Modeling Earth Systems and Environment (2024) 10:19–44
https://doi.org/10.1007/s40808-023-01835-x

REVIEW ARTICLE

A review of the application of hybrid machine learning models to improve rainfall prediction

Sam-Quarcoo Dotse[1] · Isaac Larbi[1] · Andrew Manoba Limantol[1] · Liyanage C. De Silva[2]

Received: 3 April 2023 / Accepted: 28 June 2023 / Published online: 11 July 2023

Abstract
Rainfall is one of the most important meteorological phenomena that impacts many fields, including agriculture, energy, water resources management, and mining, among others. While machine learning (ML) models have shown great potential in rainfall forecasting as they perform well and sometimes better than some physical models, the complex physical processes involved in rainfall formation make single ML models insufficient for providing accurate rainfall estimates in most cases. Although there are comprehensive reviews of the performance evaluation of individual ML models in the literature, only a limited number of reviews exist that include hybrid models that specifically focus on rainfall forecasting. This paper presents an extensive review of the performance of hybrid ML models for rainfall forecasting. The vital information on the forecasting time scales, model inputs, and evaluation methods used for constructing these models has been analysed and discussed. The findings revealed that hybrid ML models composed by integrating data pre-processing techniques and optimisation algorithms may be a successful and efficient solution to enhance rainfall predictions at various timescales. Hybrid ML models used for rainfall predictions are capable of producing comparatively more accurate forecasts and reducing uncertainty for both short and longer lead times. Recent advances in physical-ML hybrid models for weather forecasting have also been highlighted. Overall, this review article provides useful information to researchers interested in developing early warning systems for precise and timely rainfall forecasting.

Keywords Rainfall forecasting · Machine learning · Hybrid models · Optimisation · Data pre-processing

Abbreviations

NWP	Numerical weather prediction
ARMA	Auto-regressive moving average
ARIMA	Auto regressive integrated moving average
SARIMA	Seasonal auto regressive integrated moving average
PARMA	Periodic autoregressive moving average
AI	Artificial intelligence
ML	Machine learning
LR	Linear regression
ANN	Artificial neural networks
RMSE	Root Mean Square Error
R (or CC)	Correlation coefficient
R^2	Coefficient of determination
CE (or NSE)	Nash–Sutcliffe efficiency coefficient
MAE	Mean absolute error
G	Gamma coefficient
AARE	Average absolute relative error
MaxAE	Maximum absolute error
MinAE	Minimum absolute error
MSE	Mean square error
NMSE	Normalized mean square error
% error	Percent error
EV	Volumetric error
RMEP	Root mean error in probability
SS	Skill score
SNR	Signal to noise ratio
RMAE	Root mean absolute error
APE	Average performance error
PP	Performance parameter
MAPE	Mean absolute percentage error

✉ Sam-Quarcoo Dotse
sqdotse@uesd.edu.gh

1 School of Sustainable Development, University of Environment and Sustainable Development, Private Mail Bag, Somanya, Ghana

2 School of Digital Science, Universiti Brunei Darussalam, Jalan Tungku Link BE1410, Brunei Darussalam

CrossProb	Crossover probability
ProbMut	Probability of mutation
HK	Hanssen kuipers
KRR	Kernel Ridge Regression
AARE	Average absolute relative error
WI	Willmott's index
RAE	Relative absolute error
PI	Persistence index
IA	Index of agreement (IA),
RMRSE	Root mean relative square error
POD	Probability of Detection
CSI	Critical Success Index
FAR	False Alarm Ratio
RMSEP	Root mean square error of prediction
CV	Coefficient of variation
RSR	Root mean square error
RRMSE	Relative root mean square error
ANFIS	Adaptive neuro-fuzzy inference system
WOA	Whale Optimisation Algorithm
MLR	Multiple linear regression
MLP	Multi-layer perceptron network (MLP)
MANFIS	Modified adaptive neuro-fuzzy inference system
WNN	Neural-wavelet or Wavelet neural network
WSVM	Wavelet-based support vector machines
WGRNN	Wavelet-based generalized regression neural networks
GRNN	Generalized regression neural networks
CWT-MP	Continuous wavelet–multilayer perceptron
DWT-MP	Discrete wavelet–multilayer perceptron
MP	Multilayer perceptron
MODWT	Maximum overlap discrete wavelet transform
SSA	Singular spectrum analysis
TVF-EMD	Time-varying filter-based empirical mode decomposition
REMD	Robust empirical mode decomposition
CEEMD	Complementary ensemble empirical mode decomposition
WT	Wavelet transform
ESMD	Extreme-point symmetric mode decomposition
ENN	Elman neural network
SVR	Support vector regression
SVM	Support vector machines
SAS-MP	Season-multilayer perceptron
W-MP	Wavelet-multilayer perceptron
FCM	Fuzzy C-Means
ISMR	Indian summer monsoon rainfall
GA	Genetic algorithm
PSO	Particle Swarm Optimisation
CPSO	Chaotic particle swarm optimisation algorithm
WOA	Whale optimisation algorithm
RNNs	Recurrent artificial neural networks
ACOR	Ant colony optimization for the continuous domain
DE	Differential evolution
FFA	Firefly Optimisation algorithm
MI	Mutual information (MI)
WGEP	Wavelet genetic programming
WNF	Wavelet-neuro-fuzzy techniques
GEP	Gene expression programming
MANN	Modular artificial neural network
KNN	K-nearest-neighbors (KNN)
MA	Moving average
EANN	Emotional artificial neural network
FFNN	Feedforward neural networks
AEEMD	Adaptive Ensemble Empirical Mode Decomposition
GEP-ARCH	Gene expression programming-autoregressive conditional heteroscedasticity
ANN-ARCH	Artificial neural networks-autoregressive conditional heteroscedasticity
ELM	Extreme learning machine
CLA	Learning-Cellular Automation
WSVR	Wavelet-support vector machine
GP	Genetic programming
MOGA	Multi-objective genetic algorithm
EEMD	Ensemble empirical mode decomposition
SAM	Seasonal adjustment method
PI	Projection pursuit technology
BP-NN	Backpropagation neural network
AC	Ant Colony Algorithm
PSR	Phase space reconstruction
LS-SVR (or LSSVR)	Least-square support vector regression
IBOA	Improved butterfly optimisation algorithm
VMD	Variational mode decomposition
HW	Hammerstein–Weiner
RF	Random forests
CART	Classification and regression trees

MPSA	Multi-period simulated annealing
BN	Bayesian networks
GARCH	Generalized autoregressive conditional heteroscedasticity
MGGP	Multigene genetic programming

Introduction

Rainfall forecasting remains a difficult task due to the processes involved in this meteorological phenomenon. The highly spatio-temporal variability of rainfall distribution also makes its forecasting an extremely challenging task. Nonetheless, reliable prediction of this meteorological variable at any forecasting horizon (seasonal, monthly, daily, etc.) is critical, especially given the present global threat posed by climate change. Rainfall is one of the most important meteorological phenomena, influencing several economic sectors such as agriculture (Trinh 2018), hydropower generation (Haddad 2011), and water resource management (Hartmann et al. 2016). Precise and timely rainfall forecasting is critical for preventing and mitigating the consequences of natural disasters such as landslides, floods, and droughts (Pham et al. 2020). So far, several methods are available for rainfall forecasting, including numerical weather prediction (NWP), statistical, and machine learning (ML) or artificial intelligence (AI). They are generally divided into two major groups: physically based and data-driven approaches. NWP models, which are physics-based models, are the fundamental approach employed by meteorological agencies around the world. The approach involves the use of numerical methods to solve a set of equations governing the atmospheric processes responsible for rainfall to obtain forecasts. NWP models greatly enhance our understanding of the underlying complex interplay of cloud dynamics and microphysical processes involved in rainfall formation. However, although they have shown great capabilities for weather predictions (Bauer et al. 2015), they are sometimes associated with many uncertainties, in model inputs, parameterization, and the numerical algorithms (Schultz et al. 2021). Several studies have suggested gaps in the prediction capability of physical models. For example, due to a variety of issues such as a lack of region-specific parameterizations and data availability, NWP models are unable to correctly reproduce precipitation patterns in South America (Anochi et al. 2021). They typically do not perform well at forecasting medium-term rainfall (Vaze et al. 2011).

The statistical and AI methods are data-driven approaches, as predictions are made based on the analysis of historical data. Unlike physics-informed models, the data-driven methods do not require a detailed understanding of the atmospheric processes but only capture the underlying physical behaviour by identifying trends in the historical data. In that, forecasts are made after an appropriate relationship believed to be the main influence of rainfall at the location has been obtained empirically from the historical data. Historically, the main approach to seasonal and monthly rainfall forecasting was by means of classical statistical methods such as auto-regressive moving average, periodic autoregressive moving average, etc. They are mostly linear models with limited capacity to represent highly nonlinear features of rainfall time series. AI and ML approaches have over the last decades been successfully applied in rainfall forecasting to overcome the limitations of classical statistical models. Besides performing better than the statistical models due to their nonlinear nature, these AI techniques have shown great potential in rainfall modelling as they perform well and sometimes better than some physics-based models. Comparative research has shown that AI-based models may generate reliable results in precipitation predictions with regard to the physically-based models (e.g., Abbot and Marohasy 2012; Jiao et al. 2016; Anochi et al. 2021). Anochi et al. (2021) demonstrated that ML models are capable of operational forecasts with higher performance than the BAM model, which is presently employed at the National Institute for Space Research (INPE) in Brazil, and with the added benefit of not requiring supercomputers to conduct these forecasts.

The application of AI methods has progressed greatly and has now become a well-established discipline in weather and climate research. There is renewed interest in its application to enhance the understanding and improve the prediction skills of the physically-based models that are constrained by atmospheric physics and chemistry (e.g., O'Gorman and Dwyer 2018; Anderson and Lucas 2018; Scher 2018). Undoubtedly, the recent breakthroughs in ML and deep learning (DL) have led to a rise in interest in using these evolving technologies across a variety of fields (LeCun et al. 2015; McGovern et al. 2019), including weather prediction and climate modelling, potentially in conjunction with numerical models or in a standalone way (Scher 2020).

The physical processes involved in the formation of rainfall typically consist of a number of sub-processes, making it sometimes impossible to simulate them accurately using a single global model (Solomatine and Ostfeld 2008; Sumi et al. 2012). Several studies have shown that single models, whether physically based or data-driven, are not so accurate for rainfall forecasting. Thus, to overcome the limitations of ML models in order to improve the accuracy and reliability of the forecasts, the last two decades have seen an increasing trend towards combining or coupling different models. In addition to the introduction of new or advanced ML methods like DL, the combination of models has proven to be a more accurate and efficient approach. The basic goal of model combination is to capture diverse patterns in data by utilizing the unique qualities of the separate models. The approach of combining several models can be broadly categorized into hybrid or ensemble modelling. In ensemble modelling, several different or similar models are built for the same process and then integrated together to make the final prediction. On the other hand, hybrid modelling involves a combination of different models, either a physical and AI or two or more different AI models. Several recent studies on rainfall forecasting have shown that AI-based ensemble modelling (e.g., Nourani et al. 2019a) and hybridization of two or more ML models (e.g., Jiang and Wu 2013) provided more robust and efficient models and better accuracy than typical single models.

There has been an increase in the application of hybrid ML techniques for rainfall forecasting. Although comprehensive reviews of performance evaluation of individual ML models in literature exist, only a limited number of studies on hybrid models exist with no specific focus on rainfall forecasting. Additionally, the reviews mostly focus on artificial neural networks (ANN) and their variants. Fahimi et al. (2017) have provided a comprehensive review of hybrid models based on soft computing used in the modelling of hydrological variables, which included some articles on ANN-based hybrid models for rainfall prediction. Nourani et al. (2014) have also only reviewed hybrid Wavelet-Artificial Intelligence models in hydrology with no particular focus on rainfall. A study on the evaluation of different hybrid models applied in rainfall prediction is very necessary to provide useful information that will further enrich the existing literature that stressed the importance of hybrid models in enhancing ML-based models' accuracy and reliability. To the best of our knowledge, only limited studies are available, providing a comprehensive evaluation of hybrid ML models in rainfall forecasting that includes not only ANN hybrid models but also other ML hybrid models used in conjunction with numerical models. As an important meteorological variable that directly or indirectly impacts several sectors of the economies of countries, such important information will be valuable in developing an early warning system. This paper presents an extensive review of the performance of hybrid ML models for rainfall forecasting.

