Evaluating machine learning models for precipitation prediction in Casablanca City

Abdelouahed Tricha, Laila Moussaid
E2IT group research, LRI laboratory, National School of Electrical and Mechanical Engineering (ENSEM), Hassan II University, Casablanca, Morocco

Article Info

Article history:

Received Jan 16, 2024
Revised Mar 23, 2024
Accepted Apr 13, 2024

Keywords:

Artificial intelligence
Ensemble learning
Machine learning
Precipitation prediction
Water management

ABSTRACT

Accurate precipitation forecasting is a vital task for many domains, such as agriculture, water management, flood prevention, and crop yield estimation. The use of machine learning (ML) approaches has improved precipitation forecasting accuracy, exhibiting promising results in capturing the intricate connections between various meteorological variables and precipitation patterns. However, given the vast array of available ML models, a comparative analysis is imperative for identifying the most effective models for precipitation prediction. This study aims to examine the capacities of ML algorithms to forecast precipitation based on weather data for the city of Casablanca, Morocco, which faces challenges in water management and climate change adaptation. Eight different ML models' performances are compared: linear regression, polynomial regression, K-nearest neighbors (KNN), support vector machine (SVM), decision tree (DT), random forest (RF), XGBoost, and an ensemble learning model. These models are evaluated based on their mean absolute error (MAE), mean squared error (MSE), and R-squared (R^2) value to determine their effectiveness. The study showcases the potential of ML models in predicting precipitation by utilizing meteorological parameters such as temperature, humidity, wind speed, and pressure.

Corresponding Author:

Abdelouahed Tricha
E2IT group research, LRI laboratory, National High School for Electricity and Mechanics (ENSEM)
Hassan II University
Casablanca, Morocco
Email: tricha.abdelouahed@gmail.com

1. INTRODUCTION

Precipitation plays a crucial role in the water cycle and profoundly affects various aspects of our everyday life and the environment. It is a key input for hydrological modelling and forecasting and is widely used in sectors such as agriculture, flood and drought forecasting, water resource management, and so on [1], [2]. Making informed decisions about water management and preparing for floods and droughts can be made easier for communities with the help of accurate rainfall forecasts. Knowing how much rain to expect allows communities to plan and manage their water more effectively. For agriculture, knowing when and how much precipitation to expect can help farmers decide when to plant and irrigate their crops to ensure optimal growth and yields. In addition, accurate rainfall forecasts are critical for flood and drought preparedness, allowing authorities to take timely action, such as implementing flood control measures or water restrictions during droughts.

Recently, applying machine learning (ML) to climate forecasting has been investigated. Researchers have been inspired to use artificial intelligence (AI) methods to predict and categorize events under different climatic conditions by combining AI techniques with meteorological science [3], [4]. Time series data, gathered according to specific patterns, are an important field of research [5]-[10]. The intricate interactions between many meteorological data points, such as wind speed, temperature, humidity, pressure, and precipitation, may be modeled using ML techniques. These models may then be deployed to accurately forecast precipitation, providing useful information to communities and decision-makers.

This study focuses on the city of Casablanca, Morocco, which faces challenges in water management and climate change adaptation. Casablanca is the largest city and the economic hub of Morocco, with a population of about 3.7 million people [11]. It is located on the Atlantic coast and has a Mediterranean climate, with mild and wet winters and warm and dry summers. Precipitation is highly variable and unevenly distributed throughout the year, with most of the rainfall occurring between November and March. The average annual rainfall is about 412 mm [11], which is below the national average of 500 mm. Casablanca suffers from water scarcity and pollution, due to the increasing demand for water from the growing population and the industrial and agricultural sectors, as well as the inadequate water supply and sanitation infrastructure. Climate change is expected to exacerbate these problems, by reducing the availability and quality of water resources and increasing the frequency and intensity of extreme weather events, such as floods and droughts. Therefore, precipitation forecasting is crucial for Casablanca, as it can help improve water management and planning, and reduce the risks and impacts of climate change.

