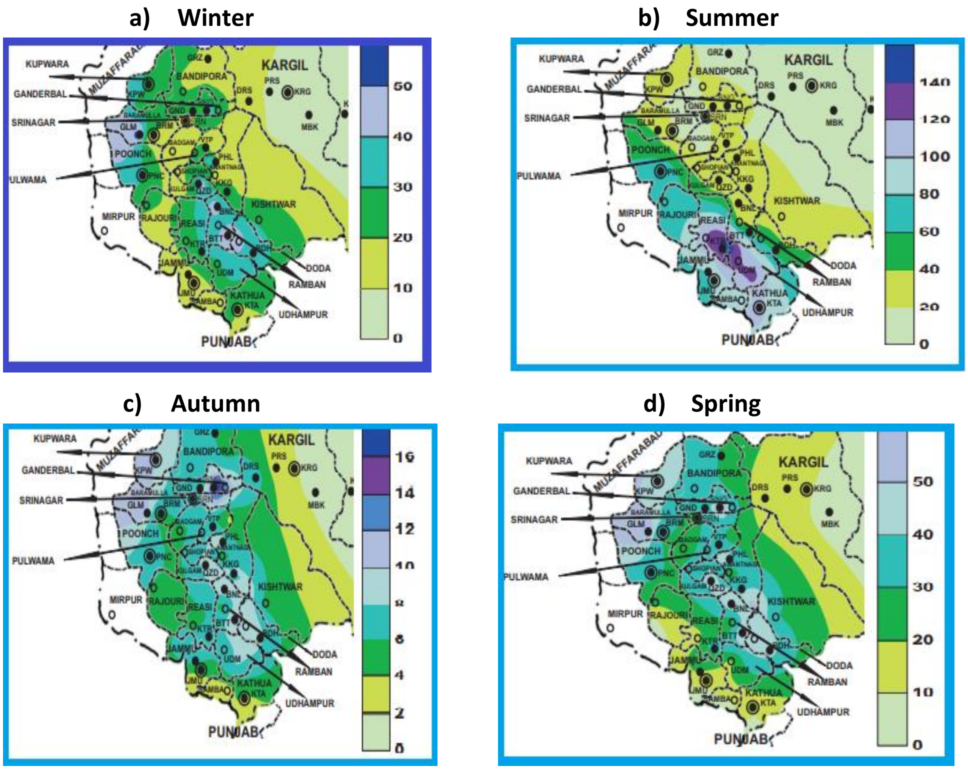

Fig. 2. Average Seasonal rainfall (cm) for each season in Kashmir province

For all six sites, the India Meteorological Department supplied data for five years (2012–2017) of daily precipitation, maximum and lowest temperatures, and humidity measurements at various time intervals. Central zone (34.0837° N, 74.7973° E), North zone (34.0837° N, 74.7973° E), and South zone (34.0484° N, 74.3805° E) are the three primary zones of Kashmir province. Figure 3 shows the overall structure of the dataset used in this study (Zaman & Butt, 2012).

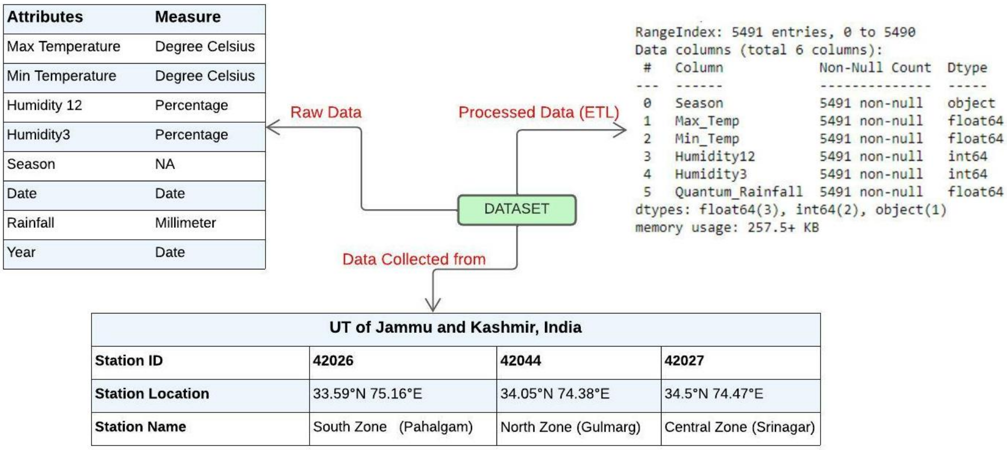

Fig. 3. Meteorological dataset of Kashmir province

To establish long-term trends and turning points of meteorological parameters with statistical significance, statistical tests such as kurtosis, cumulative deviation, and t test were used. The statistics of the data used in this study is shown in a tabular form (Table 1 and Table 2).

