Classifier comparison benchmark for machine learning weather prediction enhancement

Areen Arabiat, Mohammad Hassan
Department of Communications and Computer Engineering, Faculty of Engineering, Al-Ahliyya Amman University, Amman, Jordan

Article Info

Article history:

Received Dec 7, 2025
Revised Mar 17, 2026
Accepted Apr 19, 2026

Keywords:

Artificial intelligence
Classifiers
Confusion matrix
Data mining
Machine learning

ABSTRACT

Artificial intelligence (AI) and data mining can improve next-generation weather forecasting for urban planning, agriculture, and disaster management. This study investigates how machine learning (ML) classifiers can reduce forecast errors and support decision-making in sectors that require accurate predictions, including agriculture and transportation. We evaluate four classifiers—K-nearest neighbor (KNN), random forest (RF), Naive Bayes (NB), and multilayer perceptron (MLP)—using Waikato environment for knowledge analysis (WEKA) and Orange3 to compare their performance in identifying rain. A 10-fold cross-validation approach is applied to reduce overfitting, and model effectiveness is measured using key performance indicators including accuracy, precision, sensitivity (recall), and F-measure. Results show that classifier performance varies across tools, indicating that the analytical framework can influence outcomes. Among all models, the RF classifier performs best, achieving 99.92% accuracy in WEKA and 99.9% in Orange3. The MLP also shows strong performance with 99.20% accuracy in WEKA and 98.7% in Orange3. KNN and NB exhibit comparable performance, but lower precision and F-measure in WEKA. Overall, the findings suggest that RF is the most effective approach for rain prediction using data mining tools, with practical relevance for agriculture, transportation, and power systems.

Corresponding Author:

Areen Arabiat
Department of Communications and Computer Engineering, Faculty of Engineering
Al-Ahliyya Amman University
Al-Salt, Amman, Jordan
Email: a.arabiat@ammanu.edu.jo

1. INTRODUCTION

Weather forecasting is crucial for controlling resource demand, increasing crop productivity, predicting unfavourable conditions, and comprehending Earth's future. In addition to having an effect on daily life, agriculture, and disaster mitigation, it helps sectors like emergency management, transportation, agriculture, and energy management. However, predicting biodiversity loss, addressing global warming, and comprehending Earth's future all depend on weather forecasting. It maximizes crop yield, manages demand, and offers early warnings [1]. Daily life, farming, transportation, energy management, and disaster preparedness are all directly impacted by accurate weather forecasting. For instance, accurate rainfall forecasting can improve irrigation schedules, cut down on water waste, and boost yields in agriculture. Anticipating bad weather can enhance safety and route planning in the transportation industry. Accurate forecasts have the potential to save lives and property in disaster management by enabling early warnings and prompt evacuations. The following areas are beneficial to industry [2], [3].

Reliability of crop yield is increased and water use can be reduced by up to 20% with optimal irrigation scheduling made possible by accurate rainfall prediction. As an example early warnings help farmers in regions with regular dry seasons to pick crops that tolerate drought and to shift sowing dates. In Bangladesh besides India, forecasts built with machine learning (ML) raised rice harvests by up to 15% [4]. Transport planners report that weather models guide safer routing for buses, trains plus freight. If heavy rain or fog is spotted early, crashes and hold-ups drop. When ML forecasts feed into flight plans, weather linked incidents fall by 10 to 15% [5].

Solar but also wind plants, however, depend heavily on the weather. Precise forecasts let grid operators balance supply and demand and weave more renewable power into the network. Sharper wind speed estimates, for instance, raised wind farm output by 12% as well as cut running costs [6]. ML models also speed warnings for floods, storms and heatwaves-teams evacuate and shift supplies sooner. In 2023 flood alerts driven by ML trimmed response times in Southeast Asia by 30% [7].

Traditional forecasts rely on numerical weather prediction models that reproduce the atmosphere with physical equations. Those models demand large computing power and, because the air behaves in complex ways or data remain sparse, they sometimes fail for small areas or short time spans. Classic time series tools like ARIMA add further error when seasons change sharply, above all in climates that swing widely [8].

Deep learning and ML now serve as powerful alternatives that manage large data sets, detect nonlinear links plus improve local and short-term forecasts. When combined with optimisation routines and advanced data treatment, those methods match or exceed the skill of traditional models. For tracking time patterns in weather records, deep learning designs like long short-term memory (LSTM) nets but also their variants show clear strength.

This paper stands apart because it tests multiple ML classifiers for weather prediction through a fresh forecasting approach that relies on a new Java-based data mining platform. The paper also builds a synthetic weather prediction model that draws on new, extensive data sets. Beyond refining forecast techniques, the work delivers practical answers ready for day-to-day use. The evidence gathered here upgrades present practice and sparks further invention encouraging industry to adopt data led strategies that weigh today's climate as well as tomorrows. The study backs both economic and environmental sustainability promoting a sturdier decision framework for sectors whose fortunes hinge on the weather.

In order to capture time features and correlations among multivariate time series, Venkatachalam *et al*. [9] suggests an enhanced model that integrates self-attention mechanisms with LSTM. The model improves prediction performance by 15.8% for the LSTM model and 26.4% for the suggested model by efficiently utilizing weather forecast data. The model continuously offers the best accuracy, usefulness, and flexibility for all output sequence lengths.