Methodology

A literature search was conducted on hybrid ML for rainfall forecasting from major databases, including Google Scholar, Scopus, MDPI, IEEE Xplore, and Science Direct. Published studies from 2010 to 2021 were examined in this review paper. The review does not include hybrid deep learning and AI-based ensemble methods. Additionally, the search specifically focused on rainfall and does not include other hydrological variables. The review paper analysed and discussed the following:

- ML models for forecasting rainfall, including the forecasting time scales (forecast horizon), model inputs, and evaluation methods.
- The performance of hybrid ML models in comparison with the standalone models for rainfall forecasting.
- Recent advances in physics-ML hybrid models for weather forecasting.

Machine learning models for rainfall forecasting

As a subfield of AI, ML covers algorithms and methods that can learn from datasets, recognize patterns, and make judgments without being explicitly programmed. Based on the problem settings and the approach, ML algorithms may be classified into three types: supervised, unsupervised, and reinforcement learning. In supervised learning, an algorithm learns how to map inputs to desired outputs given a training dataset of input–output pairs. The algorithm is then trained, often by minimizing a cost or loss function, to provide output predictions that are as close as possible to the training targets. The most frequent ML technique is supervised learning, which encompasses tasks like time series forecasting and the relationship between two or more physical observations, as well as more abstract problems like image recognition (Scher 2020).

On the other hand, unsupervised learning extracts hidden patterns or features from datasets without the requirement for labeled input–output pairings. Dimensionality reduction approaches, clustering algorithms, and generative learning are a few examples of unsupervised learning. In reinforcement learning, a specific goal is pursued, and the learning process is guided by feedback from the environment (Bishop and Nasrabadi 2006). While significant in other settings, such as recent developments in computer gaming skills (Silver et al. 2017), reinforcement learning is of little significance in the domain of weather and climate forecasting (Scher 2020).

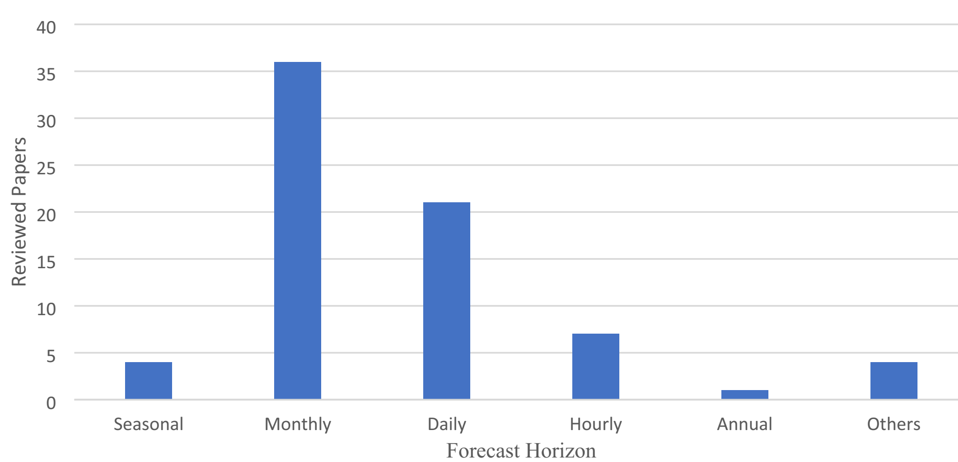
Over the last few decades, ML has shown great potential in modeling and forecasting rainfall, providing comparable results with physical models. ML models are better equipped to handle the complexity and nonlinearity of rainfall processes. Several ML algorithms have been applied to improve prediction accuracy and confidence at a low computational cost. The most widely used ML models in rainfall forecasting include support vector machine (SVM), random forest (RF), artificial neural networks (ANN), and their variants. However, despite their popularity and applications, several of these models still exhibit varying degrees of accuracy and imprecision in their rainfall predictions. Neural networks and their variants are known to have instability issues, making them unreliable (Breiman 1996) and resulting in poor generalization of rainfall prediction. It is crucial to keep in mind that models like ANN and SVM include unmeasured parameters whose values need to be determined before forecasting precipitation (Yadav and Sagar 2019). Additionally, these models are considered black-box models since there is no formal structure for explaining and analyzing the link between inputs and outcomes. Recent efforts have been made in interpreting and explaining black-box ML models (Molnar 2020). There are also various input variable selection techniques used during the data preprocessing stage to identify relevant input variables for model development in order to overcome some of these difficulties. The recent introduction of novel ML models, hybridization of existing ones, and ensemble modeling have proven to be more effective in overcoming the shortcomings or deficiencies of single ML models. The emphasis so far has been on hybrid models and novel ML or deep learning models. As provided in "Hybrid machine learning models for rainfall forecasting" section, this paper aims to examine and assess the effectiveness of hybrid ML models used in rainfall forecasting. The focus is not on the model architectures but on the performance vis-a-vis single models. The accuracy or performance assessment of different models for rainfall forecasting depends on many factors, including the forecasting horizons (time scale), model inputs (input determination), and evaluation methods, as explained in the following subsections.

Forecasting horizon

Weather and precipitation forecasts are made at different timescales to serve a variety of end users. The World Meteorological Organization (www.wmo.int) specifies meteorological forecasting ranges or time horizons, including Nowcasting with a time horizon of 0 to 2 h, Very short range with a time horizon up to 12 h, Short-range weather forecasting with a time horizon beyond 12 h and up to 72 h, Medium-range weather forecasting with a time horizon beyond 72 h and up to 240 h, Extended-range weather forecasting ranging beyond 10 days and up to 30 days (usually averaged and expressed as a departure from climate values for that period), and Long-range forecasting from 30 days up to two years. Studies have shown that, keeping forecast models and other parameters constant, the accuracy of forecasts made by ML methods usually varies significantly according to the forecast horizon or the period of prediction. Precipitation varies greatly both regionally and temporally, making precipitation forecasts with a lead time of more than a few days highly challenging (Abdusselam Altunkaynak et al. 2016). Compared to other types of forecasts, short-range forecast predictions provided for today or tomorrow (up to 48 h) are typically more accurate (Bushara and Abraham 2015). Weather forecast accuracy decreases significantly after 10 days, according to research (Jagannathan 1974). Yet, for specific applications, it is necessary to forecast precipitation at different timescales.

For many practical applications in watershed management, such as agricultural production, mine operations, drought and flood prevention measures, medium-to-long-term rainfall forecasts are necessary (Mehr et al. 2019). Rainfall predictions, both short-term and long-term, are critical components of water resource management for specific purposes. A one-month forecast or long-term timeframe is beneficial for a number of applications, including food production, environmental protection, drought control, and optimal reservoir operation. Real-time reservoir operation, flood warning systems, hydrologic models, and real-time management of wind and solar resources require short-term forecasts with lead times of a few hours to a day (Danandeh Mehr et al. 2019). In this survey, as seen in Fig. 1, most researchers prefer to use hourly, daily, monthly, and annual time scales to predict rainfall. Monthly and seasonal time scales are extremely important as they provide valuable information for agricultural planning, water resource management, and related crop insurance applications (Garbrecht

Fig. 1 Forecasting Horizon in the reviewed papers.



et al. 2010). Precise daily precipitation prediction is necessary for accurate streamflow forecasts, flooding risk assessments, and the establishment of a dependable flood control and early warning system (Altunkaynak and Nigussie 2015). Figure 1 depicts the different time scales used in the current review article, including hourly, daily, monthly, seasonal, and yearly. According to the evaluated publications, the two most often used time scales are daily and monthly.

Model inputs selection

Data-driven methods like ML models require a set of input data (predictors) in order to predict the output. The selection of adequate inputs is crucial in rainfall forecasting, particularly when using AI approaches. The performance and accuracy of the model depend on the input variables, making the selection of inputs a major consideration in rainfall predictive modeling. Therefore, it is essential to choose the most efficient input parameters and thoroughly consider them during the modeling process. Since rainfall varies considerably both temporally and spatially, inputs should be selected based on an understanding of the physical mechanism responsible for rainfall in the specific study area. Choosing the right number and type of input variables is essential in rainfall forecasting, even though any relevant set of input variables can be employed in model development. In addition, the type, number, or size of input variables used in developing the models is also important in rainfall forecasting. While there are an infinite number of possible inputs, it is necessary to select a smaller subset of potentially relevant variables related to rainfall for the study location. This improves the efficiency of the training process by minimizing the running time and uncertainty associated with generalization.

Several meteorological and climatic indices have been used as potential inputs in the cited literature. Lagged (past) rainfall records have been extensively used as model inputs for rainfall forecasting (e.g., Jiang and Wu 2013; El-Shafie et al. 2011; Kalteh 2017). Other exogenous factors, including temperature, radiation, atmospheric pressure, relative humidity, average wind speed, the number of stormy days, the number of snowy days, and the number of overcast days, have also been employed, either separately or in combination, to forecast rainfall (Pham et al. 2019). It is well known that several large-scale climatic modes significantly affect precipitation around the globe, as numerous studies have established the relationship between several climate indices and daily, monthly, and seasonal rainfall occurrence worldwide (Mekanik et al. 2013). Therefore, lagged values of several climate drivers or indices, including El Nino Southern Oscillation, Madden–Julian Oscillation, Subtropical Ridge (STR), etc., have been used as inputs to forecast rainfall at different timescales (Mekanik et al. 2013; He et al. 2015). The large-scale climatic indices are used either alone or in combination with lagged rainfall. Additionally, optical, microphysical, and textural cloud parameters generated from the multiple channels of MSG SEVIRI's high spatial and temporal resolutions have been utilized as possible inputs (Sehad et al. 2017). Tables 1 and 2 provide the various inputs used in this current review paper.

Evaluation methods

Model evaluation is conducted at various stages of model development, including during the training, validation, and testing processes. This essentially involves measuring the forecasting accuracy or reliability of the model by comparing the measured or actual values with the predicted or model outputs. In general, evaluating models requires considering factors such as accuracy, uncertainty, generalization ability, longer lead time, speed, and computing cost of implementation. The effectiveness of a model can be assessed using a variety of statistical indicators or evaluation techniques. Figure 2 shows the assessment criteria and the frequency of their usage in the literature. The survey identified a total of thirty-seven (37) types of evaluation methods. However, the majority of the reviewed publications were assessed using the Nash–Sutcliffe coefficient (CE), coefficient of determination (R^2), correlation coefficient (R), root mean square error (RMSE), and mean absolute error (MAE).

The prediction accuracy is measured by RMSE, which yields a positive number by squaring the errors. A lower RMSE indicates better predictions, with 0 indicating excellent predictions. MAE also measures the error in predictions and provides information on the goodness-of-fit for the distribution of estimate errors. A smaller MAE value indicates a better agreement between observed and predicted values. The correlation coefficient (R) ranges from −1 to 1 and represents the strength of the linear relationship between the observed and predicted values. The Nash–Sutcliffe coefficient (CE) is used to compare the goodness of fit between the observed and predicted values, and it produces a positive number. The coefficient of determination (R^2) is utilized to measure the modeling accuracy and explains how many data points are covered by the regression line. A higher R^2 indicates that the regression line passes through a greater proportion of points when plotted together with the data. A value of 1 or 0 indicates that the regression line captures all or none of the data, respectively. A larger CE value indicates a greater goodness of fit for the observations.