Tab. 1. Statistical properties of data

Attributes	t-test	Mean Difference	CI (Lower Bound)	Std. Dev	Skewness	CI (Upper Bound)	Kurtosis
Max Temp (°C)	151.8693	18.0409	17.808	8.80	-0.24	18.2738	-0.86
Min Temp (°C)	63.2285	6.3435	6.1469	7.43	0.02	6.5402	-0.84
Humid12 (%)	247.434	60.2723	59.7947	18.0	0.21	60.7498	-0.73
Humid3 (%)	396.5277	75.6416	75.2676	14.1	-0.76	76.0156	0.40
Rf (mm)	22.5342	2.7579	2.518	9.07	7.75	2.9979	99.4

Tab. 2. Correlation matrix and P value of the geographical attributes

First Column	Second Column	Correlation value	P value
Max	Min	0.879289995	0
Max	Hum12	-0.301475201	9.52E-116
Max	Hum3	-0.259730257	2.34E-85
Max	Rf	-0.190566402	4.56E-46
Min	Hum12	-0.141257364	7.16E-26
Min	Hum3	-0.107901977	1.08E-15
Min	Rf	-0.035400555	0.008704384
Hum12	Hum3	0.961517542	0
Hum12	Rf	0.008394317	0.534008313
Hum3	Rf	0.001847154	0.891152955

We employed around 6000 instances of meteorological data in this work, which included five parameters: humidity at 12 a.m., humidity at 3 p.m., maximum temperature, minimum temperature, and one goal parameter rainfall, which determines the amount of rain. Several metrics are employed in the model training, validation, and hyperparameter search to measure the validity of the machine learning algorithm's predictions. The best measure is determined totally by the task at hand. In binary and multiclass classification tasks, accuracy and kappa metrics are among the most often employed metrics. The proportion of properly categorized observations in relation to the total number of predictions is called accuracy, and kappa is the normalized accuracy value in relation to the predicted percentage of hits (Fayaz, Zaman, Kaul & Butt 2022).

5. ALGORITHMS FRAMEWORK

Several algorithms were tested for goodness of fit in this research. These algorithms include:

1. Decision trees: The decision tree (DT) is a data aggregation approach proposed in (Zaman & Butt, 2012) that is regarded as one of the most precise general-purpose tools. It entails making many judgments on samples from a data set obtained by random sampling with replacement (Banday et al., 2022).
2. Logistic Model Trees: At the leaves, Logistic Model Trees (lmt) mix model trees with logistic regression procedures. The logistic regression models that may identify important features in the data are built via a stage wise fitting approach (Aguasca-Colomo, Castellanos-Nieves & Méndez, 2019).
3. M5 Model trees: The M5 model tree is made up of two steps: a traditional decision tree and a linear regression function. To begin, the regression tree is constructed using the decision tree induction procedure. The standard deviation at each node will be determined to assess the predicted reduction in error for the splitting criterion. This node splitting in M5 will continue until there are very few instances left. Second, after constructing the normal regression tree, internal sub nodes are pruned and replaced with the regression plane rather than constant values (Niu & Zhang, 2015).

Gradient boosting: Gradient boosting is a flow process in which the original data used for prediction is given to the base model, which performs the first prediction. The error will be computed once this predicted output is compared to the actual output. The next decision tree is created based on the error, with only independent parameters considered and residuals for target parameters employed (error) (Barrera-Animas, 2022).

Several frameworks are now utilized to deal with predictive models, such as python packages such as TensorFlow, pytorch, Keras, or Scikit-learn. We utilized the caret software (Classification and Regression Training) in this study. Caret is an interface that combines many machine-learning tools into a single framework, making data preparation, training, optimization, and validation of predictive models easier, as well as native support for parallel computations.

