INDIAN JOURNAL OF SCIENCE AND TECHNOLOGY

REVIEW ARTICLE

Machine Learning for Drought Analysis and Forecasting: A Critical Review of Data Fusion, Extreme-Event Sensitivity, Explainability and Generalization

Ritik Mallik[1]*, **Dayal Kumar Behera**[1], **Satarupa Mohanty**[1]

1 School of Computer Science and Engineering, KIIT Deemed to be University, Bhubaneswar, Odisha, India

OPEN ACCESS

Received: 07/09/2026

Accepted: 28/09/2026

Published: 17/10/2026

Citation: Mallik R, Behera DK, Mohanty S (2026) Machine Learning for Drought Analysis and Forecasting: A Critical Review of Data Fusion, Extreme-Event Sensitivity, Explainability and Generalization. Indian Journal of Science and Technology 19(36): 2529-2545. https://doi.org/10.17485/IJST/v19i36.1263

* **Corresponding author**.

ritikmallik@gmail.com

Funding: None

Competing Interests: None

Copyright: © 2026 Mallik et al. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Published By Indian Society for Education and Environment (iSee)

ISSN
Print: 0974-6846
Electronic: 0974-5645

Abstract

Background: Machine learning is increasingly applied to drought monitoring and forecasting, but studies differ widely in how drought is defined, what data are used, and how models are validated, which makes reported performance difficult to compare directly. **Objectives:** To explain why reported machine-learning drought-model performance varies across definitions, data sources and validation schemes, and to assess whether recent studies adequately address rare severe events and cross-domain transferability. **Method:** Journal articles published January 2018 to August 2026 were reviewed, starting from a 70-record high-relevance subset of a 359-record discovery export and extended through citation chaining. Publisher and DOI verification produced a final evidence base of 55 articles: 48 primary studies and 7 reviews. Target definition, input data, model family, validation design, metrics, class-imbalance treatment, explainability and external testing were compared across studies. **Findings:** Reported performance varied with the modelling task rather than with any single algorithm: Random Forest reached 97.7% AUC-ROC for agricultural-drought mapping in Queensland, SVM reached average R^2=0.90 for hydrological drought in Algeria, and a CatBoost stacking model reached R^2=0.9065 for remote-sensing-based drought monitoring. These scores are not directly comparable because targets, data and validation protocols differ. Recent studies increasingly test cross-product and cross-region transfer, use SHAP/LIME explainability, and adopt extreme-aware losses, yet class-wise reporting of severe and extreme drought, ordinal-aware evaluation, and leakage-resistant temporal validation remain inconsistently applied across the 48 primary studies. **Novelty:** The review jointly evaluates predictive accuracy alongside rare-event sensitivity, ordinal drought severity, leakage-resistant validation, explainability, uncertainty and cross-domain generalization, criteria largely absent from earlier machine-learning drought reviews.

Keywords: Drought Forecasting; Machine Learning; Remote Sensing; Explainable AI; Extreme Drought

1 Introduction

1.1 Background

Drought prediction is not a single task, although it often gets treated as one in comparative tables. A model that reproduces a precipitation-based index is solving a different problem from a model built to anticipate soil-moisture stress, a streamflow deficit, or a categorical drought class. Each of these targets has its own memory and its own noise, so comparing scores across them the way many papers do is not really meaningful. Methodologically, the field has moved through several phases: conventional machine learning first, then recurrent networks, then hybrid decomposition pipelines and remote-sensing fusion, and now attention-based architectures[1]. A 2026 systematic review of deep-learning and remote-sensing studies still reported uneven geographic coverage, limited multisource fusion and inconsistent model reporting[2]. In other words, the methods kept advancing but the reporting standards did not keep pace.

Several earlier reviews raise related concerns. Prodhan et al. pointed to data quality, spatial bias and inconsistent evaluation in machine-learning drought studies[3]. Gyaneshwar et al., looking specifically at deep-learning applications, found data scarcity, imbalance and overfitting to be recurring practical problems[4]. Nandgude et al. made a related but more basic point: drought modelling is not a single standardized task, and it spans many indices, data types and technologies[5]. Alawsi et al., in their review of hybrid forecasting methods, note that the preprocessing and decomposition steps that improve prediction accuracy also make a pipeline harder to reproduce correctly[6]. Three additional reviews from 2025-2026 cover related ground: one on general trends and pitfalls in AI-based drought science[7], one a methodological review of machine- and deep-learning drought characterization[8], and one a systematic literature review of drought forecasting with machine learning[9]. Table 1 places the present review against these and against a scientometric analysis[10], comparing scope, focus and what each contributes that the others do not.