The usefulness of R^2 lies in its ability to estimate the likelihood of future events falling within the predicted outcomes. While R or R^2 is commonly used as a "goodness-of-fit" or relative error measure, the CE serves as an excellent substitute since it is sensitive to both the mean and variance of the observed and predicted values. Additionally, Legates

Table 1 Summary of ANN-hybrid models

Authors (Year)	Hybrid methods	Input variables	Forecast horizon	Location/country	Performance metrics	Comments
El-Shafie et al. (2011)	ANFIS	Previous rainfall data	Monthly	Malaysia	RMSE, R^2, Nash–Sutcliffe coefficient (CE), gamma test (G), R	ANFIS (RMSE:0.051, CE:0.991, R^2:0.906, R:0.896) gives high performance compared with single ANN (RMSE:0.740, CE:0.980, R^2:0.788, R:0.785); high accuracy and low error
Jeong et al. (2012)	ANFIS	Precipitation, average cloud quantity, solar light time, lowest temperature, average temperature	Monthly	Korea	RMSE, Maximum (MaxAE) and Minimum (MinAE) absolute errors, average absolute relative error (AARE)	ANFIS provides good and competitive performance
Petković et al. (2016)	ANFIS	Month in a year precipitation	Monthly	Serbia	Graphical plots	Results showed the effectiveness of the model
Choubin et al. (2016)	ANFIS	Large scale climate signals	Seasonal	Iran	RMSE, MAE, R	ANFIS and MLP performed better than MLR. MLP also performed better in the study area, though ANFIS better during training
Sojitra et al. (2015)	ANFIS	Wet bulb temperature, mean temperature, relative humidity, evaporation, previous moving average week	Daily	India	R, Mean square error (MSE), Normalized mean square error (NMSE), percent error (% error), volumetric error (EV), coefficient efficiency (CE)	Model performs well
Bushara and Abraham (2015)	ANFIS	Date, minimum temperature, relative humidity, wind direction	Long-term	Sudan	R, MAE, RMSE	ANFIS performed better than the single ANN model; giving better and more accurate results
Mekanik et al. (2016)	ANFIS	Large scale climate signals	Seasonal	Australia	RMSE, MAE, R, root mean error in probability (RMEP), skill score (SS)	ANFIS (R:0.530, RMSE:16.700, MAE:14:000) gives high performance compared with single ANN (R:0.210, RMSE:20:600, MAE:19.200)

Table 1 (continued)

Authors (Year)	Hybrid methods	Input variables	Forecast horizon	Location/country	Performance metrics	Comments
Azad et al. (2019)	ANFIS ANFIS-PSO, ANFIS-GA, ANFIS-DE, ANFIS-ACOR	Monthly rainfall	Monthly	Iran	SI, RMSE, R^2	The suggested hybrid models outperformed the standard ANFIS (SI:0.900, RMSE:9.980, R^2:0.330. ANFIS-ACOR (SI:0.210, RMSE:2.730, R^2:0.920) outperformed all models.
Akrami et al. (2013)	MANFIS	Lag monthly rainfall data	Monthly	Malaysia	RMSE, signal to noise ratio (SNR), R^2, root mean absolute error (RMAE)	MANFIS performed better with higher accuracy, low errors and lower computational complexity compared with ANFIS model
Venkata Ramana et al. (2013)	WNN	Monthly rainfall, minimum temperature, maximum temperature	Monthly	India	RMSE, R, CE	WNN (RMSE:63.010, R:0.974, CE:94.780) more effective than single ANN (RMSE:163.790, R:0.807, CE:64.730) and AR (RMSE:221.820, R:0.642, CE:34.910) models.
Solgi et al. (2014)	WNN, ANFIS	Rainfall time series	Daily	Iran	CE, RMSE	WNN (RMSE:0.021, R^2:0.903, CE:0.743) performs better than ANFIS (RMSE:0.027, R^2:0.613, CE:0.562).
Akrami et al. (2014)	Wavelet-neuro-fuzzy (W-ANFIS), ANN-Wavelet, ANFIS	Monthly rainfall	Monthly	Malaysia	RMSE, R^2, gamma coefficient (G), R	W-ANFIS (RMSE:0.041, R^2:0.982, R:0.981) performed better than ANN (RMSE:0.740, R^2:0.788, R:0.785) and the other ANFIS (RMSE:0.051, R^2:0.906, R:0.896) and ANN-Wavelet (RMSE:0.560, R^2:0.888, R:0.882) hybrid models; higher accuracy and lower error.
Kim et al. (2016)	WGRNN & WSVM	Hourly rainfall data	Aggregation of hourly rainfall	Republic of Korea	R, RMSE, CE, MAE, average performance error (APE)	WGRNN & WSVM performed better that the standalone SVM and GRNN models

Table 1 (continued)

Authors (Year)	Hybrid methods	Input variables	Forecast horizon	Location/country	Performance metrics	Comments
Altunkaynak and Ozger (2016)	DWT-MP & CWT-MP	Daily precipitation	Daily	Turkey	MSE, CE, SS	Both CWT-MP (MSE:56.370, CE:0.950) and DWT-MP (MSE:122.350, CE:0.880) show significantly improved performance than ANN (MSE:1002.980, CE:0.030). CWT-MP performed better than the DWT-MP
Shafaei et al. (2016)	wavelet-SARIMA-ANN	Monthly precipitation	Monthly	Iran	R^2, R, RMSE	wavelet-SARIMA-ANN performed better than wavelet-ANN and wavelet-SARIMA models
Kalteh (2017)	ANN-SSA	Monthly precipitation (antecedent values)	Monthly	Iran	RMSE, CE, R	ANN-SSA (RMSE:52.257, CE:0.858, R:0.731) gives high performance compared with single ANN (RMSE:91.096, CE:0.183, R:0.455).
Mohd-Safar et al. (2016)	FCM-ANN	Air mass pressure, dew point, humidity, wind speed, rainfall amount, rainfall rate	Short-Term (1 Hour)	Malaysia	MAE, RMSE, R	FCM-ANN (MAE:0.412, RMSE:1.116, R:0.808) gives better performance than ANN (MAE:0.373, RMSE:2.408, R:0.334).
Singh (2018)	Fuzzy set-entropy-ANN	Previous monthly rainfall	ISMR (Monthly and seasonal)	India	R, RMSE, performance parameter (PP)	Compared to other models, the fuzzy set-entropy-ANN model is computationally resilient and efficient.
Gomes and Blanco (2021)	MODWT-ANN	Rainfall data	Daily	Brazil	MSE, CE, bias	The hybrid model performed better
Song and Chen (2021)	TVF-EMD-ENN, REMD-ENN, CEEMD-ENN, WT-ENN & ESMD-ENN	Annual precipitation past three days	Annual	China	RMSE, MAE, mean absolute percentage error (MAPE), R^2	The Elman neural network (ENN) hybrid models performed better
Kuo et al. (2010)	GA-ANN	Climate indices – sea surface temperature	Seasonal	Taiwan	R, Hanssen kuipers (HK), SS, RMSE	In general, ANN-GA gives good performance

Table 1 (continued)

Authors (Year)	Hybrid methods	Input variables	Forecast horizon	Location/country	Performance metrics	Comments
Jiang and Wu (2013)	PSO-GA-NN, PSO-NN, GA-NN	Previous monthly rainfall	Monthly	China	Average absolute relative error (AARE), RMSE, Pearson relative coefficient	PSO-GA-NN (AARE:76.250, RMSE:67.920) provides better performance than GA-NN (AARE:118.850, RMSE:67.980), PSO-NN (AARE:107.100, RMSE:68.930) and BP-NN (AARE:117.090, RMSE:85.960) models; better generalisation capabilities with lowest prediction error values.
Wu et al. (2015)	RBF-HPSO-GA, RBF-GA	Monthly rainfall data	Monthly	China	AARE, RMSE, R	RBF-HPSO-GA (AARE:0.605, RMSE:67.729, R:0.933) performed better with higher accuracy and better generalization ability than GA combined with RBF (AARE:0.895, RMSE:111.906, R:0.848) and single RBF-NN (AARE:1.006, RMSE:170.461, R:0.698) models.
Yaseen et al. (2019)	ANFIS-GA, ANFIS-PSO, ANFIS-DE, ANFIS	Lag monthly rainfall	Monthly	Malaysia	RMSE, MAE, R, Willmott's index (WI)	ANFIS coupled with the EAs performed better than the conventional ANFIS. ANFIS-DE (RMSE: 0.730, MAE: 0.380, R: 0.993, WI: 0.996), ANFIS-GA (RMSE: 0.830, MAE: 0.0.510, R: 0.991, WI: 0.995), ANFIS-PSO (RMSE: 0.470, MAE: 0.280, R: 0.9970, WI: 0.998), ANFIS (RMSE: 0.990, MAE: 0.650, R: 0.987, WI: 0.994)
Wahyuni et al. (2017)	ANFIS-GA	Lag rainfall data	daily	Indonesia	RMSE	ANFIS-GA gives higher performance than the ANFIS model

Table 1 (continued)

Authors (Year)	Hybrid methods	Input variables	Forecast horizon	Location/country	Performance metrics	Comments
CALP (2019)	ANFIS-GA	Temperature daily mean (2 m above ground), relative humidity (2 m above ground), mean sea level pressure, sunshine duration daily sum, wind speed daily mean (20 m above ground)	daily	Switzerland	RMSE, MAPE, MSE, R^2	ANFIS-GA (RMSE:0.015, MAPE:1.045, MSE:0.003, R^2:0.990) gives higher performance than ANN (RMSE:0.062, MAPE:4.236, MSE:0.004, R^2:0.958).
Yaseen et al. (2018)	ANFIS-FFA	Antecedent monthly data	Monthly	Malaysia	R^2, CE, WI, RMSE, MAE, relative error distribution	ANFIS-FFA (R^2:0.994, CE:0.994, RMSE:0.378, MAE:0.188) performed better than the standard ANFIS (R^2:0.915, CE:0.914, RMSE:1.401, MAE:0.906) model.
He et al. (2015)	WNN-MI-PSO	Large scale indices, antecedent monthly rainfall	Monthly	Australia	Relative absolute error, CE	WNN-MI-PSO improves the forecasting accuracy in comparison to the reference models
Kisi and Shiri (2011)	WGEP & WNF	Lag rainfall values	Daily	Turkey	RMSE, scatter index (SI), R^2	The WGEP performed marginally better than the WNF and MLR, but neither model produced satisfying results.
Chau and Wu (2010)	ANN-SVR-SSA, ANN-SSA	Rainfall time series	Daily	China	RMSE, CE	ANN-SSA performed better than ANN. The proposed ANN-SVR-SSA performed considerably better than the ANN-SSA and ANN. ANN-SVR-SSA (RMSE:3.180, CE:0.920), ANN-SSA (RMSE:4.940, CE:0.810), ANN (RMSE:10.680, CE:0.090)
Pham et al. (2019)	NF- ARIMA, MLP-ARIMA, LSSVM-ARIMA, HW- ARIMA	Wind speed, relative humidity, solar radiation, Maximum and minimum temperatures	Daily	Vietnam	R^2, RMSE, MAE, R	The hybrid models are effective and efficient in improving the forecasts.
Wu et al. (2010b)	MANN-FCM-SSA	Daily and monthly rainfall data	Monthly, daily	China	CE, RMSE, persistence index (PI)	The proposed model performed better than ANN, K-NN, and LR benchmark models

Table 1 (continued)