GWLN-NRAX model: The GWLM method is a hybridization of the grey wolf optimization (GWO) and levenberg-marqueret (LM) algorithms with nonlinear autoregressive model (NARX) that is utilized for effective and adaptable rainfall forecast (Fayaz, Zaman & Butt, 2021c).

6. METHODOLOGY

In this paper, we show how the model is applied to meteorological data of kashmir province utilizing three traditional and ensemble approaches which includes Decision tree (DT), logistic model trees (LMT), Gradient Boost (GBoost), GWLM-NARX and M5 model trees (M5-MT). The datasets in these methods have been separated into (70-15-15) percent training, validation and test sets, respectively. This data splitting was done in Python using the sklearn split model. All the above three models follow the same basic approach in the implementation processes and we have provided a brief discussion of the machine learning strategies utilized in the prediction model development. These different machine learning methods are implemented on same set of geographical data of Kashmir province which includes different parameters like humidity at different intervals, temperatures and seasons and most importantly the target parameters rainfall.

Our approach includes the following steps, which result in an adequate prediction model: 1. training (apply a machine-learning algorithm to the training data set so that the model learns), 2. validation (predict the error of a statistical model with new data), and 3. prediction approach. Figure 4 shows the flowchart for the proposed methodologies (Aftab, Butt & Zaman, 2018; Afolayan, Ojokoh & Falaki, 2016).

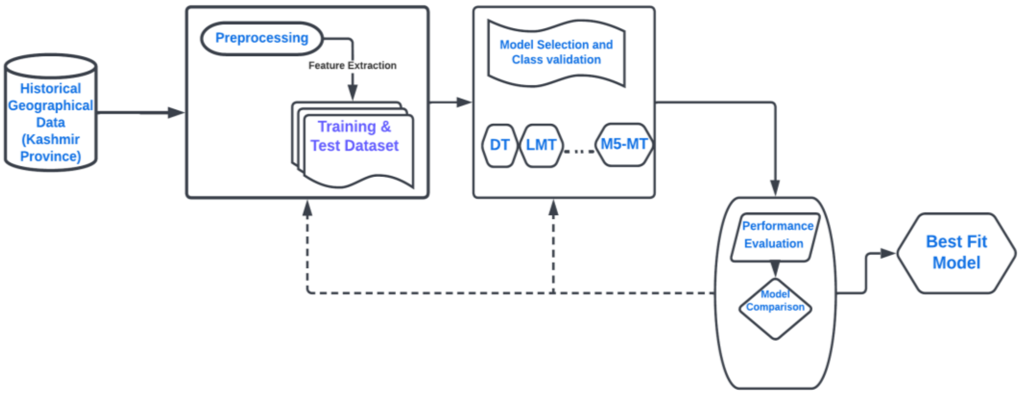

Fig. 4. Proposed Methodology

7. EXPERIMENTAL SETUP

In this work, the researchers used cutting-edge technologies on geographical datasets to test the algorithm with the greatest overall performance and accuracy. When the accuracy measure of the LMT and Gradient Boost (GBoost) were compared, it was shown that there is some increase in performance when compared to the original decision tree. Also, in the case of NARX and M5-MT models, the performance measure appears to be similar with lower mean absolute error rate (MAE), but the key difficulty with the methods is that training the data takes a long time because these models deal with continuous data streams. Table 3 displays a snapshot of the results, including accuracy, precision, recall values, and several other computations. The ML techniques chosen enable the development of a prediction model capable of representing the patterns contained in the training data set and generalizing them to new findings as shown in below table.

Tab. 3. Algortihms used with various statistical measures

Algorithms	Accuracy	Error	No of Classificati on Rules	Cohen Kappa	Precision
Original Decision Tree	80.12%	19.87%	51	0.456	0.812
LMT	87.23%	12.77%	10	0.102	0.893
Gradient Boost	87.51%	12.49%	–	0.073	0.914
M5-MT	R2= 0.478; MAE = 1.689; MSE = 6.726; RMSE = 2.593; MSD = 0.844				
GWLM-NARX	Regression (R) Testing: 0.9988%; Validation: 0.9936%; Training: 0.9899% MSE Testing: 0.311% ; Validation: 0.144%; Training: 3.12%				

The following graphical depiction (Figure 5) of geographical data from the Kashmir region aids in the simple visualization of the conclusions obtained. Other approaches are also highly efficient, but they require a huge amount of training data to train on in order to predict a relatively little amount of test data. Figure 5 is the visual representation of the table 3 where the maximum precision value of gradient boosting can be seen as 0.914 and highest accuracy level of 87.51%. Furthermore the accuracy level of LMT and GB remains head to head and remains much better than Decision tree.