Hayder *et al*. [10] suggested a forecasting model for hourly precipitation seasons is developed utilizing recurrent neural networks (RNNs) and ES-LSTM. The Australian Commonwealth Office of Meteorology's Historical Daily Weather dataset is used. According to the results, artificial neural network (ANN) and decision tree (DT) obtain 96.65% and 84.0% accuracy, respectively, while ES-LSTM and RNN achieve 3.17 and 6.42 mean absolute error (MAE). Also, LSTM and transductive (T-LSTM), a hybrid deep learning model, were presented by [11] for weather forecasting with time series data. Metrics like MAE, loss, and root mean squared error are used to assess the model. With a weather prediction accuracy of 98.2%, the T-LSTM approach performs better than other approaches and is a reliable solution for hydrological variables. In another study, Saleh and Rasel [12] used the open-source Python environment Google Colab. The three models were subjected to grid search hyperparameter optimization in order to optimize prediction accuracy. The investigation showed that the random forest (RF) model was the most accurate in Rangpur district for predicting rainfall. By producing the highest correlation (97%), RF regression further supported the remarkable correlation between the rainfall that was predicted and that was actually observed. In a similar study, Hussain *et al.* [13] using the Bangladesh weather dataset deployed ML and an ensemble-based classifier to forecast the amount, frequency, and daily average temperature of rainfall. While the ensemble regression model offers the most accurate temperature prediction, the ensemble classifier beats base methods in rain prediction with an accuracy of 83.41% and a sensitivity of 78.17%.

The study of Chen *et al*. [14] focuses on the slow processing speed of classic physical process models in urban inundation prediction. It improves forecasting by combining ML technology with a linked hydrological-hydrodynamic model, which is especially useful when training samples are limited and rain patterns fluctuate. The proposed method selects Ridge for single peak rain, K-nearest neighbor (KNN) for double peak rain besides RF for uniform rain. For a 3.68 km² area it delivers mean absolute percentage errors (MAPE) of 5.32%, 7.73%, and 2.49% respectively plus it produces each forecast in 14.07 s. The approach raises both accuracy and speed and it strongly supports urban flood forecasting but also early-warning systems. An applied ML technique is used to predict rainfall at Sathanur Dam, Tamil Nadu, India, with records from 2012. It examined four key dam variables - water level, inflow, outflow, and storage volume-to see how each affect rainfall forecasts. The algorithms tested were classification and regression tree (CART), linear discriminant analysis (LDA), KNN, support vector machine (SVM), Naïve Bayes (NB), and logistic regression (LR) besides RF. CART achieved the best accuracy, 86.11% [15].

Multiple rainfall prediction models were compared using different ML methods: LR, DT, MLP, and RF. They judged each model by its accuracy, Cohen's kappa besides receiver operating characteristic (ROC) curve. LR reached 82.80% accuracy, 82.45% ROC plus 65.05% kappa the ANN reached 82.59% accuracy, 81.94% ROC, and 64.40% kappa. Both models outperformed the rest but also revealed intricate rainfall patterns [16].

Another study examined how multiple weather factors control rainfall during wet periods at Hermannsburg besides Undoolya in the Northern Territory. They tested the links with two ML models-RF or LSTM-using input sets that cover the Indian Ocean Dipole, El Niño-Southern Oscillation plus Madden-Julian Oscillation. The results show that the main climate indices-IOD, Niño 3.4 and MJO-shape the rainfall forecasts. The LSTM model outperforms RF achieving an R^2 of 0.86 but also lower root-mean-square error and MAE values on the test data [17].

Traditional ML algorithms to create a robust analysis and prediction model for rainfall in Australia, with the goal of improving forecast accuracy and reliability were applied. The dataset contains historical meteorological data from numerous Australian locales, including temperature, humidity, wind speed, and air pressure, with monthly averages from 2008 to 2017. RF, NB, K closest neighbor, DT, and LR models were used to forecast rainfall effectively. evaluating this model was constructed on accuracy, precision, and interpretability. RF model achieved an accuracy of 0.859, which helped to inform flood risk mitigation methods [18].

However, with a MAE of 0.36 and a root mean squared error of 0.90, the ensemble regression model performed better in rainfall amount prediction than other base models, such as linear regression, RF, and support vector regression. With a MAE of 0.42 and a root mean squared error of 0.54, it also did well in daily average temperature forecasts [19].