Authors (Year)	Hybrid methods	Input variables	Forecast horizon	Location/country	Performance metrics	Comments
Nourani et al. (2019b)	WNN & EANN	Lag rainfall data	Monthly	Cyprus	RMSE, R^2	The performance of FFNN, WNN and EANN were compared. EANN model performed better than the other two models. EANN (R^2:0.807, RMSE:6.470), WNN (R^2: 0.701, RMSE:9.490), FFNN (R^2: 0.538, RMSE: 18.200)
Johny et al. (2020)	AEEMD-ANN	Monthly rainfall data	Monthly (ISMR)	India	R, MAE, index of agreement (IA), NRMSE	AEEMD-ANN performed better when compared with EEMD-ANN
Mehdizadeh et al. (2018)	GEP-ARCH & ANN-ARCH	Lag monthly rainfall	Monthly	Iran	RMSE, R^2	GEP-ARCH and ANN-ARCH outperformed GEP and ANN models. ANN-ARCH generally performed better compared GEP-ARCH model. GEP-ARCH (RMSE: 14.950, R^2: 0.966), ANN-ARCH (RMSE:16.060, R^2: 0.959), GEP (RMSE: 31.410, R^2:0.199), ANN (RMSE:30.520, R^2: 0.243)
Zeynoddin et al. (2018)	ELM hybrids (-ARMA, ARIMA, SARIMA), ANFIS-FFA	Lag monthly rainfall	Monthly	Malaysia	Root mean relative square error (RMRSE), R^2 , RMSE, MAE	Hybrid models (both linear and non-linear) performed better than single models
Mohammadpour et al. (2018)	ANN-CLA	Minimum, maximum, average temperature, humidity, maximum, wind speed, pressure, rainfall	daily	Iran	RMSE, MAE, R^2	ANN-CLA (R^2:0.880, RMSE:0.202) gives higher performance when compared with ANN (R^2:0.839, RMSE:0.222)
Huang et al. (2015)	ANFIS	Rainfall data, typhoon characteristics	Hourly	Taiwan	MAE, RMSE, R	ANFIS gives better performance compared with BPNN
Chang et al. (2014)	ANFIS	Precipitation, reservoir inflow, radar- and satellite—derived precipitation	Hourly (1–2 h ahead)	Taiwan	R, RMSE, NRMSE, MAE, SS	ANFIS (R:0.850, RMSE:3.370, MAE:2.230) performed better than LR (R:0.830, RMSE:3.450, MAE:2.360) model.

Table 1 (continued)

Authors (Year)	Hybrid methods	Input variables	Forecast horizon	Location/country	Performance metrics	Comments
Pham et al. (2020)	PSO-ANFIS	Maximum temperature, minimum temperature, wind speed, relative humidity, solar radiation	Daily	Vietnam	R, MAE, SS, Probability of Detection (POD), Critical Success Index (CSI), False Alarm Ratio (FAR)	PSO-ANFIS forecasted acceptable daily rainfall. Nonetheless, SVM outperformed the hybrid and ANN models. PSO-ANFIS (MAE:3.281, R:0.844), ANN (MAE: 3.209, R:0.829), SVM (MAE: 2.728, R: 0.863)
Xiang et al. (2018)	EEMD-SVR-ANN	Monthly rainfall datasets	Short-term, long-term	China	R, RMSE, MAE	EEMD-SVR-ANN model showed better performances than traditional methods.
Altunkaynak and Nigussie (2015)	W-SAS-MP, SAS-MP, W-MP	Rainfall time series	Daily	Turkey	R^2, RMSE, CE, SS	It was discovered that the hybrid W-SAS-MP model outperformed the W-MP and SAS-MP models.

and McCabe (1999) assert that a comprehensive evaluation of model performance should include at least one absolute error measure, such as RMSE, as a necessary complement to a relative error measure.

Hybrid machine learning models for rainfall forecasting

ANN hybrid models

ANNs are ML techniques inspired by the biological systems first introduced by McCulloch and Pitts; however, their major applications have arisen only since the development of the back-propagation method of training by Rumelhart et al. (1986). Because of their numerous advantages, neural networks have garnered a lot of interest among ML models. They are ideal for weather forecasting tasks due to their distinct qualities of adaptability, nonlinearity, and capacity for arbitrary function mapping. The most popular neural networks are feed-forward back propagation, radial basis function-based neural networks, and generalised regression neural networks. Table 1 provides a summary of the results of the main ANN-hybrid models applied for rainfall forecasting during the period 2010–2021. The table includes the hybrid model types, input variables, country, performance metrics, and comments on the comparison with the single models. Based on the available information provided, the best statistical results during validation or testing are provided. In the case of more than one station or location, the station with the best results based on the performance metrics is considered. A hybrid model combines at least two ML techniques. As indicated by Zeynoddin et al. (2018), combining ML models with pre-processing techniques and optimisation algorithms are the two broad categories into which hybrid methods are often constructed.

Hybrid models may be composed by the application of data pre-processing techniques for multiscale decomposition or transformation of the original rainfall timeseries. The non-stationarity challenges could be resolved in such hybrid models for rainfall forecasting. The wavelet neural network (WNN) model, a combination of wavelet analysis and ANN, is one common approach used in this regard. As a powerful mathematical tool that provides a time–frequency representation of an analysed signal in the time domain (Dabuechies 1990), its application allows different independent treatments on distinct time scales such as daily, monthly, and annual periods. The WNN model has been applied to estimate daily and monthly rainfall at different locations, as shown in Table 1. The findings demonstrated that WNN model estimations outperformed the single ANN and MLR models. The WNN model can predict rainfall over both short and long timescales due to the use

Table 2 Summary of other hybrid ML methods

Authors (Year)	Hybrid methods	Input variables	Forecast horizon	Location/ country	Performance metrics	Comments
Kisi and Cimen (2012)	WSVR	Decomposed preceding daily precipitation time series	Daily	Turkey	RMSE, MAE, Nash–Sutcliffe efficiency coefficient (CE), R	WSVR (RMSE:21.400, MAE:9.000, CE:0.647, R:0.815) performed better than the standalone SVR (RMSE:38.700, MAE:14.200, CE:0.154, R:0.103) model.
Feng et al. (2015)	WSVM	Lag monthly rainfall time series	Monthly	China	R, RMSE, MAE, NSE	WSVM (R:0.945, RMSE:12.689, MAE:7.828, CE: 0.892) gives accurate forecasts and outperformed the regular SVM (R: 0.875, RMSE: 18.777, MAE: 11.578, CE: 0.762) and ANN (R:0.895, RMSE:17.224, MAE:10.695, CE: 0.800) models
Shenify et al. (2016)	WSVM	Monthly precipitation data	Monthly	Serbia	MAE, MAPE, R, R^2	In comparison to ANN and GP models, the WSVM model has a higher accuracy in estimating the amount of precipitation. WSVM (MAE:0.015, MAPE:9.314, RMSE:6.728, R: 0.859, R^2:0.738), GP (MAE: 0.027, MAPE:10.847, RMSE: 7.582, R: 0.821, R^2:0.674), ANN (MAE:0.031, MAPE: 11.160, RMSE: 7.997, R: 0.798 R^2: 0.859)

Table 2 (continued)

Authors (Year)	Hybrid methods	Input variables	Forecast horizon	Location/ country	Performance metrics	Comments
Ouyang et al. (2016)	EEMD-SVR, EEMD-ANN, EEMD-ARIMA	Monthly rainfall	Monthly	China	NMSE, MAE, NSE, R	EEMD-SVR model outperformed ANN- and ARIMA- EEMD models. EEMD-SVR (NMSE:0.100, MAPE:14.900, R^2:0.910, CE:0.830), EEMD-ANN (NMSE:0.990, MAPE:25.870, R^2:0.660, CE:0.010), EEMD-ARIMA (NMSE:0.380, MAPE:47.780, R^2:0.820, CE:0.620), SVM (NMSE:0.310, MAPE:23.020, R^2:0.830, CE:0.690).
Jiao et al. (2016)	EEMD–RBFN–SVM, EEMD-RBFN, SAM-ESM-RBFN	Monthly rainfall time series data	Monthly	China	RMSE, MAE, MaxAE, R	The proposed EEMD–RBFN–SVM achieved good prediction results and a better generalization ability than RBFN, EEMD–RBFN, and SAM–ESM–RBFN. EEMD-RBFN-SVM (RMSE:11.664, MAE:8.623, MaxAE:39.193, R:0.956), EEMD-RBFN (RMSE:15.533, MAE:12.313, MaxAE:39.396, R:0.931), SAM-ESM-RBFN (RMSE:16.110, MAE:10.439, MaxAE:60.720, R:0.906), RBFN (RMSE:24.366, MAE:19.351, MaxAE:79.650, R:0.828).
Athoillah et al. (2021)	SSA-SVR	Previous rainfall data	Monthly	Indonesia	Root mean square error of prediction (RMSEP), R	SSA-SVR model yielded more accurate results than the single model.

Table 2 (continued)

Authors (Year)	Hybrid methods	Input variables	Forecast horizon	Location/ country	Performance metrics	Comments
Wu et al. (2010b)	SVM-PSO-PI, RSVRCPSO	Factors derived from several meteorological elements and physical elements	Daily	China	RMSE, NMSE, R	SVR–PSO-PI (RMSE:6.048, NMSE:0.121, R:0.972) model is superior to the BP-NN (RMSE:21.593, NMSE:0.755, R:0.517) model and RSVRCPSO (RMSE:15.852, NMSE:0.407, R:0.799) models; better forecasting accuracy and improve the prediction quality
Du et al. (2017)	PSO-SVM, GA-SVM, AC-SVM	Pressure, wind direction, wind speed, temperature, relative humidity	Hourly	China	MSE, coefficient of variation (CV)	PSO-SVM model performed better then SVM, GA-SVM and AC-SVM
Wu and Xie (2019)	SVR-GA-PSO, SVR-GA, SVR-PSO	Several meteorological elements and physical elements	Monthly	China	RMSE, MAPE, CE	SVR-GA-PSO (RMSE:12.499, MAPE:8.552, CE:0.991) have a superior generalization capability and lowest prediction error values than SVR-PSO (RMSE:16.240, MAPE:20.649, CE:0.985), SVR-GA (RMSE:26.335, MAPE:22.482, CE:0.965) and SVR (RMSE:21.608, MAPE:24.767, CE:0.976) benchmark models
Tao et al. (2018)	PSR-SVM-FFA	Previous monthly data	Monthly	India	CE, RMSE, Willmott's Index (WI)	PSR-SVM-FFA (CE:0.900, RMSE:43.130, WI:0.974) provided an accurate and reliable forecast compared with SVM (CE:0.742, RMSE:69.620, WI:0.928) and SVM-FFA (CE:0.855, RMSE:52.220, WI:0.928) models
Danandeh Mehr et al. (2019)	SVR-FFA	Lag monthly rainfall	Monthly	Iran	NSE, RMSE, Taylor diagram	SVR-FFA (RMSE:122.800, CE:0.593) model performed better than SVR (RMSE:170.000, CE:0.184) and MGGP (145.000, CE:0.410) models.

Table 2 (continued)

Authors (Year)	Hybrid methods	Input variables	Forecast horizon	Location/ country	Performance metrics	Comments
Banadkooki et al. (2019)	SVM-FFA, MLP–FFA	Lag precipitation and temperature values	Monthly	Iran	MAE, RMSE, observations standard deviation ratio (RSR) and NSE	MLP–FFA hybrid model performed better
Farajzadeh and Alizadeh (2018)	W-S-LSSVM	Previous months rainfall data	Monthly	Iran	RMSE, R^2	W-S-LSSVM (RMSE:0.019, R^2:0.920) performed better than W-LSSVM (RMSE:0.028, R^2:0.870) and LSSVM (RMSE:0.032, R^2:0.850) models.
Tao et al. (2017)	LSSVM-EMD-DE	Previous precipitation and large-scale climate indices	Monthly	China	NSE, relative absolute error (RAE)	The hybrid LSSVM model performed better than LSSVM and LSSVM–DE models
Li et al. (2020)	VMD-IBOA-LSSVM and VMD-Volterra	Monthly precipitation	Monthly	China	MAE, RMSE, R	Compared to the VMD-ELM reference model, the VMD-IBOA-LSSVM has substantially greater forecasting accuracy.
Sumi et al. (2012)	Hybrid ANN, MARS, KNN and SVR model	Lag daily and monthly rainfall series	Daily and monthly	Japan	RMSE, CE, Persistence Index (PI)	The hybrid model performed better than the individual models
Bojang et al. (2020)	SSA-LSSVR and SSA-RF	Monthly rainfall data	Monthly	Taiwan	RMSE, CE	Hybrid models performed better than single model. SSA-LSSVR (RMSE:79.480, CE:0.840), SSA-RF (RMSE:137.770, CE:0.530), LSSVR (RMSE:192.310, CE: 0.070).