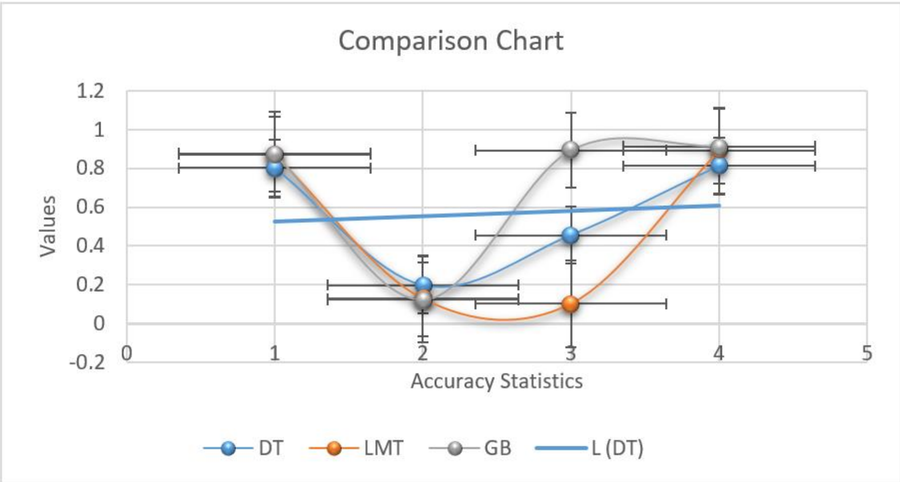

Fig. 5. Line graph: defines the accuracy statistics of each algorithm

Determining the optimum model is not an easy task; there are many of approaches, each with its own set of features and different parameters that must be changed. An unbalance in the probabilities of the observed classes is one of the classification difficulties that can have a major detrimental influence on the model's efficiency. An unbalanced data set is one in which the categorization categories are not roughly evenly represented. A potential solution to such class imbalance is to reconstruct the original the training data in a way that mitigates these issues.

To prevent having a detrimental influence on the prediction models, imbalances in the frequencies of the observed classes were handled. Because of the random sample, the minority class is the same size as the majority class. Predictors were preprocessed in order to interact with the ML algorithm or increase their performance. In order to predict how the model would perform with unknown data, the dataset was partitioned into 80-20 training and testing ratio respectively (Dhamodaran & Lakshmi, 2021; Altaf, Butt & Zaman, 2022).

8. CONCLUSIONS AND FUTURE SCOPE

In this paper, we utilized and compared various well-known ML systems for rainfall prediction. This research compares the performance of various machine-learning machines and discusses potential applications. This proposal includes prediction models that are both accurate and easy to understand. This research is driven by the need to simplify and improve the process of rainfall prediction, as well as to solve the difficulties that existing solutions involve. As a result, the following are the primary contributions of this paper: a) Generating and comparing rainfall forecast models using various machine learning approaches; and b) Determining if the combination of local meteorological factors, and the algorithms utilized affects the predictive algorithm's accuracy.

The total accuracies achieved by the original Decision tree, LMT, and Gradient boosting models are 81.12%, 87.23%, and 87.51%, respectively. As a result, we can conclude that the LMT and Gradient Boosting models show significant improvement in predicting the class labels as compared to original decision tree, and they show head-to-head accuracy and prediction and can thus be considered as promising techniques for the prediction of rainfall in temperate zones such as Kashmir province. Furthermore, in case of M5-MT and GWLM-NARX models where continuous data was taken into consideration, NARX model was proven to be better with mean squared error (MSE) and regression value (R) predictions of 3.12 percent and 0.9899 percent in training, 0.144 percent and 0.9936 percent in validation, and 0.311 percent and 0.9988 percent in testing.

Since we concluded that GWLM-NARX performs better results based on the geographical dataset of the Kashmir province, the performance on other threshold datasets, such as academic data, health data, and other geographic data, has not been determined. It will be a future suggestion of this study to check the performance of these implemented algorithms on a wide range of datasets.