In this study, precipitation will be predicted using seven ML models: linear regression (LR), polynomial regression (PR), K-nearest neighbors (KNN), support vector machine (SVM), decision tree (DT), random forest (RF) and XGBoost. Additionally, to increase the overall precision of precipitation prediction, we will assess the effectiveness of an ensemble model that integrates the results of these seven models. The research will be based on 42 years' worth of daily data, which will include surface pressure (PS), wind speed, temperature, specific humidity, and precipitation. The stages in data pre-processing and feature engineering, the model training and assessment procedure, and the prediction outcomes will all be included in the article. This study's goals are to show how ML techniques may be used to forecast precipitation and offer insights into how various algorithms and the ensemble model perform by comparing their performance.

The paper is divided into five sections. The first section provides an overview of the problem of precipitation forecasting and the motivation for using ML techniques. The second section presents a literature review of previous work on precipitation forecasting using ML. The third section outlines the methodology used to implement and evaluate the ML models. The fourth section presents the results and evaluation of the models, followed by a conclusion.

2. LITERATURE REVIEW

ML has shown great potential for improving precipitation forecasting in various regions and applications. One area of research has focused on improving the accuracy and reliability of precipitation estimates and predictions. Yang *et al.* [12] focus on using ML algorithms for accurate climate analysis and prediction of summer precipitation in China. The study conducted a multi-model ensemble prediction experiment using three tree-based ML algorithms and explored the impact of different hyperparameters on prediction accuracy. Basha *et al.* [13] compared different ML and deep learning techniques for rainfall prediction, showing that deep learning models performed better than traditional ML models. Similarly, Balamurugan and Manojkumar [14] applied ML models, including RF and artificial neural networks, to short-term rain forecasting, achieving better accuracy than traditional statistical models.

Another area of study has centered on the creation of ensemble prediction methods that incorporate the strengths of various ML models. Huang *et al.* [15] proposed a novel method for precipitation forecasting based on an enhanced KNN algorithm that takes into account temporal and spatial correlations in meteorological data. To enhance precipitation forecasting in northern Bangladesh, Nunno *et al.* [16] created a hybrid ML model based on wavelet transform-based M5P and support vector regression. Yang *et al.* [12] created an ensemble model for summer precipitation in China using RF, SVM, and KNN models, while Ahmed *et al.* [17] used an ensemble model with RF, support vector regression, and an artificial neural network to predict precipitation and temperature in the United States.

Finally, some research has looked into the application of ML to the prediction of precipitation in specific regions and applications. Using ML models, Anochi *et al.* [18] predicted precipitation over South America and compared the results to the operational general circulation atmospheric model at the National Institute of Space Research. To enhance the quality of precipitation forecasts, Zhang and Ye [19] conducted a large-scale comparison study of 21 ML algorithms with diverse model structures and regularization strategies. The optimized extra-trees regressor has been identified as the optimal algorithm for minimizing overestimation and attaining the highest level of skill in precipitation forecasting. These surveys highlight the potential of ML for practical applications in water resource management, disaster risk reduction, and climate services, as well as the relevance of taking regional and application-specific factors into account during model development and evaluation.

3. METHOD

3.1. Dataset

The dataset used to train and evaluate the ML models in this study was obtained from the NASA Langley Research Center website [20]. It contains daily data for Casablanca, Morocco, spanning 42 years for a variety of meteorological variables, including wind speed (WS2M), temperature (T2M), specific humidity (QV2M), precipitation (PRECTOTCORR), and PS. The dataset also contains the date of each observation, with columns for the year, month, and day. More than 15,000 observations are included in the dataset, which spans from January 1, 1979, to January 1, 2023. Various instruments and methodologies, including ground-based observations, satellites, and reanalysis, were used to collect the data. Various quality control and homogenization procedures were implemented to ensure the accuracy and consistency of the data. The descriptive statistics of the meteorological variables are presented in Table 1. Including the mean, standard deviation, minimum, and maximum values. The data shows that there is a great deal of variability in the meteorological variables, especially precipitation.

For this study, we used the daily PRECTOTCORR as the target variable and the remaining variables as the predictors. To prevent overfitting and improve the performance of the models, we split the dataset into a training set (80%) and a test set (20%). The training set was used to train and tune the models, and the test set was used to evaluate their performance. The data was first preprocessed to handle outliers and missing values. The missing data were handled using the mean imputation approach, which substituted the mean value of the appropriate column for the missing values. To address outliers, the z-score method was utilized to find and remove values that significantly deviated from the mean.