Table 2 (continued)

Authors (Year)	Hybrid methods	Input variables	Forecast horizon	Location/ country	Performance metrics	Comments
Mehdizadeh (2020)	KNN-AR, KNN-MA, ARMA-KNN	Temperature, dew point temperature, station pressure, vapor pressure, relative humidity, wind speed, antecedent precipitation data	Monthly	Iran	RMSE, MAE, NSE, R^2	In comparison to single models, the hybrid MARS-AR, MARS-MA, MARS-ARMA, KNN-AR, KNN-MA, and KNN-ARMA models outperformed them. KNN-AR (RMSE:2.020, MAE:1.570, R^2:0.999) KNN-MA (RMSE:2.870, MAE:2.210, R^2:0.998), ARIMA-KNN (RMSE:13.910, MAE:10.960, R^2:0.952), KNN (RMSE:60.720, MAE:45.400, R^2:0.194).
Mehr et al. (2019)	MPSA-MGGP	Lag monthly rainfall values	Monthly	Iran	NSE, RMSE	MPSA-MGGP perform better than GP and MGGP models
Mehdizadeh et al. (2017)	MARS-GARCH, BN-GARCH and GEP-GARCH	Previous months precipitation data	Monthly	Iran	RMSE, relative root mean square error (RRMSE), MAE, R^2	The overall accuracy of the three proposed hybrid models is excellent or good. They performed better than the single models. MARS-GARCH (RMSE:0.100, RRMSE:0.200, MAE:0.100, R^2:1.000) BN-GARCH (RMSE:6.100, RRMSE:7.600, MAE:3.400, R^2:0.994), GEP-GARCH (RMSE:0.900, RRMSE:1.100, MAE:0.700, R^2:1.000), MARS (RMSE:55.400, RRMSE:68.900, MAE:37.2, R^2:0.573), BN (RMSE:57.600, RRMSE:71.700, MAE:38.000, R^2:0.518), GEP (RMSE:57.100, RRMSE:71.100, MAE:37.700, R^2: 0.515).

Table 2 (continued)

Authors (Year)	Hybrid methods	Input variables	Forecast horizon	Location/ country	Performance metrics	Comments
Pham et al. (2019)	MLP-ARIMA, LSSVM-ARIMA, NF- ARIMA, HW- ARIMA	Rainfall data	Daily	Vietnam	RMSE, MAE, R^2, R, Taylor diagram	Hybrid models are effective and efficient in improving forecasts accuracy
Dabhi and Chaudhary (2014)	Wavelet-postfix-GP	Temperature, relative humidity, evaporation, previous daily rainfall	Daily	India	MAE, MSE, R	Wavelet-postfix-GP model fairly competitive to ANN model
Lin and Jhong (2015)	MGSVM	Rainfall, air pressure, air temperature, wind velocity, wind direction, duration of sunshine	Hourly (1 to 6hr ahead)	Taiwan	MAE, CE	MGSVM performs better than SVM for most of the events, especially for long lead time forecasting.
Ali et al. (2020)	CEEMD-RF-KRR, CEEMD-RF, RF, KRR	Rainfall	Monthly	Pakistan	RMSE, MAE, R, WI	CEEMD-RF-KRR performed best: CEEMD-RF-KRR (RMSE:2.52, MAE:1.98, R:0.986), CEEMD-RF (RMSE:3.86, MAE:2.95, R:0.974), RF (RMSE:8.99, MAE:6.24, R:0.826), KRR (RMSE:19.82, MAE:12.91, R:0.047)

of multiscale time series as input data to the neural network. For instance, for time series modelling of monthly rainfall, Venkata Ramana et al. (2013) have observed that the efficiency index for the WNN model is above 94%, while it is just 64% for the ANN model. The WNN model also outperformed ANFIS for daily precipitation forecasting at Hamedan, Iran (Solgi et al. 2014). Wavelet analysis has been combined with ANFIS to form the wavelet-neuro-fuzzy (W-ANFIS) model. The W-ANFIS model prediction results are shown to be considerably superior than those of WNN and ANFIS (e.g., Akrami et al. 2014); however, the hybrid models performed better than the ANN model. Similarly, wavelet analysis, seasonal auto-regressive integrated moving average (SARIMA), and WNN hybrid models to accurately forecast monthly precipitation have been investigated (Shafaei et al. 2016). The hybrid wavelet-SARIMA-ANN, WNN, and wavelet-SARIMA models were compared. The wavelet-SARIMA-ANN model outperformed the other models, according to the experimental findings. Kim et al. (2016) developed and applied two different hybrid models for rainfall aggregation and spatial disaggregation, wavelet-based support vector machines (WSVM) and wavelet-based generalized regression neural networks (WGRNN), and evaluated them in the Bocheong-stream catchment, Republic of Korea. The findings revealed that wavelet-based models (WSVM and WGRNN) outperformed the SVM and GRNN single models. Kisi and Shiri (2011) investigated the accuracy of two distinct wavelet conjunction models, namely wavelet genetic programming (WGEP) and wavelet-neuro-fuzzy approaches (WNF), for forecasting daily precipitation. In comparison, the GEP model fared marginally better than the others, although neither produced satisfying results. Nevertheless, it was discovered that the wavelet conjunction models (WGEP and WNF) greatly improved the prediction accuracy of the GEP and NF models. Discrete and continuous wavelet transformations are the two types of wavelet analysis; however, the discrete wavelet transformation is the most commonly used wavelet transform. In order to predict daily precipitation, Altunkaynak and Ozger (2016) compared discrete (DWT-MP) and continuous (CWT-MP) wavelet-MLP algorithms. The CWT-MP and DWT-MP are built and evaluated alongside the ANN model. The findings showed that applying wavelet transformations considerably enhanced the prediction performance of MLP. Based on the findings of the assessment criteria utilised, the CWT-MP model exhibited superior performance to the DWT-MP model. Gomes and Blanco (2021) have developed a maximum overlap discrete wavelet transform—ANN (MODWT-ANN) hybrid model that produces better performance in daily rainfall forecasting in Brazil than the ANN model.

Singular spectrum analysis (SSA) is another very successful data pre-processing technique that comprises two complementary steps of decomposition and reconstruction

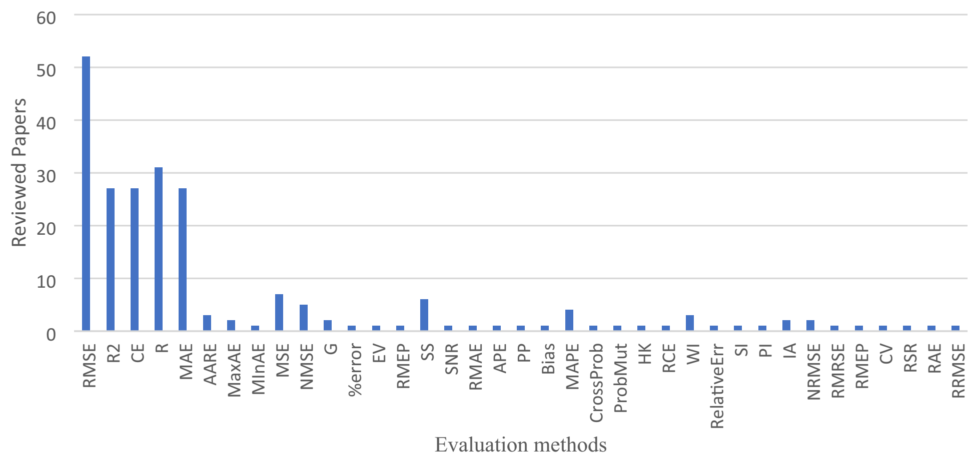


Fig. 2 Evaluation methods used in the reviewed papers.

(Golyandina et al. 2001). When compared with the ANN model, the findings indicate the hybrid SSA-ANN model is more accurate than the ANN model (e.g., Kalteh 2017). Additional findings demonstrated that the hybrid combination of the three models (ANN-SVR-SSA) performed best as it exhibited considerable accuracy in rainfall forecasting. Pre-processing the input data into components without random components using SSA allowed for a better representation of the precipitation processes in the conjunction models. Other decomposition techniques, including ensemble empirical mode decomposition (EEMD), Wavelet transform (WT), robust empirical mode decomposition (REMD), complementary ensemble empirical mode decomposition (CEEMD), time-varying filter-based empirical mode decomposition (TVF-EMD), extreme-point symmetric mode decomposition (ESMD), and season decomposition algorithms, have combined ML techniques to forecast rainfall. The various proposed hybrid models, as indicated in Table 1, outperformed the standard models.

Introduced by Jang (1993) as a universal approximator, any real continuous function on a compact set may be approximated to any degree of precision using the adaptive neuro-fuzzy inference system (ANFIS). The ANFIS, which is a hybrid model that integrates a neural network with a fuzzy system, is a common technique that has been widely used for modelling several types of nonlinear systems, including rainfall. As presented in Table 1, it has been applied to forecast rainfall in different regions of the world and at various timescales, including hourly, daily, monthly, and seasonal. When compared with the standard ANN model via a number of evaluation criteria, the ANFIS model was shown to have greater rainfall prediction accuracy and lower error than the single model at the various settings. It also performed better than multiple linear regression (MLR). However, in seasonal precipitation forecasting in Iran, Choubin et al. (2016) observed that the single ANN (i.e., multi-layer perceptron network, MLP) model demonstrated superior accuracy over ANFIS in the study location. In order to increase the ANFIS model's prediction accuracy, the authors suggested that the optimal combination and length of data be obtained by utilising additional nonlinear approaches such as genetic algorithms (GA), gamma tests, and particle swarm optimisation (PSO). A modified adaptive neuro-fuzzy inference system (MANFIS), proposed by Jovanovic et al. (2004), has been identified to increase the performance of the ANFIS model. The MANFIS has been applied to forecast monthly rainfall, and the experimental findings revealed that the MANFIS outperformed the ANFIS model in terms of forecasting accuracy, error, and computing complexity (Akrami et al. 2013). In Chau and Wu (2010), the fuzzy C-Means (FCM) soft clustering method, which gives membership to each data point that corresponds to each cluster centre based on the distance between the cluster centre and the data point (Bezdek et al. 1984), has been combined with ANN, SVM, and SSA. The FCM-ANN model performed better than single ANN models, producing 80% for a 1-h forecast (Mohd-Safar et al. 2016). In Singh (2018), a hybrid methodology that combined fuzzy set, entropy, and ANN is applied to forecast the Indian summer monsoon rainfall (SMR). Although forecasting ISMR may be accomplished using monthly time series data and ANN, the author stated that ANN-based models alone cannot deal with the unpredictable or dynamic character of ISMR. As a consequence, the applications of fuzzy set for uncertain representations of ISMR values, entropy for inherited information in ISMR values, and ANN for final outcomes, in contrast to current models, show that the suggested model is computationally resilient and efficient.

Several hybrid models are also obtained by applying different optimisation algorithms to improve the speed of convergence and the accuracy of precipitation forecasting. The findings provided in Table 1 showed that the ML models coupled with optimisation techniques consistently performed better compared to the standalone models (e.g., Kuo et al.

