International Journal of Mathematics and
Computer Science, **19**(2024), no. 1, 171–179

Nonlinear Model for Precipitation Forecasting in Northern Iraq using Machine Learning Algorithms

Muzahem M. Al-Hashimi, Heyam A. A. Hayawi

Department of Statistics and Informatics
College of Computer Science and Mathematics
University of Mosul
Mosul, Iraq

email: muzahim_alhashime@uomosul.edu.iq, he.hayawi@uomosul.edu.iq

(Received July 2, 2023, Accepted August 3, 2023,
Published August 31, 2023)

Abstract

Precipitation forecasting is a challenging task in meteorology, especially in Iraq, where the complex and nonlinear nature of precipitation requires powerful computational capabilities. Our study focuses on monthly rainfall in five northern provinces of Iraq: Nineveh, Dahuk, Erbil, Kirkuk, and Al-Sulaymaniyah. We propose bagging ensemble learning model incorporating Random Trees (RT), Locally Weighted Learning (LWL), and k-Nearest Neighbor (kNN) as base models. We utilize meteorological data and evaluate various input parameters. The results, measured using MSE, MAE, and RMSE, demonstrate the effectiveness and efficiency of our approach. The BG-LWL-RT model performs best in precipitation forecasting for Dahuk, Erbil, and Kirkuk, while the BG-LWL-kNN model achieves the highest accuracy for predicting precipitation in the Nineveh and Al-Sulaymaniyah datasets.

Key words and phrases: Bagging ensemble model, LWL, kNN, RT, Precipitation, Iraq.
AMS (MOS) Subject Classifications: 60, 62.
ISSN 1814-0432, 2024, http://ijmcs.future-in-tech.net

1 Introduction

Iraq, located in southwest Asia, is a country spanning 438,320 square kilometers [1] with diverse climatic regions. It experiences hot and dry summers, cold winters, and distinct variations across different areas. Rainfall plays a vital role in Iraq's socioeconomic landscape, and the country can be divided into three climatic regions [2]. The northeast mountainous region has moderate summers, cold winters, and annual rainfall exceeds 400 mm. The terrain region, located between the mountains and southern desert areas, has a transitional climate with annual rainfall ranging from 200 to 400 mm. The sedimentary plain and western plateau region, covering the majority of Iraq's land area, has a hot desert climate with extreme temperature variations and annual rainfall less than 200 mm [2].

Reliable rainfall data is crucial for the sustainable economic growth, water resource management, and agricultural practices of Iraq, especially as the agricultural sector faces challenges due to water scarcity intensified by upstream water policies of neighboring countries Turkey and Iran. Informed risk management and decision-making depend on accurate rainfall information.

Over the past decades, traditional statistical techniques like ARMA, ARIMA, SARIMA, and others have been extensively utilized in precipitation forecasting. However, these conventional models have inherent limitations, particularly when confronted with the nonlinear characteristics of precipitation time series [3]. Additionally, the prediction of precipitation time series can pose challenges due to its decomposition into multiple components, each representing a distinct underlying pattern category [4]. To tackle these challenges, non-linear and emerging machine learning models such as RT, LWL, and kNN have emerged as powerful solutions for nonlinear precipitation prediction.

The objective of this article was to forecast monthly precipitation in the northern provinces of Iraq (Nineveh, Dohuk, Erbil, Kirkuk, and Sulaymaniyah) until the year 2026, utilizing meteorological data. To achieve this, bagging ensemble approach was proposed by coupling three machine learning algorithms: LWL, kNN, and RT.

2 Data and methods

2.1 Study Area and Data Sources

The study focused on five provinces situated in the northern region of Iraq, which were selected as the study areas (Figure 1). These provinces include Nineveh, Kirkuk, Dahuk, Erbil, and Sulaymaniyah. These provinces were selected due to their importance for water resources and relatively regular precipitation patterns compared to other parts of Iraq.