Table 1. Descriptive statistics of meteorological variables in Casablanca, Morocco

	WS2M (m/s)	T2M (C)	QV2M (g/kg)	PRECTOTCORR (mm/day)	PS (kPa)
Count	14976	14976	14976	14976	14976
Mean	2.949663	18.590029	9.656567	1.210107	99.961595
Std	1.036048	4.440644	2.276189	4.458472	0.478710
Min	0.800000	5.430000	3.230000	0.000000	97.120000
Max	11.2600	35.71000	16.2400	207.870000	101.78000

3.2. ML algorithms

ML algorithms optimize tasks by learning from experience through training data. When the learned results satisfy the mathematical and statistical performance metrics, the algorithms can make accurate predictions for new data, or test data, by applying the learned rules during the training process [21]. Before importing the training samples into the taught model, it is essential to preprocess and organize the dataset appropriately, which involves data cleaning, data integration, and data conversion [22]. This study utilizes eight established ML algorithms to predict the amount of daily rainfall.

3.2.1. Linear regression

LR is a supervised ML method that seeks to establish quantitative relationships among multiple variables [21]. LR is simple, fast, and interpretable, but it may not capture the nonlinear and complex relationships in the data [23]. The mathematical expression for LR is defined as (1).

$$\hat{y} = w_0x_0 + w_1x_1 + w_1x_1 + \dots + w_nx_n \quad (1)$$

In the given equation, $\hat{y}$ presents the output of the model, $x_0, \dots, x_\mathrm{n}$ represent the input variables of the model and $w_0, \dots, w_\mathrm{n}$ represent the regression parameters.

3.2.2. Polynomial regression

PR, an extension of LR, is a versatile ML model used for capturing nonlinear relationships between variables. While LR assumes a linear relationship between the independent and dependent variables, PR accommodates curvature and nonlinear patterns by introducing polynomial features. The model can be represented as (2).

$$\hat{y}_\mathrm{i} = w_0 + w_1x_i + w_2x_i^2 + \dots + w_nx_i^n \quad (2)$$

In this equation, $\hat{y}_{\mathrm{i}}$ represents the predicted value for the i-th sample, x_i denotes the input feature, and $w_0, w_1, \ldots, w_n$ are the regression coefficients. PR empowers us to capture complex relationships in the data, making it a valuable tool for scenarios where simple linear models fall short.

3.2.3. KNN

KNN is a simple yet effective algorithm used for both regression and classification tasks. It operates on the principle that similar data points tend to have similar outcomes. KNN assigns a data point's target value based on the majority class or the average of the K-nearest data points. The choice of K, the number of neighbors, influences the model's performance [17], [24]. Mathematically, KNN can be expressed as (3).

$$\hat{y}_i = \frac{1}{k} \sum_{j=1}^{k} y_j \tag{3}$$

In this equation, $\hat{y}_i$ is the predicted value for the i-th sample, K is the number of nearest neighbors, and $\hat{y}_i$ represents the target values of the K-nearest data points. KNN is advantageous when dealing with data that lacks a clear structure and can provide competitive results with suitable parameter tuning.

3.2.4. SVM

SVM is a powerful supervised learning algorithm used for classification and regression tasks. It aims to find a hyperplane that maximizes the margin between data points of different classes. In the context of regression, SVM seeks to find a hyperplane that best fits the data while minimizing errors. The mathematical representation of SVM regression is as (4):

$$\hat{y}_i = w_0 + w_1 x_i + \ldots + w_n x_i^n \tag{4}$$

where, $\hat{y}_{\mathrm{i}}$ represents the predicted value for the i-th sample, x_i represents the input feature, and $w_0, w_1, \ldots, w_n$ are the model parameters. SVM excels in capturing complex patterns and can handle high-dimensional data effectively.

3.2.5. DT

A DT is a versatile model used for both classification and regression tasks. It represents data as a tree structure, where internal nodes represent features, branches represent decisions based on features, and leaves represent the target values. For regression, a DT aims to create a model that minimizes the error between predicted and actual values. The model can be expressed as (5).

$$\hat{y}_i = \sum (y_i \; in \; leaf \; i \;) y_i \tag{5}$$

In this equation, $\hat{y}_i$ stands for the predicted value for the i-th sample, and y_i represents the target values in the leaf node. DTs are interpretable and suitable for capturing complex relationships in the data.