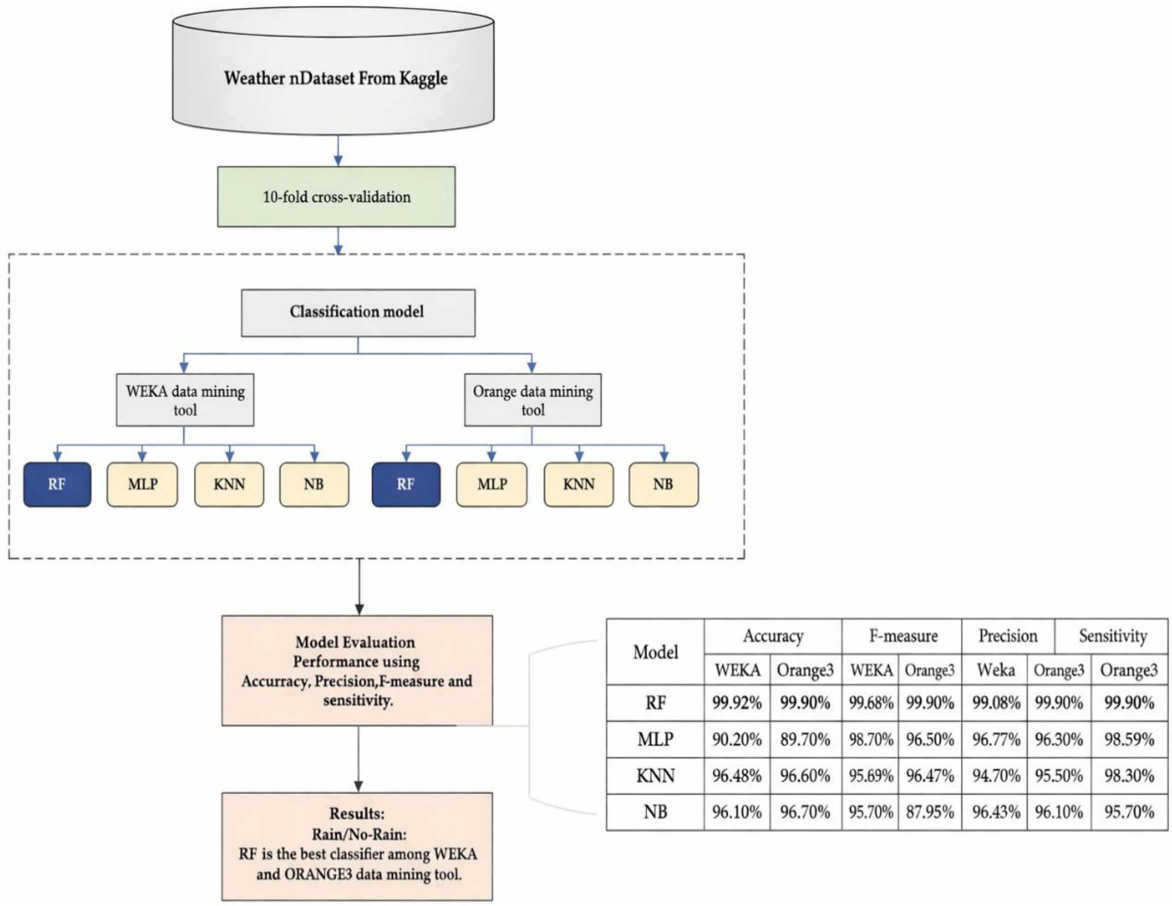
To ensure reproducibility in comparisons across Waikato environment for knowledge analysis (WEKA) and Orange3, initial experiments utilized the default learning settings of each classifier. These settings are detailed explicitly, and it is suggested that hyperparameter optimization techniques, such as grid search or Bayesian optimization, could enhance the workflow when applied to larger, temporally structured datasets. The paper addresses a notable gap in research regarding the absence of a concise, tool-agnostic baseline benchmark that compares classical classifiers under a uniform workflow, as previous studies often presented inconsistencies in data preprocessing, evaluation methods, and toolchains, complicating replication efforts and educational use. The primary contributions of this paper are: a reproducible baseline benchmarking workflow for classical ML classifiers (RF, multilayer perceptron (MLP), KNN, and NB) focused on weather-event prediction using two data-mining platforms, WEKA and Orange3; a cross-tool comparison illustrating the impact of implementation details on reporting metrics; a comprehensive error analysis utilizing confusion matrices and other metrics, with insights on the robustness of tree-ensembles like RF for tabular meteorological data, and a transparent discussion on limitations and deployment, positioning this study as a foundational step towards modern sequence models and hybrid ML–numerical weather prediction methods.

2. METHOD

The goal of this study was to predict weather conditions using multiple ML classifiers, and datasets from Kaggle were used to ensure comprehensive analysis. In order to prepare the data for analysis, data preprocessing included addressing missing values, standardizing features, and encoding categorical variables. Four classifiers were chosen for evaluation: MLP, NB, RF, and KNN. To reduce the risk of overfitting, a 10-fold cross-validation technique was used, which divides the dataset into ten subsets, nine for training and one for testing in each iteration, providing a robust assessment of model performance. Experiments were conducted utilizing both WEKA and Orange3 tools, facilitating the creation of algorithms and performance measurement. The classifiers' efficacy was assessed using a variety of metrics, including accuracy, precision, sensitivity, and the F-measure. The findings of these assessments emphasized each model's strengths and flaws, allowing for comparisons that will influence practical weather forecasting applications in areas such as agriculture and energy. This organized methodology provided useful insights into weather prediction capabilities as well as the efficiency of ML techniques. The workflow includes data acquisition, preprocessing, model training, validation, and performance evaluation Figure 1.

2.1. Dataset

The Kaggle dataset used here contains 2,500 records and represents a restricted sampling of weather conditions [14]. Accordingly, results should be interpreted as a baseline on a curated, tabular dataset rather than a substitute for operational forecasting. Moreover, the present task focuses on event occurrence (rain vs no-rain). Multi-class and regression forecasting of continuous variables (e.g., rainfall amount, temperature, and wind speed) requires additional modeling choices and will be addressed in future work with temporally ordered data. However, the dataset has multiple features, which include: temperature, humidity, wind speed, cloud cover, and the feature of the data set can be demonstrated as following: temperature, humidity, wind speed, rain/snow quantity, and cloud cover are key weather features that indicate the day's temperature, humidity, wind speed, and cloud coverage. In the other hand, data preprocessing is essential in this stage; this process includes removing or fixing inaccurate or distorted data from a dataset. To eliminate redundant data from our weather prediction system, we used excel's remove duplicates tool. Usually saved as a comma-separated values (CSV) file, data preprocessing guarantees that the data satisfies analytical requirements. However, Figure 2 shows the attributes of dataset in WEKA while Figure 3 shows the attributes of dataset in Orange.



Model	Accuracy		F-measure		Precision		Sensitivity
	WEKA	Orange3	WEKA	Orange3	Weka	Orange3	Orange3
RF	99.92%	99.90%	99.68%	99.90%	99.08%	99.90%	99.90%
MLP	90.20%	89.70%	98.70%	96.50%	96.77%	96.30%	98.59%
KNN	96.48%	96.60%	95.69%	96.47%	94.70%	95.50%	98.30%
NB	96.10%	96.70%	95.70%	87.95%	96.43%	96.10%	95.70%