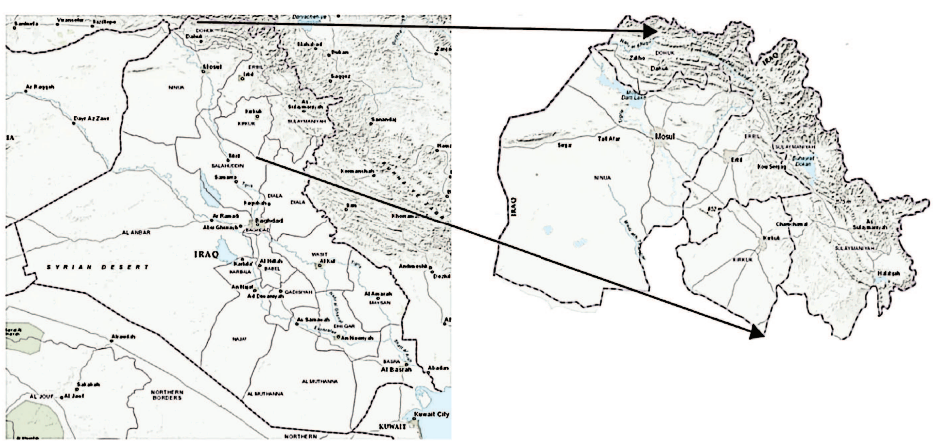

Figure 1: The Study Area's Location of the Five Provinces: Nineveh, Kirkuk, Dahuk, Erbil, and Sulaymaniyah on a Map.

The dataset used for modeling monthly precipitation spans a period of 41 years, from January 1982 to December 2022 (492 instances), and includes meteorological variables, precipitation, minimum and maximum temperature, pressure, humidity, wind direction, wind speed, and cloud cover obtained from the Iraqi Agrometeorological Center. During the data pre-processing phase, cleaning and normalization techniques were applied to handle missing values and enhance forecasting efficiency [5]. To develop and evaluate the models, the dataset was partitioned into two subsets: a training set, which encompassed 70% of the data from January 1982 to August 2010, and a test set, which consisted of the remaining 30% of the data spanning September 2010 to December 2022. The purpose of the test set was to evaluate the models' performance on unseen data and assess their ability to generalize beyond the training data.

2.2 Methods

Bagging ensemble method was used as the primary algorithm to integrate the predictions of multiple learners. Bagging was highlighted as a specific algorithm that reduces variance and overfitting by merging multiple learners into a single classifier [6].
Locally Weighted Learning (LWL) is a non-parametric algorithm capable of handling both linear and non-linear regression scenarios. It incorporates nearby training samples into the neighborhood of test samples [7] and utilizes a simple linear model to estimate non-linear functions [8]. Additionally, LWL utilized two additional algorithms: k-Nearest Neighbor (kNN) and Random Trees (RT).
k-Nearest Neighbor (kNN) is widely used as a non-parametric method that assigns weights to neighbors based on proximity and utilizes majority voting for class membership determination [9]. In addition to kNN, the Random Tree method utilizes random attribute selection at each analysis node, enabling the calculation of class probabilities. This method produces randomized trees, where each tree is chosen randomly and has an equal probability of being tested with different permutations [10].
The analyses were carried out using the WEKA program. The performance of the machine learning methods was evaluated using evaluation metrics, including Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and Mean Absolute Error (MAE).

3 Results

To identify the variables that demonstrate a correlation with precipitation, Table 1 presents the Pearson correlation coefficients between environmental features and rainfall. The analysis revealed that humidity did not have a significant impact on precipitation forecasts in Nineveh, Dahuk, Erbil, and Sulaymaniyah. Conversely, pressure, wind speed, wind direction, minimum temperature, maximum temperature, and cloud amount demonstrated a strong correlation with precipitation prediction in all the five provinces.

Attributes	Provinces				
	Nineveh	Dahuk	Erbil	Kirkuk	Sulaymaniyah
Temperature-min	0.631*	0.647*	0.614*	0.599*	0.589*
Temperature-max	0.632*	0.637*	0.611*	0.590*	0.596*
Pressure	0.537*	0.467*	0.536*	0.529*	0.473*
Humidity	0.043	0.021	0.042	0.148*	0.079
Wind Speed	0.528*	0.462*	0.418*	0.453*	0.534*
Wind Direction	0.496*	0.450*	0.479*	0.466*	0.267*
Cloud amount	0.715*	0.693*	0.664*	0.614*	0.642*
Precipitation	1.000	1.000	1.000	1.000	1.000