3.2.6. RF

RF is a powerful ensemble learning technique utilized in ML [25], [26]. It goes beyond individual DTs by constructing a multitude of trees during training and aggregating their outputs for robust predictions. This approach mitigates overfitting and enhances prediction accuracy by capturing complex relationships within the data [27]. Each tree in the RF ensemble is trained on a bootstrapped subset of the data and employs a random subset of features at each split. The final prediction is determined through averaging or voting across all trees. Mathematically, the RF prediction $\hat{y}$ for a given input X can be expressed as (6):

$$\hat{y} = \frac{1}{N} \sum_{i=1}^{N} f_i(x) \tag{6}$$

where N is the number of trees in the ensemble, and $f_i(x)$ represents the prediction of the i-th tree.

3.2.7. XGBoost

XGBoost, short for XGBoosting, is a gradient boosting algorithm known for its exceptional performance in various ML tasks. It leverages the power of boosting, which combines weak learners iteratively to create a robust model. XGBoost optimizes a user-defined loss function by iteratively adding DTs to the model. Each tree aims to minimize the residual error of the previous iteration, resulting in a sequence of trees that progressively improve prediction accuracy. Regularization techniques and a specialized splitting criterion further enhance XGBoost's predictive capabilities. Mathematically, the XGBoost model can be represented as (7):

$$\hat{y}_i = \phi(x_i) + \sum_{k=1}^{K} f_k(x_i) \tag{7}$$

where $\hat{y}_i$ is the predicted value for the i-th sample, $\phi(x_i)$ is the global bias term, K is the number of trees, and $f_k(x_i)$ is the prediction of the k-th tree for the i-th sample.

3.2.8. Ensemble model

In pursuit of accurate and robust precipitation prediction, we employ an ensemble model that unifies the strengths of LR, PR, RF, XGBoost, KNN, SVM, and DT models. The ensemble strategy combines the predictions of these diverse ML techniques to enhance the quality and reliability of precipitation forecasts.

Mathematically, the ensemble model can be expressed as (8).

$$\hat{y}_{EM} = \alpha_{LR}\hat{y}_{LR} + \alpha_{PR}\hat{y}_{PR} + \alpha_{KNN}\hat{y}_{KNN} + \alpha_{SVM}\hat{y}_{SVM} + \alpha_{DT}\hat{y}_{DT} + \alpha_{RF}\hat{y}_{RF} + \alpha_{XGBoost}\hat{y}_{XGBoost} \tag{8}$$

Where:
- $\hat{y}_{EM}$ is the ensemble model's predicted value for a given sample.
- $\hat{y}_{LR}$, $\hat{y}_{PR}$, $\hat{y}_{KNN}$, $\hat{y}_{SVM}$, $\hat{y}_{DT}$, $\hat{y}_{RF}$ and $\hat{y}_{XGBoost}$ represent the predictions of the individual models (LR, PR, KNN, SVM, DT, RF, XGBoost).
- α_{LR}, α_{PR}, α_{KNN}, α_{SVM}, α_{DT}, α_{RF} and $\alpha_{XGBoost}$ are the weighting coefficients applied to each model's predictions.

By combining these models, the ensemble model aims to minimize biases, reduce errors, and elevate the overall quality of precipitation forecasts. This collaborative approach exemplifies the potential of ML techniques to deliver accurate and reliable predictions, offering valuable insights into the complex interplay of meteorological factors in precipitation forecasting.