Conflicts of Interest

The authors have no conflicts of interest to declare.

REFERENCES

Adnan, R. M., Petroselli, A., Heddam, S., Santos, C. A. G., & Kisi, O. (2021). Comparison of different methodologies for rainfall–runoff modeling: machine learning vs conceptual approach. *Natural Hazards*, *105*(3), 2987–3011.

Afolayan, H. A., Ojokoh, B. A., & Falaki, S. O. (2016). Comparative analysis of rainfall prediction models using neural network and fuzzy logic. *International Journal of Soft Computing and Engineering*, *5*(6), 4–7.

Aftab, S., Ahmad, M., Hameed, N., Bashir, M. S., Ali, I., & Nawaz, Z. (2018). Rainfall prediction using data mining techniques: A systematic literature review. *International journal of advanced computer science and applications*, *9*(5), 143–150.

Aguasca-Colomo, R., Castellanos-Nieves, D., & Méndez, M. (2019). Comparative Analysis of Rainfall Prediction Models Using Machine Learning in Islands with Complex Orography: Tenerife Island. *Applied Sciences*, *9*(22), 4931. https://doi.org/10.3390/app9224931

Altaf, I., Butt, M. A., & Zaman, M. (2021). A Pragmatic Comparison of Supervised Machine Learning Classifiers for Disease Diagnosis. In *2021 Third International Conference on Inventive Research in Computing Applications (ICIRCA)* (pp. 1515–1520). IEEE. https://doi.org/10.1109/ICIRCA51532.2021.9544582

Banday, I.R., Zaman, M., Quadri, S.M.K., Fayaz, S.A., Butt, M.A. (2022). Big data in academia: A proposed framework for improving students performance. Revue d'Intelligence Artificielle, Vol. 36, No. 4, pp. 589–595. https://doi.org/10.18280/ria.360411

Barrera–Animas, A. Y., Oyedele, L. O., Bilal, M., Akinosho, T. D., Delgado, J. M. D., & Akanbi, L. A. (2022). Rainfall prediction: A comparative analysis of modern machine learning algorithms for time-series forecasting. *Machine Learning with Applications*, *7*, 100204. https://doi.org/10.1016/j.mlwa.2021.100204

Dhamodaran, S., & Lakshmi, M. (2021). Comparative analysis of spatial interpolation with climatic changes using inverse distance method. *Journal of Ambient Intelligence and Humanized Computing*, *12*(6), 6725–6734. https://doi.org/10.1007/s12652-020-02296-1

Fayaz, S. A., Kaul, S., Zaman, M., & Butt, M. A. (2022). An adaptive gradient boosting model for the prediction of rainfall using ID3 as a base estimator. *Revue d'Intelligence Artificielle*, *36*(2), 241–250. https://doi.org/10.18280/ria.360208

Fayaz, S. A., Zaman, M., & Butt, M. A. (2021a). To ameliorate classification accuracy using ensemble distributed decision tree (DDT) vote approach: An empirical discourse of geographical data mining. *Procedia Computer Science*, *184*, 935–940. https://doi.org/10.1016/j.procs.2021.03.116

Fayaz, S. A., Zaman, M., & Butt, M. A. (2021b). An application of logistic model tree (LMT) algorithm to ameliorate Prediction accuracy of meteorological data. *International Journal of Advanced Technology and Engineering Exploration*, *8*(84), 1424–40.

Fayaz, S. A., Zaman, M., & Butt, M. A. (2021c). A hybrid adaptive grey wolf Levenberg-Marquardt (GWLM) and nonlinear autoregressive with exogenous input (NARX) neural network model for the prediction of rainfall. *International Journal of Advanced Technology and Engineering Exploration*, *9*(89), 509–522. https://doi.org/10.19101/IJATEE.2021.874647

Fayaz, S. A., Zaman, M., & Butt, M. A. (2022a). Numerical and Experimental Investigation of Meteorological Data Using Adaptive Linear M5 Model Tree for the Prediction of Rainfall. *Review of Computer Engineering Research*, *9*(1), 1–12.

Fayaz, S. A., Zaman, M., & Butt, M. A. (2022b). Knowledge Discovery in Geographical Sciences—A Systematic Survey of Various Machine Learning Algorithms for Rainfall Prediction. In *International Conference on Innovative Computing and Communications* (pp. 593–608). Springer.