2010; Wu et al. 2015; Azad et al. 2019). When combined with the ANFIS hybrid model, the hybrid models performed better than ANFIS with a lower error rate and a higher accuracy level in prediction (e.g., Wahyuni et al. 2017; Azad et al. 2019; Yaseen et al. 2019; Pham et al. 2020). For instance, Yaseen et al. (2018) found that the ANFIS merged with the firefly optimisation algorithm (FFA), an algorithm inspired by the social behaviour of fireflies (Yang 2010), on average increased the CE value by roughly 8%. Furthermore, the ANFIS-FFA model appears to lower the RMSE value by 53% during the testing phase, which is consistent with the training phase findings. He et al. (2015) presented a hybrid mix of WNN, mutual information (MI), and PSO for monthly rainfall forecasting. PSO is a population-based stochastic optimization technique based on studies of animal social behavior such as bird flocking, fish schooling, and swarm theory (Kennedy and Mendes 2006). The MI algorithm, which is a measure of statistical dependence, is used for input identification in the proposed model. The findings revealed that the proposed model increases monthly rainfall forecasting accuracy across Australia in comparison to the reference models. Diop et al. (2020) developed an MLP-WOA hybrid model by combining MLP with the Whale Optimisation Algorithm, an innovative heuristic algorithm that belongs to the family of stochastic population-based algorithms proposed by Mirjalili and Lewis (2016). When compared to MLP for annual rainfall forecasting, results show that MLP-WOA slightly improved the accuracy of the standard MLP models. Ant colony optimisation (ACO) for the continuous domain, chaotic particle swarm optimisation technique (CPSO), and differential evolution (DE) are the other types of optimisation algorithms applied in this paper. Other hybrid models as presented in Table 1 are shown to perform better than single models in rainfall forecasting, including heteroscedasticity (GEP-ARCH) and ANN-autoregressive conditional heteroscedasticity (ANN-ARCH) (Mehdizadeh et al. 2018), emotional artificial neural network (EANN) hybrid models (Lotfi and Akbarzadeh-T 2014), adaptive ensemble empirical mode decomposition-artificial neural network (AEEMD-ANN) (Johny et al. 2020), nonlinear extreme learning machine (ELM) (Zeynoddin et al. 2018), and Learning-Cellular Automation (CLA)-ANN model (ANN-CLA) (Mohammadpour et al. 2018).

Other hybrid ML models

A summary of hybrid methods other than a combination of ANN with other techniques for rainfall forecasting (2010–2021) is presented in Table 2. SVM, initially developed for classifying binary data (Cortes and Vapnik 1995), can be used for tasks requiring classification as well as regression; the regression variant is known as support vector regression (SVR). Due to its robustness and generalisation performance, several studies have considered SVM-based hybrid models in order to improve rainfall prediction. The Wavelet-SVR conjunction model was found to perform better than the ANN, SVM, and genetic programming (GP) models in both long- and short-term rainfall predictions (e.g., Kisi and Cimen 2012; Feng et al. 2015; Shenify et al. 2016). Hybrid EEMD-SVM based on phase-space reconstruction has improved rainfall prediction when compared with ARIMA, ANN, and SVR benchmark models (Ouyang et al. 2016). In Jiao et al. (2016), EEMD has been combined with radial basis function neural networks (RBFN) and SVM to forecast precipitation in Qinghai Province, China. The proposed model was compared with RBFN, EEMD–RBFN, and SAM–ESM–RBFN (seasonal adjustment method, SAM) models. The findings demonstrated the effectiveness of EEMD–RBFN–SVM for predicting precipitation compared to the benchmark models. Athoillah et al. (2021) showed that coupling SVM with SSA data pre-processing techniques yields a more accurate forecast than the single SVM model in forecasting monthly rainfall. Optimisation techniques, including PSO, FFA, and GA, have also been combined with SVM to forecast rainfall. The PSO-SVM hybrid model outperformed the benchmark models SVM, GA-SVM, and AC-SVM (SVM with the Ant Colony Algorithm) (Du et al. 2017). A hybrid SVM-FFA model delivered more precise and reliable results when compared to a single SVM model (e.g., Tao et al. 2018). Danandeh Mehr et al. (2019) have compared SVR–FFA with SVR and multigene genetic programming (MGGP) models to forecast rainfall in a semiarid region of Iran. The authors concluded from the results that the SVR-FFA model has a greater ability than SVR to capture nonlinear features of monthly rainfalls in the region, and forecast accuracy has been greatly enhanced when employing MGGP models. In Banadkooki et al. (2019), the new flow regime optimization algorithm (FRA), an algorithm created based on fluid flow concepts (Tahani and Babayan 2018), is applied to optimise the MLP and SVM. To predict monthly precipitation, the proposed hybrid models (MLP-FRA and SVM-FRA) were compared with the decision tree model (M5T). The outcomes demonstrated that the MLP-FRA model performed better than the other models with a smaller uncertainty band width. Wu et al. (2010a, b) proposed a hybrid SVM-PSO-based projection pursuit technology (PI), a technique for projecting high-dimensional data onto a low-dimensional subspace by numerically optimising a certain objection function or projection index (Friedman and Tukey 1974). The empirical findings using observed data of daily rainfall values in Guangxi, China, revealed that the SVM-PSO-PI model outperformed the BP-NN and RSVRCPSO models. Wu and Xie (2019) applied a hybrid combination of SVR and parallel co-evolution algorithms based on GA and PSO for forecasting monthly rainfall. The findings demonstrated that the SVR-GA-PSO model has a higher generalisation capacity with the lowest forecasting error values, which may greatly enhance rainfall forecasting accuracy. SVM combined with a multi-objective genetic algorithm (MOGA) is utilised in Lin and Jhong (2015) to forecast the hourly rainfall in Taiwan's Tsengwen river basin. The proposed MGSVM has been shown to not only effectively improve forecast accuracy but also reduce the negative impact of increased prediction lead time.

Due to its capacity to generalise and minimise errors, least-squares support vector regression (LSSVR), a variation of the conventional SVR, has been proven to perform better. For instance, Farajzadeh and Alizadeh (2018) suggested a hybrid model (W-S-LSSVM) based on the discrete wavelet transform, SARIMAX, and LSSVM to handle linear, nonlinear, and seasonal rainfall time series. According to the findings of the monthly rainfall forecasts across the Iranian watershed around Urmia Lake, the W-S-LSSVM model performed about 7%–8% better than traditional models and about 5%–6% better than the W-LSSVM model. The suggested model is capable of making accurate predictions two, three, six, and twelve months in the future. Tao et al. (2017) proposed a hybrid LSSVM technique that combines the empirical mode decomposition (EMD) algorithm for preprocessing, the partial information (PI) algorithm for input identification, and the differential evolution (DE) algorithm for optimisation. When applied for monthly precipitation forecasting at the Yangtze River basin in China, the proposed approach performed better than the LSSVM–DE and LSSVM models. Additionally, the model exhibited significant spatial variability in forecast performance. The hybrid LSSVM–DE also performed better than the standalone LSSVM model. Li et al. (2020) proposed a hybrid LSSVM model with an error correction strategy that improved the performance by at least 43% for monthly mean precipitation forecasting in Shaanxi Province, China.

Random forests (RF) and k-nearest neighbour (KNN) learning algorithms are the other ML models applied in rainfall prediction (e.g., Sumi et al. 2012; Mehdizadeh 2020; Bojang et al. 2020). RF is a tree-based algorithm based on the classification and regression trees proposed by Breiman (2001). KNN, proposed by Fix and Hodges (1952) for pattern classification, is a non-parametric statistical method for both classification and regression tasks. Only a few RF and KNN hybrid models exists; however, the performance comparison analysis also revealed that hybrid models were more accurate than single RF and KNN models (see Table 2). In Ali et al. (2020), a hybrid combination of complete ensemble empirical mode decomposition (CEEMD), RF, and Kernel Ridge Regression (KRR) algorithms has been used to forecast rainfall on a monthly time scale. The CEEMD-RF-KRR model performed best with R (0.97–0.99), WI (0.94–0.97), and NSE (0.94–0.97). Other advanced hybrid models other than ANN, SVM, RF, and KNN hybrid models have been used to forecast rainfall. Mehr et al. (2019) applied a multi-period simulated annealing optimiser with a multigene genetic programming symbolic regression (MPSA-MGGP) model, and Mehdizadeh et al. (2017) developed three novel hybrid models by combining MARS, Bayesian networks (BN), and gene expression programming (GEP) with GARCH. The hybrid models performed better.

Physical—ML hybrid models

Almost all operational rainfall and weather forecasting is currently based on Numerical Weather Prediction (Marchuk 2012). While the development of NWP models has reached remarkable accuracy in predicting the future atmospheric state, the fundamental approach has not changed in this time (Scher 2020) and still suffers several challenges. The main scientific challenges have continuously seen improvement (Bauer et al. 2015), though they remain sources of uncertainties that need advancement for future NWP models. In many environmental and engineering domains, there is a growing agreement for innovative methodologies capable of integrating traditional physics-based modelling approaches with cutting-edge ML techniques (Willard et al. 2021). In the field of weather and climate, researchers are looking for ways to speed up the computation of the NWP model's very computationally expensive components, to enhance the performance of current algorithms, and to replicate the current codes with ML models in order to make it easy for a model to run on a computer cluster with GPU accelerators (Bochenek and Ustrnul 2022). Incorporating physics and domain knowledge into ML models has proved beneficial in simulating, downscaling, and forecasting weather and climate processes, according to a recent analysis of 10 case studies (Kashinath et al. 2021). Kashinath et al. (2021) observed that physics-informed ML models can achieve increased physical consistency, higher accuracy, faster training, better convergence, data efficiency, improved generalization, greater interpretability, and increased scalability to more complex physical systems and larger computational platforms. According to Kashinath et al. (2021), physics-informed machine learning models can boost physical consistency, accuracy, speed of training, convergence, data efficiency, generalisation, interpretability, and scalability to more advanced physical systems and larger computing platforms. O'Gorman and Dwyer (2018) recently used ML to parameterise moist convection, which led to precise simulations of climate and precipitation extremes in an atmospheric circulation model. Krasnopolsky et al. (2013) have published a technique for parameterising stochastic convection in numerical weather predictions and climate models. Moreover, ML has been used to develop statistical models that mimic the behaviour of climate models (Anderson and Lucas 2018; Scher 2018). To the best of our knowledge, there are no explicit studies on combining physics-based models with ML models to forecast rainfall. Yet, the preceding and several other publications have demonstrated that ML approaches have unparalleled capabilities for improving the predictive understanding of existing weather and climate models.

Summary and conclusion

Accurate and reliable rainfall predictions at various timeframes are critical for a variety of practical applications, including agriculture, energy, mining, and water resource management, among others. Disasters caused by climate and hydro-meteorological extremes can be prevented or mitigated by timely predictions. Consequently, several methods, including NWP, statistical, and ML or AI approaches, have been used to provide forecasts of this important meteorological phenomenon. ML methods have, over the last decades, been applied successfully and exhibited great potential in rainfall forecasting. However, the complex processes involved in rainfall formation and the non-stationary nature of rainfall time series data make single ML models insufficient for providing accurate forecasts in most cases. The introduction of novel or advanced ML methods like deep learning and the hybridisation of different ML techniques have proven to be more accurate and efficient. In this paper, an extensive review of hybrid ML methods for rainfall forecasting that includes the earliest work and recent advances is presented. The focus was not on the detailed description of the architecture or structure of these models but on the type of ML techniques involved in the hybridisation, model evaluation, and the performance of hybrid models compared with single ones.