Table 1. Scope comparison with major recent drought machine-learning reviews.

Review	Coverage	Principal focus	Distinct contribution
Prodhan et al. (2022)[3]	ML drought monitoring/forecasting	Methods, data limitations, evaluation	Present review adds rare-event sensitivity, ordinal severity, leakage and transfer.
Gyaneshwar et al. (2023)[4]	123 selected studies	Deep-learning drought prediction	Present review spans conventional ML, ensembles, transfer, XAI and event-sensitive evaluation.
Nandgude et al. (2023)[5]	Broad drought-prediction literature	Models, indices and technologies	Present review narrows to 2018-2026 and emphasizes reliability of validation.
Osman et al. (2025)[1]	2000-June 2025	ML/DL/hybrid across drought types	Present review extends into 2026 and foregrounds ordinal/extreme-event evaluation.
de Oliveira Luz et al. (2026)[2]	87 DL/remote-sensing studies	DL, data sources, geography	Present review also integrates conventional ML, XAI, uncertainty and cross-product transfer.
Ascenso et al. (2026)[7]	Trends and pitfalls in AI-based drought science	AI-based drought science, priorities	Present review adds a verified 48-primary-study matrix with qualitative reliability assessment.
Naresh & Mathew (2026)[8]	ML/DL drought characterization and prediction	Methods and characterization	Present review restricts to journal-verified, DOI-checked evidence and adds ordinal/leakage criteria.
Oyarzabal et al. (2025)[9]	Systematic literature review	ML drought forecasting	Present review narrows coverage to 2018-2026 and foregrounds event-level and transfer evaluation.
Present review	2018-August 2026; 55 articles	Reliability-centered comparison	48 primary studies + 7 reviews; source and DOI verification applied.

1.2 Research Gaps

Three gaps recur across this prior literature and motivate the present review. First, reported scores are compared across studies as though they measured the same quantity, even though drought target, lead time, spatial resolution and validation design differ from paper to paper; none of the eight prior reviews in Table 1 systematically re-interprets published scores against these confounds. Second, severe and extreme drought events, the categories most relevant to early warning, are rarely evaluated separately from the overall accuracy figure, so a model can appear strong on average while missing the events that matter most. Third, validation practice for time-dependent, geographically correlated drought data is inconsistently reported: random splits,

decomposition applied before splitting, and within-domain holdouts can each overstate how a model will perform on future years, unseen basins or a different data product.

1.3 Research Questions, Novelty and Contributions

This review shifts the guiding question from "which model is best?" to "under which conditions is a reported result credible and useful?", and asks three specific questions: (i) how does reported performance vary with drought target, data source and model family; (ii) do recent studies adequately detect rare, severe and extreme drought rather than only average conditions; and (iii) how well do reported models transfer across regions, time periods and data products. The review treats predictive accuracy as one component of evidence, alongside class balance, ordered drought severity, temporal independence, external validation, explanation quality and cross-domain transfer.

The contribution of this review is threefold. It is, to the authors' knowledge, the first drought machine-learning synthesis to jointly evaluate accuracy alongside rare-event sensitivity, ordinal severity, leakage-resistant validation, explainability and cross-domain generalization within a single reliability framework (Figure 3); it verifies all 55 included articles against publisher and DOI records rather than relying on an unverified discovery export; and it derives four explicit, actionable methodological priorities (Table 4) from the comparative evidence rather than from general recommendation.

2 Review Methodology

2.1 Review Window and Search Framework

The review covers January 2018 through August 2026. This interval captures the period in which tree ensembles, recurrent deep networks, remote-sensing fusion, explainable artificial intelligence and transformer architectures became prominent in drought research. A Boolean search combined a drought term with a method term, of the form ("drought" OR "SPI" OR "SPEI" OR "PDSI" OR "SRI" OR "soil moisture" OR "streamflow drought") AND ("machine learning" OR "deep learning" OR "random forest" OR "support vector machine" OR "LSTM" OR "CNN" OR "transformer" OR "ensemble" OR "hybrid" OR "remote sensing" OR "class imbalance" OR "explainability" OR "transfer learning"). The candidate material drew from major publisher and bibliographic environments, including ScienceDirect, SpringerLink, IEEE Xplore, MDPI, Wiley Online Library ,Taylor & Francis, and was extended through backward and forward citation checking of the retained articles.