Province	Evaluation on test data	Model					
		RT	kNN	LWL - RT	LWL - kNN	BG-LWL - RT	BG-LWL - kNN
Nineveh	MSE	0.941	0.794	0.868	0.778	0.783	0.770
	MAE	0.607	0.561	0.587	0.551	0.572	0.548
	RMSE	0.970	0.891	0.932	0.882	0.884	0.877
Dahuk	MSE	0.864	0.705	0.849	0.695	0.646	0.673
	MAE	0.655	0.555	0.632	0.552	0.539	0.547
	RMSE	0.929	0.839	0.921	0.833	0.804	0.820
Erbil	MSE	0.694	0.636	0.691	0.634	0.587	0.628
	MAE	0.555	0.519	0.542	0.519	0.496	0.507
	RMSE	0.833	0.797	0.831	0.796	0.766	0.792
Kirkuk	MSE	0.747	0.674	0.700	0.669	0.662	0.666
	MAE	0.547	0.513	0.527	0.510	0.508	0.521
	RMSE	0.864	0.821	0.837	0.818	0.814	0.816
Sulaymaniyah	MSE	0.988	0.910	0.959	0.881	0.902	0.782
	MAE	0.620	0.610	0.614	0.587	0.591	0.502
	RMSE	0.994	0.954	0.979	0.938	0.950	0.884

During the experimentation phase, an extensive evaluation was conducted to determine the best learning method for the precipitation dataset. However, the results revealed varying levels of accuracy among the different learning methods. The performance metrics of RT, kNN, LWL-RT, LWL-kNN, BG-LWL-RT, and BG-LWL-kNN models during the testing phase are displayed in Table 2, highlighting the performance of the best models. Notably, the proposed model BG-LWL-RT demonstrated superior performance compared to other models when predicting monthly precipitation for Dahuk, Erbil, and Kirkuk. Similarly, the proposed model BG-LWL-kNN achieved the highest accuracy for the Nineveh and al-Sulaymaniyah datasets. These findings indicate that both the BG-LWL-RT and BG-LWL-kNN models surpassed the other models in terms of the evaluation criteria (Table 2).

Figures 2-6 display the actual and predicted precipitation data for the five provinces, covering the period from September 2010 to December 2022. The

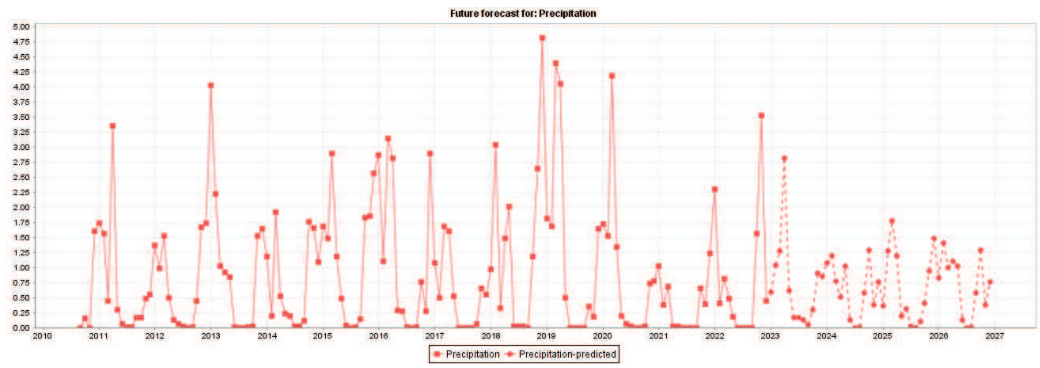

Figure 2: Precipitation forecasting of Nineveh province.

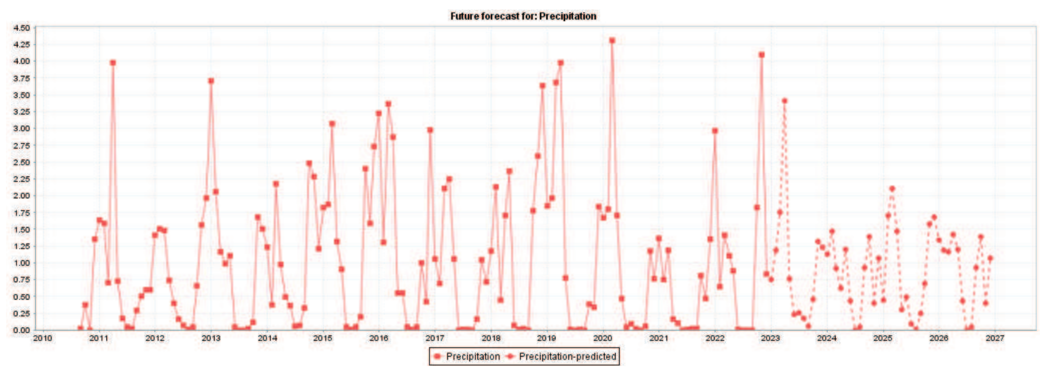

Figure 3: Precipitation forecasting of Dahuk province.

actual data represents 30% of the dataset, while the predicted data consists of 48 instances spanning from January 2023 to December 2026.