Figure 1. System architecture of weather prediction

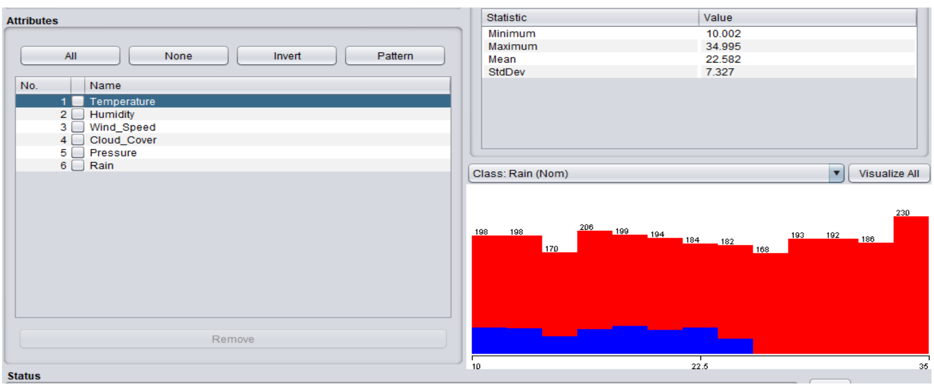


Figure 2. Attributes of the weather prediction model (WEKA)

	Name	Type	Role	Values
1	Temperature	N numeric	feature	
2	Humidity	N numeric	feature	
3	Wind_Speed	N numeric	feature	
4	Cloud_Cover	N numeric	feature	
5	Pressure	N numeric	feature	
6	Rain	C categorical	target	no rain, rain

Figure 3. Attributes of the weather prediction model (Orange 3)

2.2. Machine learning

The design and implementation of algorithms that enable computers to learn behaviors based on provided data is known as ML, and it is seen as a complement to artificial intelligence (AI) [15]. However, these days, ML models are applied to solve issues across many different fields. A ML classifier's accuracy can be considerably increased by doing the appropriate hyperparameter adjustment [16].

2.2.1. Multilayer perceptron

In AI, MLP is a non-recurrent ANN with feedforward back-propagation that is used to estimate solar radiation and generate complex predictions from non-linear data. Its input, output, and hidden layers offer a great deal of design independence for forecasting models [17], [18].

2.2.2. K-nearest neighbor

One popular and simple approach to classification problems is the KNN algorithm. Its basic idea is to use the item of interest's received data to find the closest neighbors across the entire training dataset and then classify the object to the nearest neighbor class that is most commonly represented [19].

2.2.3. Random forest

To decrease correlations and increase accuracy, the RF stochastic technique creates several decision DTs using a random vector. Every DT is separated into features, and the variety of each DT is determined by the number of attributes. The objective is to build a collection of different DT for various types of predictions, then combine the outcomes to produce a single prediction [20], [21].

2.2.4. Naive Bayes

The well-known sentiment analysis algorithm NB outperforms ML and classification techniques like KNN and DT, but it struggles to handle datasets that are unbalanced or highly correlated. It determines the likelihood that a particular event would occur given another event. Assuming that each feature contributes equally and independently, it is based on multiplying the likelihood by the evidence. The likelihood that an event will occur given that another event has previously occurred is known as the NB. The likelihood multiplied by the prior is the probability of X happening given that A happens, whereas the posterior probability is the probability that C happens given that X happens as shown in (1) [22], [23]:

$$(C|X) = \frac{P(X|C)\,P(C)}{P(X)} \quad (1)$$

where: C is target class. The data is X. $P(X)$ is the probability of the prediction (prior probability) the probability based on the hypothesis's conditions is $P(X|C)$. According to circumstances, $P(C|X)$ is hypothesis probability (posterior probability).

The dataset was split into 70% for training/validation and 30% for testing. A 10-fold cross-validation protocol was used to prevent overfitting and ensure robust performance estimates. Model evaluation metrics included accuracy, precision, sensitivity, and F-measure. By randomly dividing the dataset into training and testing subsets, cross validation assesses the performance of ML models without taking spatial dependency into account. The testing data is situated both close to and far from the training data, which is iteratively divided into k-1 sets [24], [25].

3. PERFORMANCE EVALUATION

3.1. Confusion matrix

The confusion matrix is a fundamental tool for evaluating classification models [26], [27]. It provides a detailed breakdown of true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN), which are essential for calculating performance metrics such as accuracy, precision, sensitivity, and F-measure [28], [29]. Table 1 illustrates the general structure of a confusion matrix.

In this study, confusion matrices were generated for each of the four classifiers-KNN, RF, MLP, and NB to assess their predictive performance on the weather dataset. The results using WEKA are summarized in Tables 2-5. while the results using Orange3 are summarized in Tables 2-9.

Table 1. Confusion matrix [30]

	Predicted	
Actual	TP	FN
	FP	TN

Table 2. Confusion matrix for KNN using WEKA

Confusion matrix	Rain	No rain
Rain	266	48
No rain	31	2155

Table 3. Confusion matrix for RF using WEKA

Confusion matrix	Rain	No rain
Rain	312	2
No rain	0	2186

Table 4. Confusion matrix for MLP using WEKA

Confusion matrix	Rain	No rain
Rain	300	14
No rain	6	2180

Table 5. Confusion matrix for NB using WEKA

Confusion matrix	Rain	No rain
Rain	218	96
No rain	0	2186

Table 6. Confusion matrix for KNN using orange

Confusion matrix	Rain	No rain
Rain	257	57
No rain	26	2160

Table 7. Confusion matrix for RF using orange

Confusion matrix	Rain	No rain
Rain	312	4
No rain	0	2186

Table 8. Confusion matrix for MLP using orange

Confusion matrix	Rain	No rain
Rain	291	23
No rain	9	2177

Table 9. Confusion matrix for NB using orange

Confusion matrix	Rain	No rain
Rain	259	55
No rain	52	2134

3.2. Performance matrices

In (2) to (5) describe the model's performance in detail using the metrics that are produced from the previous section, such as F1-score, sensitivity/sensitivity, accuracy, and precision [31]-[34].

$$Accuracy = (TP + TN)/(TP + TN + FP + FN) \quad (2)$$

$$Precision = TP/(TP + FP) \quad (3)$$

$$Sensitivity = TP/(TP + FN) \quad (4)$$

$$F - measure = (2 * Precision * Sensitivity)/\ (Precision\ + Sensitivity) \quad (5)$$