3.3. Model selection, implementation and evaluation

For this study, eight distinct ML models were carefully chosen to address various aspects of precipitation prediction. These models were specifically selected to encompass a wide array of predictive techniques, each suited to handling specific characteristics of the meteorological data. The objective was to provide a holistic analysis capable of capturing linear, nonlinear, and complex relationships within the data.
- Model selection: The selection process involved a thoughtful evaluation of the models' capabilities to manage the intricate relationships between predictors and the target variable. These considerations led to the inclusion of models with the ability to tackle non-linear correlations effectively.
- Model implementation: The selected models were implemented using popular Python libraries such as scikit-learn and XGBoost. Model implementation involved data preprocessing, feature engineering, and training to optimize their predictive capabilities.
- Model evaluation: The performance of these models was rigorously assessed using three primary metrics: mean squared error (MSE), mean absolute error (MAE), and R-squared (R^2). MSE measured the average squared difference between predicted and actual values, serving as an indicator of prediction accuracy. MAE quantified the average absolute difference between predictions and observed data. R^2 indicated the proportion of variance in the target variable that could be explained by the predictor variables. The evaluation process provided critical insights into the predictive abilities of these models. The model with the lowest MSE and MAE, along with the highest R^2 value, was chosen as the optimal model for precipitation prediction.

4. RESULTS AND DISCUSSION

To visually assess the performance of the models. We plotted the predicted precipitation values against the actual precipitation values as shown in Figure 1. This graphical representation provided valuable insights into each model's ability to capture the trends and patterns in the actual precipitation data.

Remarkably, the PR model, in particular, emerges as the standout performer. As seen in Figure 1, it exhibits a superior ability to capture intricate precipitation patterns and fluctuations, offering predictions closely aligned with the actual data. This achievement is significant, especially when considering the impact of rainfall events on agriculture, water resource management, and various sectors. PR excels in this aspect, effectively foreseeing spikes and dips in precipitation.

It is important to note that other models also provide noteworthy performance. Ensemble model, RF, and XGBoost display commendable accuracy in capturing precipitation trends and patterns, making them valuable tools for precipitation prediction. These models, too, show skill in predicting spikes and dips in actual precipitation values. In contrast, LR, KNN, SVM, and DT, while competent models, are less proficient at capturing the complex non-linear relationships present in the data. However, their use in the ensemble model contributes to enhanced accuracy.

In addition to visual analysis, we conducted a quantitative assessment of the model's performance using MAE, MSE, and R^2 values. These metrics provide a numerical measure of the predictive accuracy and overall goodness-of-fit for each model. The results, presented in Table 2, offer valuable insights into the models' performance.