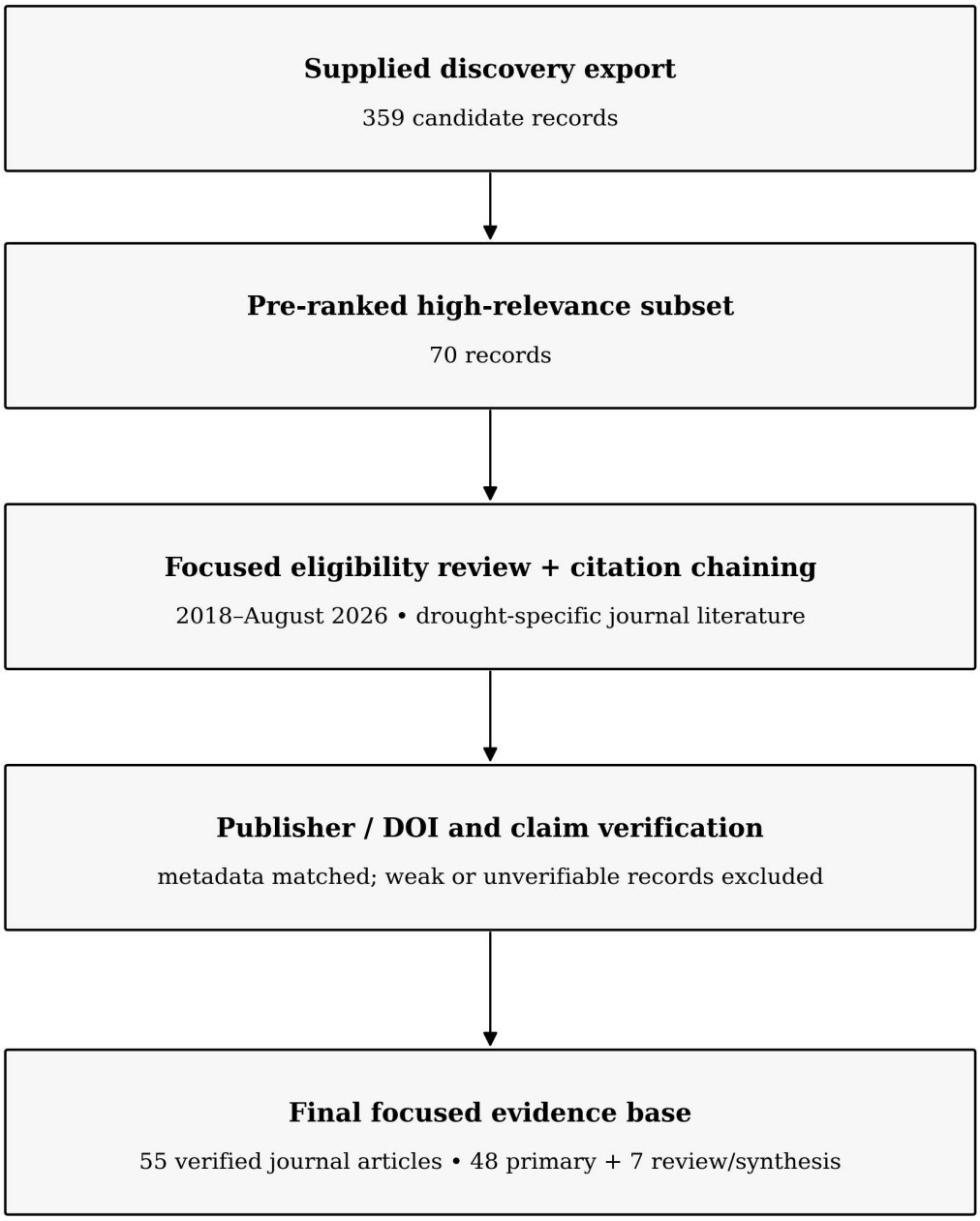
2.2 Inclusion and Exclusion Criteria

Articles were included if they (i) were published in a peer-reviewed journal between January 2018 and August 2026; (ii) reported a machine-learning, deep-learning or directly comparable data-driven model applied to a drought-related target (a standardized index, soil-moisture or streamflow deficit, or a categorical drought class); and (iii) reported at least one quantitative performance metric with a traceable publisher record or DOI. Articles were excluded if they were conference-only or preprint-only records, editorials or commentaries without new evidence, duplicate reports of the same dataset and model without added analysis, or records whose metadata could not be verified against the publisher or DOI system. Review articles and scientometric syntheses were excluded from the primary-study evidence matrix (Table 2a- 2c) but retained separately for the comparative positioning in Table 1 and for citation chaining.

2.3 Screening, Source Verification and Study-Quality Assessment

The initial discovery export held 359 records, of which 70 had already been marked highly relevant. This ranking served only as a starting pool and was not treated as a substitute for an independently run database search. Each candidate record was then checked against the Section 2.2 criteria: publication type, drought relevance, whether it fell in the review window, and whether it could be traced bibliographically at all. Metadata problems were resolved using publisher pages, DOI records and recognized indexing services. Records that were weak, duplicated, preprint-only, conference-only, or simply could not be verified were removed at this stage. Additional studies were located through citation chaining, since some relevant journal articles had not appeared prominently in the original ranking.

A further quality check was applied to the primary studies that remained. Three things were examined for each one: whether the train/test split was kept separate from any preprocessing fitted on the full dataset, whether the target definition and study region were stated clearly, and whether at least one quantitative metric was reported in enough detail to be compared with other studies. An article that failed more than one of these checks was dropped. After this process, 55 articles remained: 48 primary studies and 7 reviews or methodological syntheses. Figure 1 shows this workflow.



The 359/70 counts originate from the supplied discovery export; final inclusion was independently checked.

Fig 1. Screening and verification workflow used to construct the focused drought evidence base.

2.4 Data Extraction and Comparative Reading