4. RESULTS AND DISCUSSIONS

Four classifiers are evaluated in the study: RF, MLP, KNN, and NB. With an accuracy of 99.92% and a precision of 99.36%, RF had the highest overall efficacy in producing accurate predictions. It is extremely dependable for crucial applications because of its ideal sensitivity of 100%, which guarantees that it properly detects all real positive situations. Compared to RF, the MLP has a greater proportion of FP despite its impressive accuracy of 99.20%. Although not as excellent as RF's, MLP's sensitivity was still useful. With an accuracy of 96.84% and a precision of 84.71%, KNN showed a decrease in performance, indicating a significant number of FP. KNN showed good detection capabilities for real positives, maintaining a significant sensitivity of 98.59%. With a troubling precision of 69.43% and an accuracy of 96.16%, NB performed the worst overall, indicating substantial FP rates. However, the comparison of the classifiers of the weather prediction model is shown in Tables 6 and 7 compares the accuracy of the proposed classifiers with recent deep learning and ensemble-based weather forecasting models reported in the literature. Table 10 depicts the result of performance metrices using WEKA data mining tool.

Table 10. Weather prediction model results using WEKA

Model	Accuracy (%)	Precision (%)	Sensitivity (%)	F-measure (%)
RF	99.92	99.36	100	99.68
MLP	99.20	95.59	98.04	96.77
KNN	96.84	84.71	98.59	87.07
NB	96.16	69.43	100	87.95

The results of employing the orange data mining tool demonstrate the effectiveness of several ML models in weather prediction. The RF model produced outstanding metrics, with 99.9% accuracy, precision, sensitivity, and F-measure, proving its capacity to manage complicated datasets and identify non-linear correlations found in meteorological data. The MLP model likewise performed admirably, with metrics of 98.7%, demonstrating remarkable flexibility to complicated patterns while being a dependable option for prediction tasks. The KNN model performed well, with accuracy metrics ranging from 96.6% to 96.7%, however it may suffer with extremely variable datasets. Also, NB model was the least successful across measures, with a 95.7% accuracy rate, owing to high independence assumptions that limit its ability to capture interdependent weather patterns. Overall, the RF and MLP models were the most effective, indicating that future research should focus on enhancing these models using potential ensemble techniques or additional characteristics from related datasets in order to increase predictive accuracy and operational weather forecasting. Table 11 depicts the result of performance metrices using orange data mining tool while Table 12. Comparison of the classifier's performance assessment for both tools. However, Figure 4 depicts comparative analysis of different classifiers' performances. On the other hand, Table 13 shows a comparison of the classifier's performance assessment from earlier research and suggested model. Also, Figure 5 depicts high-resolution accuracy comparison across WEKA and Orange3.

Table 11. Weather prediction model results using Orange3

Model	Accuracy (%)	Precision (%)	Sensitivity (%)	F-measure (%)
RF	99.9	99.9	99.9	99.9
MLP	98.7	98.7	98.7	98.7
KNN	96.7	96.6	96.7	96.6
NB	95.7	95.7	95.7	95.7

Table 12. Comparison of the classifier's performance assessment for both tools

Model	Accuracy		F-measure		Precision		Sensitivity	
	WEKA (%)	Orange3 (%)	WEKA (%)	Orange3 (%)	WEKA (%)	Orange3 (%)	WEKA (%)	Orange3 (%)
RF	99.92	99.90	99.68	99.90	99.36	99.90	100	99.90
MLP	99.20	98.70	96.77	98.70	95.59	98.70	98.04	98.70
KNN	96.84	96.70	87.07	96.60	84.71	96.60	98.59	96.70
NB	96.16	95.70	87.95	95.70	69.43	95.70	100	95.70

By combining numerous decorrelated DT, RF enhances generalization on tabular data while lowering variance. It is reasonably resilient to noisy attributes and outliers, and it can capture nonlinear feature interactions (such as humidity×cloud cover) without the need for feature scaling. Heterogeneous meteorological readings are a good fit for these characteristics. Accuracy warning: because carefully selected benchmark datasets may contain substantial separability signals and may not fully capture the spatiotemporal variability of actual weather, the exceptionally high accuracy values seen here should be evaluated cautiously. Performance should be reassessed on bigger, chronologically ordered datasets with stringent leakage control and extra calibration analyses for operational deployment. Finally, in operational settings, the proposed workflow can be embedded in an end-to-end pipeline that ingests observations from weather stations and IoT sensors, performs automated quality control, and serves probabilistic predictions via an API. Classical classifiers are computationally light and can support edge or near-real-time analytics, for example in agriculture (irrigation scheduling), disaster management (flood early warnings), and renewable-energy scheduling. However, operational agencies typically rely on numerical weather prediction (NWP) outputs; therefore, a realistic next step is to fuse ML classifiers with NWP predictors (e.g., model output statistics and bias correction) and to benchmark against NWP baselines. Hybrid ensembles (RF/GBDT+sequence models such as LSTM/GRU) are also promising when temporal sequences are available.