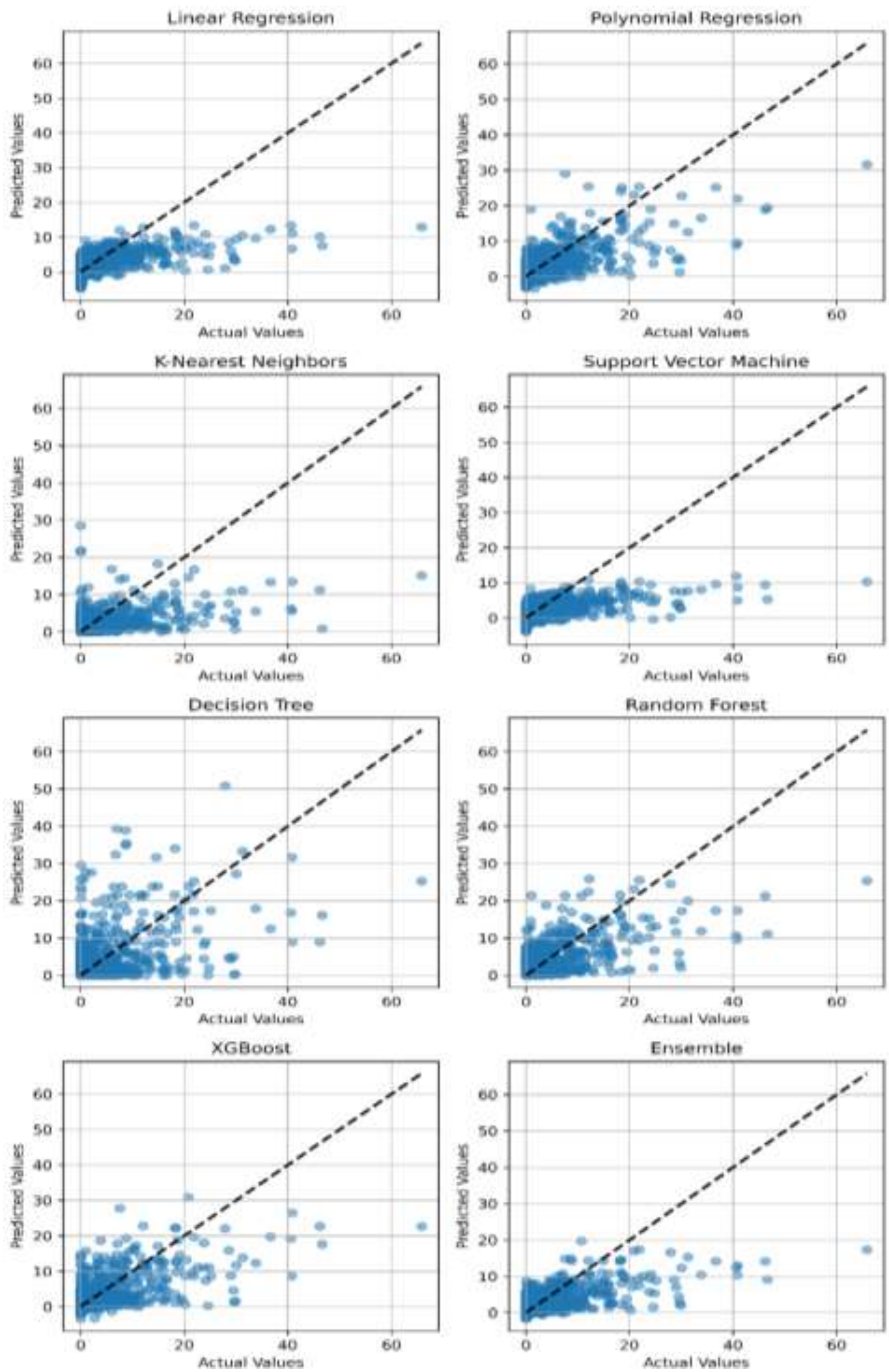

Figure 1. Comparison of predicted vs actual values for different models

Table 2. MAE, MSE, and R^2 values of the models

Model	MAE	MSE	R^2 value
LR	1.747	10.852	0.326
PR	1.102	7.880	0.510
KNN	1.215	11.587	0.280
SVM	1.590	11.321	0.297
DT	1.297	14.028	0.129
RF	1.084	8.091	0.497
XGBoost	1.133	8.572	0.467
Ensemble	1.239	8.051	0.463

Based on these metrics, the RF model demonstrated the most accurate predictions with the lowest MAE (1.084), followed closely by the PR and XGBoost models. The ensemble model also performed remarkably well, highlighting the advantage of integrating multiple models to enhance prediction accuracy. In terms of MSE, the PR model achieved the lowest value of 7.880, reinforcing its ability to minimize prediction errors. The ensemble model exhibited the second-best MSE, further emphasizing the strength of combining multiple models. The R^2 values provide insights into the proportion of variance in the target variable explained by each model. Here, the PR model achieved the highest R^2 value of 0.510, indicating its capability to explain a substantial portion of the variance in precipitation data. The results offer valuable insights for developing more accurate precipitation prediction systems. The models, particularly the PR, RF, XGBoost, and ensemble models, hold the potential to significantly enhance precipitation forecasting, benefiting various stakeholders such as farmers, water resource managers, and disaster preparedness agencies.

5. CONCLUSION

This work compares the performance of various ML models for precipitation forecasting using meteorological data for Casablanca, Morocco. Casablanca is a city that faces challenges in water management and climate change adaptation and needs accurate rainfall forecasts for various applications. We used a dataset containing 42 years of weather data and eight different ML models and evaluated them using MAE, MSE, and R2 values. The results showed that PR, RF, XGBoost, and ensemble models performed better than the other models, achieving the lowest errors and the highest R^2 values. This work demonstrates the potential of ML models for precipitation forecasting and identifies the best-performing model for practical applications. Future work could involve exploring other ML models and techniques and expanding the study to include more meteorological variables.

REFERENCES

[1] J. Kang, H. Wang, F. Yuan, Z. Wang, J. Huang, and T. Qiu, "Prediction of precipitation based on recurrent neural networks in jingdezhen, jiangxi province, China," *Atmosphere (Basel)*, vol. 11, no. 3, Mar. 2020, doi: 10.3390/atmos11030246.