The same set of questions went to every primary paper: what form of drought, which index or physical variable as the target, what data fed the model, how training and test data were kept apart, which metrics got reported, whether severe events were checked separately from the average case, and whether the model was ever interpreted, calibrated or tested outside the domain it was built on. Study design mattered here as much as the headline number. Reviews served mainly to frame the broader trends and turn up additional primary literature — they were never counted as performance evidence in their own right. [Table 1] lays out how the principal reviews compare in scope, and [Table 2a], [Table 2b] and [Table 2c] carry the full 48-study comparative matrix.

Table 2. a. Verified primary studies in the comparative evidence matrix (part 1).

Study	Drought/region	Target/data	Model/strategy	Key relevance
Tufaner & Özbeyaz (2020) [11]	Meteorological / Türkiye	PDSI-related values + meteorological variables	LR, ANN, SVM, DT	Demonstrated ML-based estimation of Palmer drought values from meteorological data.
Mokhtar et al. (2021) [12]	Meteorological / multi-region	SPEI + climate variables	Multiple ML algorithms	Showed nonlinear ML estimation of SPEI from climatic predictors.
Achite et al. (2022) [13]	Hydrological / Algeria	SPI/SRI	ANN, ANFIS, SVM, DT	SVM average RMSE 0.28, MAE 0.19, NSE 0.86, R^2 0.90; 12-month SRI R^2=0.95.
Kan et al. (2023) [14]	Agricultural / Sweden	Soil moisture, PDSI, climate and land variables	Seven ML/DL/statistical models	Temporal soil-moisture prediction R^2=0.79; multi-feature RF R^2=0.59 nationally.
Pande et al. (2023) [15]	Meteorological / Upper Godavari	Multi-scale SPI	SVM with alternative kernels	Demonstrated substantial dependence of drought skill on kernel selection.
Wang et al. (2024) [16]	Grassland drought / Inner Mongolia	Remote-sensing and drought predictors	RF and comparison models	RF showed strong regional drought-monitoring performance.
Rahmati et al. (2020) [17]	Agricultural / Queensland	RDSM + hydro-environmental factors	CART, BRT, RF, MARS, FDA, SVM	RF AUC-ROC 97.7%, TSS 0.873, efficiency 0.929, F-score 0.898.
Jehanzaib et al. (2021) [18]	Hydrological	Standardized runoff drought	Multiple ML models	Six-model seasonal SRI comparison; DT average MAE=0.26, RMSE=0.34, NSE=0.87, R^2=0.89.
Oruç et al. (2024) [19]	Meteorological / Norway	Regional drought time series	LSTM, SVM, GPR, RF + wavelet	Wavelet decomposition improved representation but adds pipeline complexity.
Wei et al. (2025) [20]	Meteorological / Songliao River Basin	Multi-scale SPEI	VMD-JAYA-Informer	Four-station 1-month NSE: 0.8663, 0.8765, 0.8822, 0.8416.
Karbasi et al. (2024) [21]	Meteorological / Eastern Canada	SPEI12	TVF-EMD + LASSO + ensemble deep RVFL	Hybrid design targeted multistep nonstationary drought forecasting.