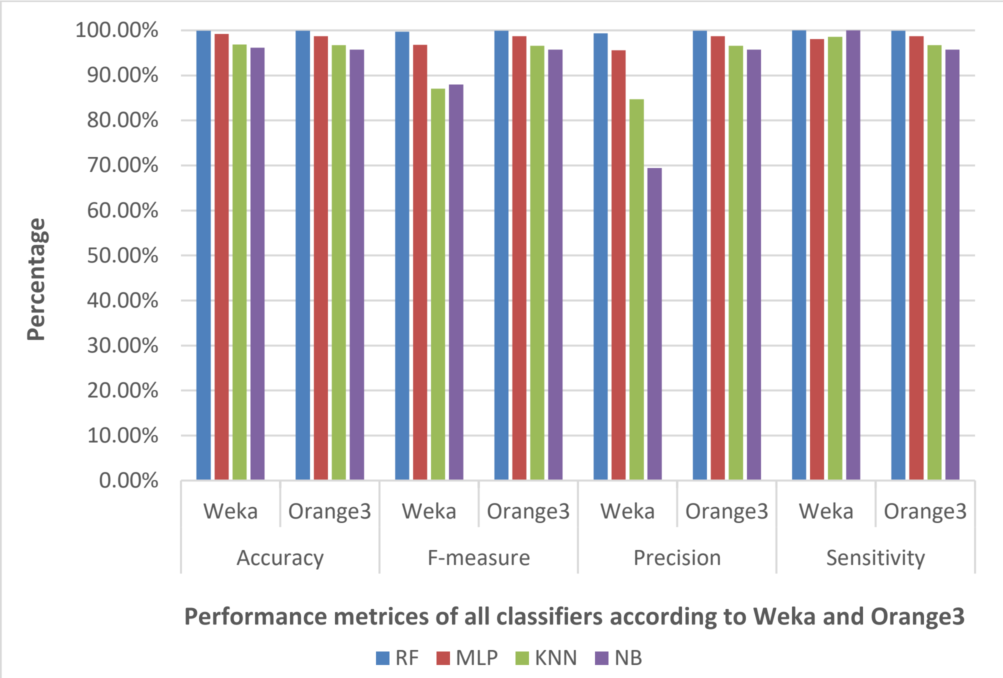


Figure 4. Comparative analysis of different classifiers' performances

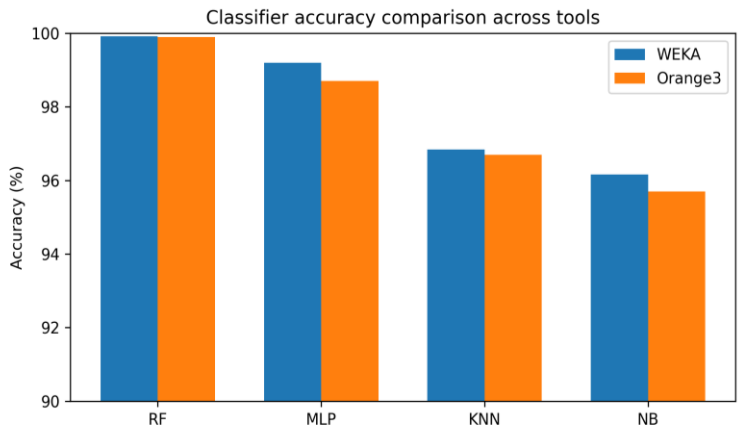


Figure 5. High-resolution accuracy comparison across WEKA and Orange3

Table 13. Comparison of the classifier's performance assessment from earlier research and suggested model

Study	Methods/model	Performance metrics	Percentage (%)
[9]	RNNs and ES-LSTM	Accuracy	84.0
[10]	T-LSTM approach	Accuracy	98.2
[11]	RF regression	Accuracy	97.0
[12]	Ensemble regression model	Accuracy	83.41
[15]	Ridge, KNN, and RF with hydrological model	MAPE	5.32 (single-peak), 7.73 (double-peak), 2.49 (uniform)
[16]	CART among various classifiers for rainfall estimation	Accuracy	86.11
[17]	LR, DT, MLP, and RF	Accuracy	82.80 (LR) and 82.59 (ANN)
[18]	RF and LSTM models	R^2 Root mean square error (RMSE) and MAE	0.86 (LSTM)
[19]	RF, NB, KNN, and DT	Accuracy	85.9
[20]	Ensemble-based classifier and regressor models	Rainfall accuracy MAPE	83.41 0.42
Proposed	RF, MLP, NB, and KNN using WEKA data mining tool	Accuracy	99.92

5. CONCLUSION

This work offers a reproducible baseline comparison of four classical machine-learning classifiers (RF, MLP, KNN, and NB) for predicting rain/no-rain events using two popular data-mining platforms (WEKA and Orange3). In the analyzed tabular dataset, RF demonstrated the best performance and lowest error rates in the confusion-matrix analysis, aligning with the known robustness of tree ensembles for handling diverse features. However, the findings should be viewed as a benchmark on a curated dataset rather than a direct indicator of operational forecasting accuracy. Future research will expand to datasets that are temporally ordered and geographically diverse, broaden the task to include multi-class and regression forecasting (e.g., predicting rainfall amounts, temperatures, and wind speeds), incorporate advanced sequence models (LSTM/GRU) and hybrid ensembles, and integrate ML predictors with outputs from numerical weather predictions using bias correction/model-output-statistics and calibration analysis. These initiatives aim to enhance practical applications in agriculture, disaster management, and renewable energy scheduling.

ACKNOWLEDGMENT

The authors would like to thank Al Ahliyya Amman University for their support.

FUNDING INFORMATION

Authors state no funding involved.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Areen Arabiat	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓	
Mohammad Hassan	✓				✓	✓			✓	✓	✓	✓	✓	

C : **C**onceptualization	I : **I**nvestigation	Vi : **Vi**sualization
M : **M**ethodology	R : **R**esources	Su : **Su**pervision
So : **So**ftware	D : **D**ata Curation	P : **P**roject administration
Va : **Va**lidation	O : Writing - **O**riginal Draft	Fu : **Fu**nding acquisition
Fo : **Fo**rmal analysis	E : Writing - Review & **E**diting	

CONFLICT OF INTEREST STATEMENT

Authors state no conflict of interest.

DATA AVAILABILITY

The dataset that supports the findings of this study is openly available in Kaggle at. https://www.kaggle.com/datasets/zeeshier/weather-forecast-dataset,reference number [20].