Fayaz, S. A., Zaman, M., & Butt, M. A. (2022c). Performance Evaluation of GINI Index and Information Gain Criteria on Geographical Data: An Empirical Study Based on JAVA and Python. In *International Conference on Innovative Computing and Communications* (pp. 249–265). Springer.

Fayaz, S. A., Zaman, M., Kaul, S., & Butt, M. A. (2022). Is Deep Learning on Tabular Data Enough? An Assessment. *International Journal of Advanced Computer Science and Applications*, *13*(4), 2022. http://dx.doi.org/10.14569/IJACSA.2022.0130454

Kaul, S., Fayaz, S. A., Zaman, M., & Butt, M. A. (2022). Is decision tree obsolete in its original form? A burning debate. *Revue d'Intelligence Artificielle*, *36*(1), 105–113.

Kaul, S., Zaman, M., Fayaz, S. A., & Butt, M. A. (2023). Performance Stagnation of Meteorological Data of Kashmir. In *International Conference on Innovative Computing and Communications. Lecture Notes in Networks and Systems* (vol. 471). Springer. https://doi.org/10.1007/978-981-19-2535-1_63

Mohd, R., Butt, M. A., & Baba, M. Z. (2020). GWLM–NARX: grey wolf levenberg–marquardt-based neural network for rainfall prediction. *Data Technologies and Applications*, *54*(1), 85–102. https://doi.org/10.1108/DTA-08-2019-0130. 2020.

Mohd, R., Butt, M. A., & Baba, M. Z. (2022). Grey Wolf-Based Linear Regression Model for Rainfall Prediction. *International Journal of Information Technologies and Systems Approach*, *15*(1), 1-18.

Niu, J., & Zhang, W. (2015). Comparative analysis of statistical models in rainfall prediction. In *2015 IEEE International Conference on Information and Automation* (pp. 2187-2190). IEEE.

Pucheta, J. A., Cristian, M. R. R., Martín, R. H., Carlos, A. S., Patiño, H. D., & Benjamín, R. K. (2009). A feed-forward neural networks-based nonlinear autoregressive model for forecasting time series. *Comput y Sistemas*, *14*(4), 423–435.

Rezaie-balf, M., Naganna, S. R., Ghaemi, A., & Deka, P. C. (2017). Wavelet coupled MARS and M5 Model Tree approaches for groundwater level forecasting. *Journal of hydrology*, *553*, 356–373.

Singh, P., & Borah, B. (2013). Indian summer monsoon rainfall prediction using artificial neural network. *Stochastic Environmental Research and Risk Assessment*, *27*(7), 1585–1599.

Singh, U., Chauhan, S., Krishnamachari, A., & Vig, L. (2015). Ensemble of deep long short term memory networks for labelling origin of replication sequences. In *2015 IEEE International Conference on Data Science and Advanced Analytics (DSAA)* (pp.1–7). IEEE. http://dx.doi.org/10.1109/DSAA.2015.7344871

Wu, C., & Chau, K.-W. (2013). Prediction of rainfall time series using modular soft computing methods. *Engineering Applications of Artificial Intelligence*, *26*(3), 997–1007. https://doi.org/10.1016/j.engappai.2012.05.023

Xiang, Y., Gou, L., He, L., Xia, S., & Wang, W. (2018). A SVR–ANN combined model based on ensemble EMD for rainfall prediction. *Applied Soft Computing*, *73*, 874–883. https://doi.org/10.1016/j.asoc.2018.09.018

Yang, Y., Lin, H., Guo, Z., & Jiang, J. (2007). A data mining approach for heavy rainfall forecasting based on satellite image sequence analysis. *Comput Geosci*, *33*(1), 20–30.

Zaman, M., & Butt, M. A. (2012). Information translation: a practitioners approach. In *World Congress on Engineering and Computer Science* (WCECS).

Zaz, S. N., Romshoo, S. A., Krishnamoorthy, R. T., & Viswanadhapalli, Y. (2019). Analyses of temperature and precipitation in the Indian Jammu and Kashmir region for the 1980–2016 period: implications for remote influence and extreme events. *Atmospheric Chemistry and Physics*, *19*(1), 15-37. https://doi.org/10.5194/acp-19-15-2019