While hybrid ML models are broadly composed by integrating with data pre-processing techniques and optimisation algorithms, hybrid models are formed by the integration of two or more ML algorithms that are not necessary optimisers or pre-processing techniques. Several pre-processing or decomposition techniques are employed in the study to convert the original time series data into new series or to decompose the original time series data. They include wavelet analysis, ensemble empirical mode decomposition, singular spectrum analysis, time-varying filter-based empirical mode decomposition, robust empirical mode decomposition, complementary ensemble empirical mode decomposition, wavelet transform, extreme-point symmetric mode decomposition, adaptive ensemble empirical mode decomposition, seasonal adjustment method, season decomposition, variational mode decomposition, etc. In addition to fuzzy inference systems, Fuzzy C-Means, which is a data clustering technique, has also been used for preprocessing. The optimisation algorithms applied in this study include genetic algorithm, particle swarm optimisation, chaotic particle swarm optimisation algorithm, whale optimisation algorithm, ant colony algorithm, ant colony optimisation for the continuous domain, differential evolution, firefly optimisation algorithm, improved butterfly optimisation algorithm, etc. The experimental results in this study revealed that the hybrid models produced higher rainfall forecast accuracy, low error, improved speed of convergence, better generalisation and reduced computational complexity compared with the single model. Furthermore, the hybrid models performed well at the various forecasting timescales. Daily and monthly timescales are the two most frequently considered in rainfall forecasting, according to the studies reviewed in this paper. Rainfall and other exogenous variables such as temperature, sunlight, air pressure, relative humidity, wind speed, and wind direction are useful model inputs for rainfall forecasting. Large-scale climate indices have also been employed as model inputs. Different combinations of these variables, including their antecedent values, are used for rainfall prediction. In terms of evaluation methods used to assess the performance and uncertainty of the models, the majority of the studies employed root mean square error, mean absolute error, correlation coefficient, coefficient of determination, and Nash–Sutcliffe efficiency coefficient.

Machine learning techniques are now being combined with physical models to enhance the predictive performance of the existing models that are constrained by atmospheric physical and chemical processes. Overall, the empirical findings showed that integrating several ML techniques may be a successful and efficient solution to enhance rainfall predictions at various timescales. The ML hybrid models are capable of producing comparatively more accurate forecasts for both short and longer lead times. Ensemble-based ML models such as RF, gradient boosting machines, extreme gradient boosting algorithms, etc. generally perform better than individual learners with better generalisation capabilities. For instance, RF is an efficient algorithm with unique qualities, including good generalisation, a faster learning rate, handling datasets of various sizes, robustness to noise, and missing data. However, these models have been less explored in this review paper. The authors would like to suggest more investigation be done into such models for rainfall forecasting to take advantage of their unique qualities. In recent years, advanced ML methods like deep learning or deep structures have been developed, which are producing remarkable results. It is important to benchmark this with the standard hybrid ML models.

Declarations

Conflict of interest The authors do not have any conflict of interest.

References

Abbot J, Marohasy J (2012) Application of artificial neural networks to rainfall forecasting in Queensland, Australia. Adv Atmos Sci 29(4):717–730

Abdusselam Altunkaynak PE, ASCE A, Ozger M (2016) Comparison of discrete and continuous wavelet–multilayer perceptron methods for daily precipitation prediction. J Hydrol Eng 21(7):04016014

Akrami SA, El-Shafie A, Jaafar O (2013) Improving rainfall forecasting efficiency using modified adaptive neuro-fuzzy inference system (MANFIS). Water Resour Manag 27(9):3507–3523

Akrami SA, Nourani V, Hakim SJS (2014) Development of nonlinear model based on wavelet-ANFIS for rainfall forecasting at Klang Gates Dam. Water Resour Manag 28(10):2999–3018

Ali M, Prasad R, Xiang Y, Yaseen ZM (2020) Complete ensemble empirical mode decomposition hybridized with random forest and kernel ridge regression model for monthly rainfall forecasts. J Hydrol 584:124647

Altunkaynak A, Ozger M (2016) Comparison of discrete and continuous wavelet–multilayer perceptron methods for daily precipitation prediction. J Hydrol Eng 21(7):04016014

Altunkaynak A, Nigussie TA (2015) Prediction of daily rainfall by a hybrid wavelet-season-neuro technique. J Hydrol 529:287–301

Anderson GJ, Lucas DD (2018) Machine learning predictions of a multiresolution climate model ensemble. Geophys Res Lett 45(9):4273–4280

Anochi JA, de Almeida VA, de Campos Velho HF (2021) Machine learning for climate precipitation prediction modeling over South America. Remote Sens 13(13):2468

Athoillah I, Wigena AH, Wijayanto H (2021) Hybrid modeling of singular spectrum analysis and support vector regression for rainfall prediction. J Phys Conf Ser 1863(1):012054

Azad A, Manoochehri M, Kashi H, Farzin S, Karami H, Nourani V, Shiri J (2019) Comparative evaluation of intelligent algorithms to improve adaptive neuro-fuzzy inference system performance in precipitation modelling. J Hydrol 571:214–224

Banadkooki FB, Ehteram M, Ahmed AN, Fai CM, Afan HA, Ridwam WM, Sefelnasr A, El-Shafie A (2019) Precipitation forecasting using multilayer neural network and support vector machine optimization based on flow regime algorithm taking into account uncertainties of soft computing models. Sustainability 11(23):6681

Bauer P, Thorpe A, Brunet G (2015) The quiet revolution of numerical weather prediction. Nature 525(7567):47–55

Bezdek JC, Ehrlich R, Full W (1984) FCM: The fuzzy c-means clustering algorithm. Comput Geosci 10(2–3):191–203

Bishop CM, Nasrabadi NM (2006) Pattern recognition and machine learning, vol 4(4). Springer, Berlin, p 738

Bochenek B, Ustrnul Z (2022) Machine learning in weather prediction and climate analyses—applications and perspectives. Atmosphere 13(2):180

Bojang PO, Yang TC, Pham QB, Yu PS (2020) Linking singular spectrum analysis and machine learning for monthly rainfall forecasting. Appl Sci 10(9):3224

Breiman L (1996) Bagging predictors. Mach Learn 24(2):123–140

Breiman L (2001) Random forests. Mach Learn 45:5–32

Bushara NO, Abraham A (2015) Using adaptive neuro-fuzzy inference system (anfis) to improve the long-term rainfall forecasting. J Network Innov Comput 3(2015):146–158

Calp MH (2019) A hybrid ANFIS-GA approach for estimation of regional rainfall amount. Gazi Univ J Sci 32(1):145–162

Chang FJ, Chiang YM, Tsai MJ, Shieh MC, Hsu KL, Sorooshian S (2014) Watershed rainfall forecasting using neuro-fuzzy networks with the assimilation of multi-sensor information. J Hydrol 508:374–384

Chau KW, Wu CL (2010) A hybrid model coupled with singular spectrum analysis for daily rainfall prediction. J Hydroinf 12(4):458–473

Choubin B, Khalighi-Sigaroodi S, Malekian A, Kişi Ö (2016) Multiple linear regression, multi-layer perceptron network and adaptive neuro-fuzzy inference system for forecasting precipitation based on large-scale climate signals. Hydrol Sci J 61(6):1001–1009

Cortes C, Vapnik V (1995) Support-vector networks. Mach Learn 20:273–297

Dabhi VK, Chaudhary S (2014) Hybrid wavelet-postfix-GP model for rainfall prediction of Anand Region of India. Adv Artif Intell 2014

Dabuechies I (1990) The wavelet transform, time-frequency localization and signal analysis. IEEE Trans Inf Theory 36:6–7

Danandeh Mehr A, Nourani V, Karimi Khosrowshahi V, Ghorbani MA (2019) A hybrid support vector regression–firefly model for monthly rainfall forecasting. Int J Environ Sci Technol 16(1):335–346

Diop L, Samadianfard S, Bodian A, Yaseen ZM, Ghorbani MA, Salimi H (2020) Annual rainfall forecasting using hybrid artificial intelligence model: integration of multilayer perceptron with whale optimization algorithm. Water Resour Manage 34:733–746

Du J, Liu Y, Yu Y, Yan W (2017) A prediction of precipitation data based on support vector machine and particle swarm optimization (PSO-SVM) algorithms. Algorithms 10(2):57

El-Shafie A, Jaafer O, Seyed A (2011) Adaptive neuro-fuzzy inference system based model for rainfall forecasting in Klang River. Malays Int J Phys Sci 6(12):2875–2888

Fahimi F, Yaseen ZM, El-shafie A (2017) Application of soft computing based hybrid models in hydrological variables modeling: a comprehensive review. Theor Appl Climatol 128(3):875–903

Farajzadeh J, Alizadeh F (2018) A hybrid linear–nonlinear approach to predict the monthly rainfall over the Urmia Lake watershed using wavelet-SARIMAX-LSSVM conjugated model. J Hydroinf 20(1):246–262

Feng Q, Wen X, Li J (2015) Wavelet analysis-support vector machine coupled models for monthly rainfall forecasting in arid regions. Water Resour Manag 29(4):1049–1065

Fix E, Hodges JL (1952) Discriminatory analysis-nonparametric discrimination: small sample performance. California University Berkeley, Berkeley

Friedman JH, Tukey JW (1974) A projection pursuit algorithm for exploratory data analysis. IEEE Trans Comput 100(9):881–890

Garbrecht JD, Zhang XC, Schneider JM, Steiner JL (2010) Utility of seasonal climate forecasts in management of winter-wheat grazing. Appl Eng Agric 26(5):855–866

Golyandina N, Nekrutkin V, Zhigljavsky AA (2001) Analysis of time series structure: SSA and related techniques. CRC Press, London

Gomes EP, Blanco CJC (2021) Daily rainfall estimates considering seasonality from a MODWT-ANN hybrid model. J Hydrol Hydromech 69(1):13–28

Haddad MS (2011) Capacity choice and water management in hydroelectricity systems. Energy Econ 33(2):168–177

Hartmann H, Snow JA, Stein S, Su B, Zhai J, Jiang T, Krysanova V, Kundzewicz ZW (2016) Predictors of precipitation for improved water resources management in the Tarim River basin: creating a seasonal forecast model. J Arid Environ 125:31–42

He X, Guan H, Qin J (2015) A hybrid wavelet neural network model with mutual information and particle swarm optimization for forecasting monthly rainfall. J Hydrol 527:88–100

Huang CL, Hsu NS, Wei CC, Lo CW (2015) Using artificial intelligence to retrieve the optimal parameters and structures of adaptive network-based fuzzy inference system for typhoon precipitation forecast modeling. Adv Meteorol 2015

Jang JS (1993) ANFIS: adaptive-network-based fuzzy inference system. IEEE Trans Syst Man Cybern 23(3):665–685

Jagannathan P (1974) Long range weather forecasting, forecasting manual part IV-comprehensive articles on selected topics. India Meteorological Department

Jeong C, Shin JY, Kim T, Heo JH (2012) Monthly precipitation forecasting with a neuro-fuzzy model. Water Resour Manag 26(15):4467–4483

Jiao G, Guo T, Ding Y (2016) A new hybrid forecasting approach applied to hydrological data: a case study on precipitation in Northwestern China. Water 8(9):367

Jiang L, Wu J (2013) Hybrid PSO and GA for neural network evolutionary in monthly rainfall forecasting. In: Asian conference on intelligent information and database systems. Springer, Berlin, pp 79–88

Johny K, Pai ML, Adarsh S (2020) Adaptive EEMD-ANN hybrid model for Indian summer monsoon rainfall forecasting. Theoret Appl Climatol 141(1):1–17

Jovanovic BB, Reljin IS, Reljin BD (2004) Modified ANFIS architecture-improving efficiency of ANFIS technique. In: 7th seminar on neural network applications in electrical engineering, 2004. NEUREL 2004. 2004. IEEE, pp 215–220.

Kalteh AM (2017) Enhanced monthly precipitation forecasting using artificial neural network and singular spectrum analysis conjunction models. INAE Lett 2(3):73–81

Kashinath K, Mustafa M, Albert A, Wu JL, Jiang C, Esmaeilzadeh S, Azizzadenesheli K, Wang R, Chattopadhyay A, Singh A, Manepalli A (2021) Physics-informed machine learning: case studies for weather and climate modelling. Philos Trans R Soc A 379(2194):20200093

Kennedy, J., & Mendes, R. (2006). Neighborhood topologies in fully informed and best-of-neighborhood particle swarms. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, *36*(4), 515–519.