[2] H. Wu, R. F. Adler, Y. Hong, Y. Tian, and F. Policelli, "Evaluation of global flood detection using satellite-based rainfall and a hydrologic model," *J Hydrometeorol*, vol. 13, no. 4, pp. 1268–1284, Aug. 2012, doi: 10.1175/JHM-D-11-087.1.

[3] J. A. Anochi, V. A. de Almeida, and H. F. de Campos Velho, "Machine learning for climate precipitation prediction modeling over South America," *Remote Sens (Basel)*, vol. 13, no. 13, Jul. 2021, doi: 10.3390/rs13132468.

[4] S. Q. Dotse, I. Larbi, A. M. Limantol, and L. C. De Silva, "A review of the application of hybrid machine learning models to improve rainfall prediction," *Modeling Earth Systems and Environment*. Springer Science and Business Media Deutschland GmbH, 2023. doi: 10.1007/s40808-023-01835-x.

[5] S. Aftab, M. Ahmad, N. Hameed, M. S. Bashir, I. Ali, and Z. Nawaz, "Rainfall prediction in lahore city using data mining techniques," *International Journal of Advanced Computer Science and Applications*, 2018, [Online]. Available: www.ijacsa.thesai.org

[6] S. Aftab, M. Ahmad, N. Hameed, M. S. Bashir, I. Ali, and Z. Nawaz, "Rainfall prediction using data mining techniques: a systematic literature review," *International Journal of Advanced Computer Science and Applications* 2018, [Online]. Available: www.ijacsa.thesai.org

[7] M. A. Nayak and S. Ghosh, "Prediction of extreme rainfall event using weather pattern recognition and support vector machine classifier," *Theor Appl Climatol*, vol. 114, no. 3–4, pp. 583–603, 2013, doi: 10.1007/s00704-013-0867-3.

[8] T. Yue, S. Zhang, J. Zhang, B. Zhang, and R. Li, "Variation of representative rainfall time series length for rainwater harvesting modelling in different climatic zones," *J Environ Manage*, vol. 269, Sep. 2020, doi: 10.1016/j.jenvman.2020.110731.

[9] N. Mishra, H. K. Soni, S. Sharma, and A. K. Upadhyay, "A comprehensive survey of data mining techniques on time series data for rainfall prediction," *Journal of ICT Research and Applications*, vol. 11, no. 2, pp. 167–183, 2017, doi: 10.5614/itbj.ict.res.appl.2017.11.2.4.

[10] A. U. Rahman *et al.*, "Rainfall prediction system using machine learning fusion for smart cities," *Sensors*, vol. 22, no. 9, May 2022, doi: 10.3390/s22093504.

[11] "Regional directorate of casablanca settat." Accessed: Dec. 06, 2023. [Online]. Available: www.hcp.ma

[12] J. Yang, Y. Xiang, J. Sun, and X. Xu, "Multi-model ensemble prediction of summer precipitation in china based on machine learning algorithms," *Atmosphere (Basel)*, vol. 13, no. 9, 2022, doi: 10.3390/atmos13091424.

[13] C. Z. Basha, N. Bhavana, N. Bhavana, and N. Bhavana, "Rainfall prediction using machine learning and deep learning techniques," in *Proceedings of the International Conference on Electronics and Sustainable Communication Systems (ICESC 2020)*, 2020.

[14] M. S. Balamurugan and R. Manojkumar, "Study of short-term rain forecasting using machine learning based approach," *Wireless*