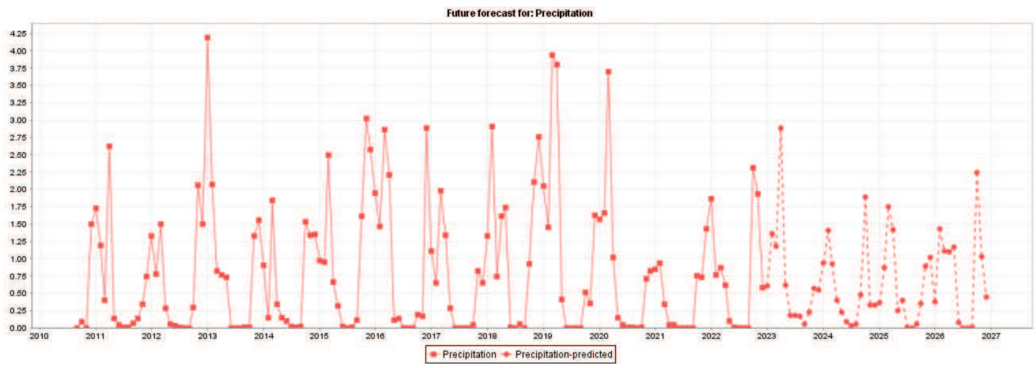

Figure 4: Precipitation forecasting of Erbil province.

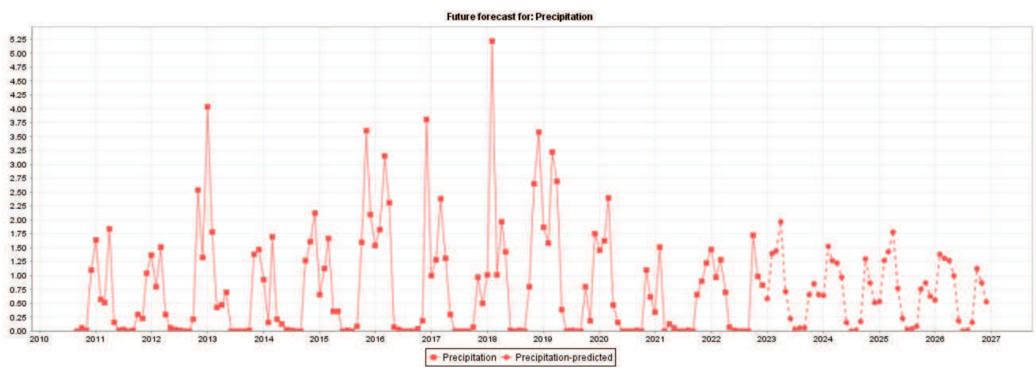

Figure 5: Precipitation forecasting of Kirkuk province.

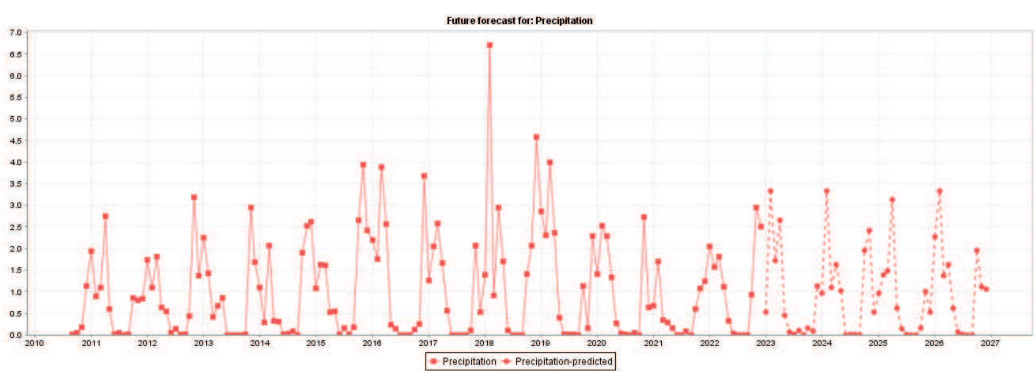

Figure 6: Precipitation forecasting of Sulaymaniyah province.