REFERENCES

[1] L. Chen, B. Han, X. Wang, J. Zhao, W. Yang, and Z. Yang, "Machine Learning Methods in Weather and Climate Applications: A Survey," *Applied Sciences (Switzerland)*, vol. 13, no. 21, p. 12019, Nov. 2023, doi: 10.3390/app132112019.

[2] S. Dalal, B. Seth, M. Radulescu, T. F. Cilan, and L. Serbanescu, "Optimized Deep Learning with Learning without Forgetting (LwF) for Weather Classification for Sustainable Transportation and Traffic Safety," *Sustainability (Switzerland)*, vol. 15, no. 7, p. 6070, Mar. 2023, doi: 10.3390/su15076070.

[3] V. Kumar, N. Kedam, O. Kisi, S. Alsulamy, K. M. Khedher, and M. A. Salem, "A Comparative Study of Machine Learning Models for Daily and Weekly Rainfall Forecasting," *Water Resources Management*, vol. 39, no. 1, pp. 271–290, Jan. 2025, doi: 10.1007/s11269-024-03969-8.

[4] S. B. Nuthalapati and A. Nuthalapati, "Accurate weather forecasting with dominant gradient boosting using machine learning," *International Journal of Science and Research Archive*, vol. 12, no. 2, pp. 408–422, Jul. 2024, doi: 10.30574/ijsra.2024.12.2.1246.

[5] A. Ayoub, H. M. Wainwright, and G. Sansavini, "Machine learning-enabled weather forecasting for real-time radioactive transport and contamination prediction," *Progress in Nuclear Energy*, vol. 173, p. 105255, Aug. 2024, doi: 10.1016/j.pnucene.2024.105255.

[6] K. Calvin *et al.*, "IPCC, 2023: Climate Change 2023: Synthesis Report. Contribution of Working Groups I, II and III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change [Core Writing Team, H. Lee and J. Romero (eds.)]. IPCC, Geneva, Switzerland," Jul. 2023, doi: 10.59327/IPCC/AR6-9789291691647.

[7] Z. Hu, Y. Gao, S. Ji, M. Mae, and T. Imaizumi, "Improved multistep ahead photovoltaic power prediction model based on LSTM and self-attention with weather forecast data," *Applied Energy*, vol. 359, p. 122709, Apr. 2024, doi: 10.1016/j.apenergy.2024.122709.

[8] S. Q. Dotse, I. Larbi, A. M. Limantol, and L. C. D. Silva, "A review of the application of hybrid machine learning models to improve rainfall prediction," *Modeling Earth Systems and Environment*, vol. 10, no. 1, pp. 19–44, Feb. 2024, doi: 10.1007/s40808-023-01835-x.

[9] K. Venkatachalam, P. Trojovský, D. Pamucar, N. Bacanin, and V. Simic, "DWFH: An improved data-driven deep weather forecasting hybrid model using Transductive Long Short Term Memory (T-LSTM)," *Expert Systems with Applications*, vol. 213, p. 119270, Mar. 2023, doi: 10.1016/j.eswa.2022.119270.

[10] I. M. Hayder *et al.*, "An Intelligent Early Flood Forecasting and Prediction Leveraging Machine and Deep Learning Algorithms with Advanced Alert System," *Processes*, vol. 11, no. 2, p. 481, Feb. 2023, doi: 10.3390/pr11020481.

[11] G. Rudrappa and N. Vijapur, "Survey on the research of various machine learning and deep learning techniques for precipitation forecasting," *International Journal of Industrial and Systems Engineering*, vol. 52, no. 2, pp. 190-211, 2026, doi: 10.1504/IJISE.2026.151667.

[12] Md. A. Saleh and H. M. Rasel, "Performance evaluation of Machine Learning based regression models for rainfall forecasting," Research Square, preprint, 2024, doi: 10.21203/rs.3.rs-3856741/v1.

[13] A. Hussain, A. Aslam, S. Tripura, V. Dhanawat, and V. Shinde, "Weather Forecasting Using Machine Learning Techniques: Rainfall and Temperature Analysis," *Journal of Advances in Information Technology*, vol. 15, no. 12, pp. 1329-1338, 2024, doi: 10.12720/jait.15.12.1329-1338.

[14] G. Chen *et al.*, "Urban inundation rapid prediction method based on multi-machine learning algorithm and rain pattern analysis," *Journal of Hydrology*, vol. 633, p. 131059, Apr. 2024, doi: 10.1016/j.jhydrol.2024.131059.

[15] S. Singarasubramanian, M. Anbumani, and K. Kaniyaiah, "Rainfall prediction around Sathanur dam by Naive Bayes classifier, logistic regression models and various classification and regression machine learning techniques," *Multidisciplinary Science Journal*, vol. 6, no. 10, p. 2024200, Apr. 2024, doi: 10.31893/multiscience.2024200.

[16] M. U. S. Khan, K. M. Saifullah, A. Hussain, and H. M. Azamathulla, "Comparative analysis of different rainfall prediction models: A case study of Aligarh City, India," *Results in Engineering*, vol. 22, p. 102093, Jun. 2024, doi: 10.1016/j.rineng.2024.102093.

[17] R. Farooq, M. A. Imteaz, D. Shangguan, and K. Hlavčová, "Machine learning algorithms to forecast wet-period rainfall using climate indices in Northern Territory of Australia," *Science Talks*, vol. 12, p. 100397, Dec. 2024, doi: 10.1016/j.sctalk.2024.100397.

[18] S. S. Band *et al.*, "Hybrid machine learning and deep learning models for river suspended sediment load forecasting," *Engineering Applications of Computational Fluid Mechanics*, vol. 20, no. 1, Dec. 2026, doi: 10.1080/19942060.2025.2591799.

[19] C. Xu and Z. Fan, "A performance-driven multi-stage KNN approach for local adaptive classification," *Applied Soft Computing*, vol. 175, p. 113070, Mar. 2025, doi: 10.1016/j.asoc.2025.113070.