Networks, vol. 27, no. 8, Nov. 2019, doi: 10.1007/s11276-019-02168-3.
[15] M. Huang, R. Lin, S. Huang, and T. Xing, “A novel approach for precipitation forecast via improved K-nearest neighbor algorithm,” *Advanced Engineering Informatics*, vol. 33, pp. 89–95, Aug. 2017, doi: 10.1016/j.aei.2017.05.003.
[16] F. Di Nunno, F. Granata, Q. B. Pham, and G. de Marinis, “Precipitation forecasting in Northern Bangladesh using a hybrid machine learning model,” *Sustainability (Switzerland)*, vol. 14, no. 5, Mar. 2022, doi: 10.3390/su14052663.
[17] K. Ahmed, D. A. Sachindra, S. Shahid, Z. Iqbal, N. Nawaz, and N. Khan, “Multi-model ensemble predictions of precipitation and temperature using machine learning algorithms,” *Atmos Res*, vol. 236, May 2020, doi: 10.1016/j.atmosres.2019.104806.
[18] J. A. Anochi, V. A. de Almeida, and H. F. de Campos Velho, “Machine learning for climate precipitation prediction modeling over South America,” *Remote Sens (Basel)*, vol. 13, no. 13, Jul. 2021, doi: 10.3390/rs13132468.
[19] Y. Zhang and A. Ye, “Machine learning for precipitation forecasts postprocessing: multimodel comparison and experimental investigation,” *J Hydrometeorol*, 2021, doi: 10.1175/JHM-D-21.
[20] “NASA langley research center,” Accessed: Aug. 07, 2023. [Online]. Available: https://power.larc.nasa.gov/
[21] K. G. Liakos, P. Busato, D. Moshou, S. Pearson, and D. Bochtis, “Machine learning in agriculture: a review,” *Sensors (Switzerland)*, vol. 18, no. 8. MDPI AG, Aug. 14, 2018. doi: 10.3390/s18082674.
[22] L. Benos, A. C. Tagarakis, G. Dolias, R. Berruto, D. Kateris, and D. Bochtis, “Machine learning in agriculture: a comprehensive updated review,” *Sensors*, vol. 21, no. 11. MDPI AG, Jun. 01, 2021. doi: 10.3390/s21113758.
[23] W. T. Abebe and D. Endalie, “Artificial intelligence models for prediction of monthly rainfall without climatic data for meteorological stations in Ethiopia,” *J Big Data*, vol. 10, no. 1, Dec. 2023, doi: 10.1186/s40537-022-00683-3.
[24] S. Zhang, X. Li, M. Zong, X. Zhu, and R. Wang, “Efficient kNN classification with different numbers of nearest neighbors,” *IEEE Trans Neural Netw Learn Syst*, vol. 29, no. 5, pp. 1774–1785, May 2018, doi: 10.1109/TNNLS.2017.2673241.
[25] C. M. Liyew and H. A. Melese, “Machine learning techniques to predict daily rainfall amount,” *J Big Data*, vol. 8, no. 1, Dec. 2021, doi: 10.1186/s40537-021-00545-4.
[26] C. D. Usman, A. P. Widodo, K. Adi, and R. Gernowo, “Rainfall prediction model in Semarang City using machine learning,” *Indonesian Journal of Electrical Engineering and Computer Science (IJEECS)*, vol. 30, no. 2, pp. 1224–1231, May 2023, doi: 10.11591/ijeecs.v30.i2.pp1224-1231.
[27] H. Jdi and N. Falih, “Precipitation forecasting using machine learning in the region of Beni Mellal-Khenifra,” *Indonesian Journal of Electrical Engineering and Computer Science (IJEECS)*, vol. 31, no. 1, pp. 451–458, Jul. 2023, doi: 10.11591/ijeecs.v31.i1.pp451-458.

BIOGRAPHIES OF AUTHORS

Abdelouahed Tricha received his M.Sc. degree in smart industry from the Faculty of Sciences Dhar El Mahraz, University Sidi Mohamed Ben Abdellah, Fez, Morocco, in 2021. He is currently a Ph.D. student at the National School of Electrical and Mechanical Engineering (ENSEM), Casablanca, Morocco. His research interests include the internet of things (IoT), machine learning, and embedded systems. He can be contacted at email: tricha.abdelouahed@gmail.com.

Dr. Laila Moussaid is a professor of computer engineering at the ENSEM engineering school in Morocco. She was the coordinator of the Master in BIG DATA between 2015-2018. From 2019, she was elected head of the computer engineering department of the National School of Electrical and Mechanical Engineers of the Hassan II University of Casablanca, Morocco. In 2020, she was elected head of the E2IT team at the LRI laboratory, also coordinator of the “data engineering and artificial intelligence” sector. since 2018, she has been a member of the scientific committees and organizers of numerous international conferences. Professor Laila Moussaid’s main research focuses on data, IoT, embedded systems, smart city, and digitalization. She can be contacted at email: l.moussaid@yahoo.com.