4 Discussion and Conclusion

In this paper we presented bagging ensemble models for forecasting precipitation in five northern provinces of Iraq: Nineveh, Dahuk, Erbil, Kirkuk, and Sulaymaniyah. The datasets used in this study were obtained from the Iraqi Agrometeorological Center. The experimental results demonstrated that the BG-LWL-RT precipitation forecasting model performs best for Dahuk, Erbil, and Kirkuk, while the BG-LWL-kNN model yields the highest accuracy rates for forecasting precipitation in the Nineveh and al-Sulaymaniyah datasets. These findings highlight the effectiveness of the proposed models in accurately predicting the considered time series.

Ensemble methods provide a supervised learning approach that has achieved remarkable success across numerous real-world tasks. It represents a cutting-edge learning methodology that involves training multiple learners and subsequently combining their outputs. Bagging is a prominent example of ensemble techniques used to amalgamate the individual models and achieve superior performance compared to using a single learner alone [11].

Accurate rainfall forecasting is crucial to address challenges in Iraq. Machine learning techniques leverage historical rainfall patterns to predict future precipitation with exceptional accuracy, benefiting decision-makers, policymakers, and farmers in effective planning and preparation for potential impacts like floods or droughts. These forecasts also assist farmers in optimizing water usage and strategically planning their agricultural activities.

Our case study provided compelling evidence that the bagging ensemble models exhibited superior performance compared to the individual LWL, RT, and kNN models. The evaluation metrics MSE, MAE, and RMSE clearly demonstrated their remarkable accuracy and outperformance in precipitation forecasting. This achievement can be attributed to the inherent nonlinear characteristics of precipitation. By utilizing bagging ensemble models that leverage the strengths of multiple models, we successfully captured the complexity and nonlinearity inherent in precipitation. Consequently, these bagging ensemble models showcased exceptional capability and efficiency in accurately predicting monthly precipitation, surpassing the performance of individual models.

References

[1] Ayad M. Al-Quraishi, Fadhil, Abdelazim M. Negm. "Environmental remote sensing and GIS in Iraq", Springer-Water, 2019.

[2] K. A. Rahi, A. S. Al-Madhhachi, S. N. Al-Hussaini. Assessment of surface water resources of eastern Iraq. Hydrology, **6,** no. 3, (2019), 57.

[3] C. L. Wu, K. W. CL, Chau, Prediction of rainfall time series using modular soft computing methods. Engineering applications of artificial intelligence, **26,** no. 3, (2013), 997–1007.

[4] R. Mahmood, S. Jia, W. Zhu. Analysis of climate variability, trends, and prediction in the most active parts of the Lake Chad basin, Africa. Scientific Reports, **9,** no. 1, (2019), Article no. 6317.

[5] F. Di Nunno, F. Granata, Q. B. Pham, G. de Marinis, Precipitation Forecasting in Northern Bangladesh Using a Hybrid Machine Learning Model, Sustainability, **14,** no. 5, (2022), Article no. 2663.

[6] O. Sagi, L. Rokach. Ensemble learning: A survey. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, **8,** no. 4, (2018), e1249.

[7] O. Reyes, A. Cano, H. M. Fardoun, S. Ventura. A locally weighted learning method based on a data gravitation model for multi-target regression. International Journal of Computational Intelligence Systems, **11,** no. 1, (2018), 282–295.

[8] C. G. Atkeson, A. W. Moore, S. Schaal. Locally weighted learning. Lazy learning, **1997,** 11–73.

[9] H. Saadatfar, S. Khosravi, J. H. Joloudari, A. Mosavi, S. A. Shamshirband. A new K-nearest neighbors classifier for big data based on efficient data pruning. Mathematics. **8,** no. 2, (2020), 286.

[10] O. Moscoso-Zea, P. Saa, S. Lujan-Mora. Evaluation of algorithms to predict graduation rate in higher education institutions by applying educational data mining. Australasian Journal of Engineering Education **24,** no. 1, (2019), 4–13.

[11] Z. H. Zhou. Ensemble methods: foundations and algorithms, CRC press, 2012.