Kim S, Seo Y, Lee CJ (2016) Modeling of rainfall by combining neural computation and wavelet technique. Proc Eng 154:1231–1236

Kisi O, Shiri J (2011) Precipitation forecasting using wavelet-genetic programming and wavelet-neuro-fuzzy conjunction models. Water Resour Manag 25(13):3135–3152

Kisi O, Cimen M (2012) Precipitation forecasting by using wavelet-support vector machine conjunction model. Eng Appl Artif Intell 25(4):783–792

Kuo CC, Gan TY, Yu PS (2010) Wavelet analysis on the variability, teleconnectivity, and predictability of the seasonal rainfall of Taiwan. Mon Weather Rev 138(1):162–175

Krasnopolsky VM, Fox-Rabinovitz MS, Belochitski AA (2013) Using ensemble of neural networks to learn stochastic convection parameterizations for climate and numerical weather prediction models from data simulated by a cloud resolving model. In: Advances in artificial neural systems

Legates DR, McCabe GJ Jr (1999) Evaluating the use of "goodness-of-fit" measures in hydrologic and hydroclimatic model validation. Water Resour Res 35(1):233–241

LeCun Y, Bengio Y, Hinton G (2015) Deep learning. Nature 521(7553):436–444

Li G, Chang W, Yang H (2020) A novel combined prediction model for monthly mean precipitation with error correction strategy. IEEE Access 8:141432–141445

Lin GF, Jhong BC (2015) A real-time forecasting model for the spatial distribution of typhoon rainfall. J Hydrol 521:302–313

Lotfi E, Akbarzadeh-T MR (2014) Practical emotional neural networks. Neural Netw 59:61–72

Marchuk G (2012) Numerical methods in weather prediction. Elsevier, Amsterdam

Mekanik F, Imteaz MA, Gato-Trinidad S, Elmahdi A (2013) Multiple regression and Artificial Neural Network for long-term rainfall forecasting using large scale climate modes. J Hydrol 503:11–21

Mekanik F, Imteaz MA, Talei A (2016) Seasonal rainfall forecasting by adaptive network-based fuzzy inference system (ANFIS) using large scale climate signals. Clim Dyn 46(9):3097–3111

Mehr AD, Jabarnejad M, Nourani V (2019) Pareto-optimal MPSA-MGGP: a new gene-annealing model for monthly rainfall forecasting. J Hydrol 571:406–415

Mehdizadeh S, Behmanesh J, Khalili K (2018) New approaches for estimation of monthly rainfall based on GEP-ARCH and ANN-ARCH hybrid models. Water Resour Manag 32(2):527–545

Mehdizadeh S (2020) Using AR, MA, and ARMA time series models to improve the performance of MARS and KNN approaches in monthly precipitation modeling under limited climatic data. Water Resour Manag 34(1):263–282

Mehdizadeh S, Behmanesh J, Khalili K (2017) A comparison of monthly precipitation point estimates at 6 locations in Iran using integration of soft computing methods and GARCH time series model. J Hydrol 554:721–742

McGovern A, Lagerquist R, Gagne DJ, Jergensen GE, Elmore KL, Homeyer CR, Smith T (2019) Making the black box more transparent: Understanding the physical implications of machine learning. Bull Am Meteor Soc 100(11):2175–2199

Mirjalili S, Lewis A (2016) The whale optimization algorithm. Adv Eng Softw 95:51–67. https://doi.org/10.1016/j.advengsoft.2016.01.008

Mohammadpour R, Asaie Z, Shojaeian MR, Sadeghzadeh M (2018) A hybrid of ANN and CLA to predict rainfall. Arab J Geosci 11(18):1–9

Molnar C (2020) Interpretable machine learning. Lulu.com

Mohd-Safar NZ, Ndzi D, Sanders D, Noor HM,& Kamarudin LM (2016) Integration of fuzzy c-means and artificial neural network for short-term localized rainfall forecasting in tropical climate. In: Proceedings of SAI intelligent systems conference, pp 325–348. Springer, Cham

Nourani V, Baghanam AH, Adamowski J, Kisi O (2014) Applications of hybrid wavelet–artificial intelligence models in hydrology: a review. J Hydrol 514:358–377

Nourani V, Uzelaltinbulat S, Sadikoglu F, Behfar N (2019a) Artificial intelligence-based ensemble modeling for multi-station prediction of precipitation. Atmosphere 10(2):80

Nourani V, Molajou A, Uzelaltinbulat S, Sadikoglu F (2019b) Emotional artificial neural networks (EANNs) for multi-step ahead prediction of monthly precipitation; case study: northern Cyprus. Theoret Appl Climatol 138(3):1419–1434

O'Gorman PA, Dwyer JG (2018) Using machine learning to parameterize moist convection: potential for modeling of climate, climate change, and extreme events. J Adv Model Earth Syst 10(10):2548–2563

Ouyang Q, Lu W, Xin X, Zhang Y, Cheng W, Yu T (2016) Monthly rainfall forecasting using EEMD-SVR based on phase-space reconstruction. Water Resour Manag 30(7):2311–2325

Petković D, Gocić M, Shamshirband S (2016) Adaptive neuro-fuzzy computing technique for precipitation estimation. Facta Univ Ser Mech Eng 14(2):209–218

Pham QB, Abba SI, Usman AG, Linh NT, Gupta V, Malik A, Costache R, Vo ND, Tri DQ (2019) Potential of hybrid data-intelligence algorithms for multi-station modelling of rainfall. Water Resour Manag 33:5067–5087

Pham BT, Le LM, Le TT, Bui KTT, Le VM, Ly HB, Prakash I (2020) Development of advanced artificial intelligence models for daily rainfall prediction. Atmos Res 237:104845

Rumelhart DE, Hinton GE, Williams RJ (1986) Learning representations by back-propagating errors. Nature 323(9):533–536
Scher S (2018) Toward data-driven weather and climate forecasting: approximating a simple general circulation model with deep learning. Geophys Res Lett 45(22):12–616
Scher S (2020) Artificial intelligence in weather and climate prediction: learning atmospheric dynamics. Doctoral dissertation, Department of Meteorology, Stockholm University
Schultz MG, Betancourt C, Gong B, Kleinert F, Langguth M, Leufen LH, Mozaffari A, Stadtler S (2021) Can deep learning beat numerical weather prediction? Philos Trans R Soc A 379(2194):20200097
Sehad M, Lazri M, Ameur S (2017) Novel SVM-based technique to improve rainfall estimation over the Mediterranean region (north of Algeria) using the multispectral MSG SEVIRI imagery. Adv Space Res 59(5):1381–1394
Shafaei M, Adamowski J, Fakheri-Fard A, Dinpashoh Y, Adamowski K (2016) A wavelet-SARIMA-ANN hybrid model for precipitation forecasting. J Water Land Dev 28:27–36
Shenify M, Danesh AS, Gocić M, Taher RS, Abdul Wahab AW, Gani A, Shamshirband S, Petković D (2016) Precipitation estimation using support vector machine with discrete wavelet transform. Water Resour Manag 30:641–652
Silver D, Schrittwieser J, Simonyan K, Antonoglou I, Huang A, Guez A, Hubert T, Baker L, Lai M, Bolton A, Chen Y (2017) Mastering the game of go without human knowledge. Nature 550(7676):354–359
Singh P (2018) Indian summer monsoon rainfall (ISMR) forecasting using time series data: a fuzzy-entropy-neuro based expert system. Geosci Front 9(4):1243–1257
Solomatine DP, Ostfeld A (2008) Data-driven modelling: some past experiences and new approaches. J Hydroinf 10(1):3–22
Sojitra MA, Purohit RC, Pandya PA (2015) Comparative study of daily rainfall forecasting models using adaptive-neuro fuzzy inference system (ANFIS). Curr World Environ 10(2):529
Solgi A, Nourani V, Pourhaghi A (2014) Forecasting daily precipitation using hybrid model of wavelet-artificial neural network and comparison with adaptive neurofuzzy inference system (case study: Verayneh station, Nahavand). Adv Civ Eng 2014
Song C, Chen X (2021) Performance comparison of machine learning models for annual precipitation prediction using different decomposition methods. Remote Sens 13(5):1018
Sumi SM, Zaman M, Hirose H (2012) A rainfall forecasting method using machine learning models and its application to the Fukuoka city case. Int J Appl Math Comput Sci 22(4):841–854
Tahani M, Babayan N (2018) Flow Regime Algorithm (FRA): a physics-based meta-heuristics algorithm. Knowl Inf Syst 60:1001–1038
Tao H, Sulaiman SO, Yaseen ZM, Asadi H, Meshram SG, Ghorbani MA (2018) What is the potential of integrating phase space reconstruction with SVM-FFA data-intelligence model? Application of rainfall forecasting over regional scale. Water Resour Manag 32(12):3935–3959
Tao L, He X, Wang R (2017) A hybrid LSSVM model with empirical mode decomposition and differential evolution for forecasting monthly precipitation. J Hydrometeorol 18(1):159–176
Trinh TA (2018) The impact of climate change on agriculture: findings from households in Vietnam. Environ Resour Econ 71(4):897–921
Vaze J, Teng J, Chiew FHS (2011) Assessment of GCM simulations of annual and seasonal rainfall and daily rainfall distribution across south-east Australia. Hydrol Process 25(9):1486–1497
Venkata Ramana R, Krishna B, Kumar SR, Pandey NG (2013) Monthly rainfall prediction using wavelet neural network analysis. Water Resour Manag 27(10):3697–3711
Wahyuni I, Mahmudy WF, Iriany A (2017) Rainfall prediction using hybrid adaptive neuro fuzzy inference system (ANFIS) and genetic algorithm. J Telecommun Electron Comput Eng (JTEC) 9(2–8):51–56
Willard J, Jia X, Xu S, Steinbach M, Kumar V (2021) Integrating scientific knowledge with machine learning for engineering and environmental systems. ACM Comput Surv 55:1–37
Wu J, Liu M, Jin L (2010a) A hybrid support vector regression approach for rainfall forecasting using particle swarm optimization and projection pursuit technology. Int J Comput Intell Appl 9(02):87–104
Wu CL, Chau KW, Fan C (2010b) Prediction of rainfall time series using modular artificial neural networks coupled with data-preprocessing techniques. J Hydrol 389(1–2):146–167
Wu J, Long J, Liu M (2015) Evolving RBF neural networks for rainfall prediction using hybrid particle swarm optimization and genetic algorithm. Neurocomputing 148:136–142
Wu J, Xie Y (2019) Hybrid support vector regression with parallel co-evolution algorithm based on GA and PSO for forecasting monthly rainfall. J Softw Eng Appl 12(12):524–539
Xiang Y, Gou L, He L, Xia S, Wang W (2018) A SVR–ANN combined model based on ensemble EMD for rainfall prediction. Appl Soft Comput 73:874–883
Yadav P, Sagar A (2019) Rainfall prediction using artificial neural network (ANN) for Tarai region of Uttarakhand. Curr J Appl Sci Technol 33:1–7
Yang X-S (2010) Firefly algorithm, stochastic test functions and design optimization. Int J Biomed Comput 2(2):78–84. https://doi.org/10.1504/IJBIC.2010.032124
Yaseen ZM, Ebtehaj I, Kim S, Sanikhani H, Asadi H, Ghareb MI, Bonakdari H, Mohtar WW, Al-Ansari N, Shahid S (2019) Novel hybrid data-intelligence model for forecasting monthly rainfall with uncertainty analysis. Water 11(3):502
Yaseen ZM, Ghareb MI, Ebtehaj I, Bonakdari H, Siddique R, Heddam S, Yusif AA, Deo R (2018) Rainfall pattern forecasting using novel hybrid intelligent model based ANFIS-FFA. Water Resour Manag 32(1):105–122
Zeynoddin M, Bonakdari H, Azari A, Ebtehaj I, Gharabaghi B, Madavar HR (2018) Novel hybrid linear stochastic with non-linear extreme learning machine methods for forecasting monthly rainfall a tropical climate. J Environ Manag 222:190–206

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.