[20] Z. Ahmad, "Weather Forecast Dataset," Kaggle. [Online]. Available: https://www.kaggle.com/datasets/zeeshier/weather-forecast-dataset, (Accessed: Dec 21, 2025).

[21] E. Elgeldawi, A. Sayed, A. R. Galal, and A. M. Zaki, "Hyperparameter tuning for machine learning algorithms used for arabic sentiment analysis," *Informatics*, vol. 8, no. 4, p. 79, Nov. 2021, doi: 10.3390/informatics8040079.

[22] I. M. Mfetoum *et al.*, "A multilayer perceptron neural network approach for optimizing solar irradiance forecasting in Central Africa with meteorological insights," *Scientific Reports*, vol. 14, no. 1, p. 3572, Feb. 2024, doi: 10.1038/s41598-024-54181-y.

[23] O. Almomani *et al.*, "A Robust Model for Android Malware Detection via ML and DL classifiers," *Mesopotamian Journal of Big Data*, vol. 2025, pp. 261–277, Sep. 2025, doi: 10.58496/MJBD/2025/017.

[24] C. Wu, I. C. Wong, Y. Wang, W. Ke, and X. Yang, "Experimental Study of Bluetooth Indoor Positioning Using RSS and Deep Learning Algorithms," *Mathematics*, vol. 12, no. 9, p. 1386, May 2024, doi: 10.3390/math12091386.

[25] J.-L. Zheng and S.-E. Fang, "NGBoost-Naïve Bayes collaborative deep learning for structural safety evaluation of bridges," *Reliability Engineering & System Safety*, vol. 271, p. 112230, Jan. 2026, doi: 10.1016/j.ress.2026.112230.

[26] A. Suleymanov *et al.*, "Random Forest Modeling of Soil Properties in Saline Semi-Arid Areas," *Agriculture (Switzerland)*, vol. 13, no. 5, p. 976, Apr. 2023, doi: 10.3390/agriculture13050976.

[27] S. Naiem, A. E. Khedr, A. M. Idrees, and M. I. Marie, "Enhancing the Efficiency of Gaussian Naïve Bayes Machine Learning Classifier in the Detection of DDOS in Cloud Computing," *IEEE Access*, vol. 11, pp. 124597–124608, 2023, doi: 10.1109/ACCESS.2023.3328951.

[28] C. Kumar, G. Walton, P. Santi, and C. Luza, "Random Cross-Validation Produces Biased Assessment of Machine Learning Performance in Regional Landslide Susceptibility Prediction," *Remote Sensing*, vol. 17, no. 2, p. 213, Jan. 2025, doi: 10.3390/rs17020213.

[29] B. Al-Naami, H. Fraihat, H. A. Owida, K. Al-Hamad, R. D. Fazio, and P. Visconti, "Automated Detection of Left Bundle Branch Block from ECG Signal Utilizing the Maximal Overlap Discrete Wavelet Transform with ANFIS," *Computers*, vol. 11, no. 6, p. 93, Jun. 2022, doi: 10.3390/computers11060093.

[30] O. Almomani, A. Alsaaidah, A. A. A. Shareha, A. Alzaqebah, and M. Almomani, "Performance Evaluation of Machine Learning Classifiers for Predicting Denial-of-Service Attack in Internet of Things," *International Journal of Advanced Computer Science and Applications*, vol. 15, no. 1, pp. 263–271, 2024, doi: 10.14569/IJACSA.2024.0150125.

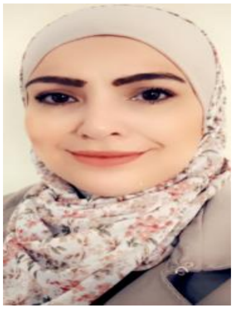
[31] G. Zeng, "Invariance Properties and Evaluation Metrics Derived from the Confusion Matrix in Multiclass Classification," Mathematics, vol. 13, no. 16, p. 2609, Aug. 2025, doi: 10.3390/math13162609.

[32] A. M. Arabiat, "Intelligent Model for Detecting GAN-Generated Images Based on Multi-Classifier and Advanced Data Mining Techniques," *International Journal of Electrical and Electronic Engineering and Telecommunications*, vol. 14, no. 3, pp. 147–157, 2025, doi: 10.18178/ijeetc.14.3.147-157.

[33] N. Alshdaifat et al., "Automated blood cancer detection models based on EfficientNet-B3 architecture and transfer learning," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 36, no. 3, p. 1731, Oct. 2024, doi: 10.11591/ijeecs.v36.i3.pp1731-1738.

[34] H. A. Owida *et al.*, "A deep learning-based dual-branch framework for automated skin lesion segmentation and classification via dermoscopic Images," *Scientific Reports*, vol. 15, no. 1, p. 37823, Oct. 2025, doi: 10.1038/s41598-025-21783-z.

BIOGRAPHIES OF AUTHORS



Areen Arabiat earned her B.Sc. in Computer Engineering in 2009 from al Balqaa Applied University (BAU), and her M.Sc. in Intelligent Transportation Systems (ITS) from Al Ahliyya Amman University (AAU) in 2022. She is currently a computer lab supervisor and (part time) lecturer in the Department of Communications and Computer Engineering at Al-Ahliyya Amman University since 2013. Her research interests are focused on the following areas: machine learning, data mining, image processing, internet of things (IoT), and artificial intelligence. She can be contacted at email: a.arabiat@ammanu.edu.jo.

Mohammad Hassan has completed his Ph.D. from Baku State University, Azerbaijan. He is an Associate Professor in the Computer Engineering Department at the Faculty of Engineering at Al-Ahliyya Amman University. He is a member of the Jordanian Engineering Association. He has published numerous research papers in various journals and conferences, covering topics such as machine learning, computer networks, intelligent transportation systems, and mobile learning adaptation models. He can be contacted at email: mhassan@ammanu.edu.jo.