Continued on next page

Table 2 continued

Moghaddasi et al. (2025) [22]	Meteorological / Iran	SPEI at 1-12 month lead	WSO-optimized DFFNN	Reported r=0.9961 at 1-month and r=0.7840 at 12-month lead at Chabahar.
Xu et al. (2024) [23]	Seasonal drought	Remote sensing + environmental data	19 learners + CatBoost stacking	Stack R^2=0.9065 and 0.8218 in eastern and western regions; map comparison >0.8 in multiple cases.
Ferchichi et al. (2024) [24]	Agricultural / Africa	1999-2022 multivariate remote sensing	GAN based on CNN-LSTM	Integrated spatiotemporal DL and climate-change-sensitive drought forecasting.
Dikshit & Pradhan (2021) [25]	Meteorological / New South Wales	SPI-12 + climate drivers	LSTM + SHAP	Combined forecasting with feature-attribution analysis.
Ozupek et al. (2025) [26]	Meteorological / Türkiye	SPI/SPEI + climate variables	Six ensemble learners + SHAP/LIME	768 scenarios; R^2=0.513-0.864; CatBoost best in the reported top scenario.
b. Verified primary studies in the comparative evidence matrix (part 2).				
Verma et al. (2026) [27]	Agricultural / India	Soil-moisture percentile + local/non-local drivers	Stacked LSTM + extreme-aware loss	Extreme-aware SLSTM with chronological testing; strongest 1-month regions: POD/percentage correct 0.6–1, FAR 0–0.3.
Pathania & Gupta (2026) [28]	Meteorological / India	IMDAA source + 11 target precipitation products; six climate zones	Transformer transfer learning	Transfer tested on unseen 2017–2019 period; skill varied by season and climate; fine-tuning remained useful.
Dikshit et al. (2021) [29]	Meteorological / New South Wales	SPEI-1 and SPEI-3 + hydro-meteorological predictors	LSTM compared with RF and ANN	R^2 >0.99 for drought intensity; multiclass ROC-AUC 0.83 and 0.82.
Neeti et al. (2021) [30]	Meteorological / Central India	CHIRPS, MODIS, topography, SPI-3	RF downscaling + PCA composite index	Downscaled rainfall from 5 km to 1 km and validated against 65 rain gauges.
Lotfirad et al. (2022) [31]	Meteorological / Iran	SPI and SPEI at multiple time scales	Random Forest	SPEI generally improved drought-prediction accuracy relative to SPI; skill increased with time scale.
Feng et al. (2019) [32]	Agricultural / South-Eastern Australia	30 TRMM/MODIS drought factors targeting SPEI	Bias-corrected RF, SVM, MLP	Bias-corrected RF achieved R^2=0.9 and RMSE <0.28 and outperformed SVM/MLP.

Continued on next page

Table 2 continued

He et al. (2023) [33]	Meteorological / China	Multiple precipitation products + stations + topography	Gaussian Process Regression blending	Generated high-spatial-resolution multiscale SPI by blending heterogeneous precipitation information.
Zhang et al. (2023) [34]	Integrated drought / Xinjiang	MODIS + CHIRPS precipitation, vegetation, temperature, soil moisture	ConvLSTM + comparison models	Composite index correlated >0.5 with multiscale SPI and >0.45 with 10-cm soil relative humidity.
Li et al. (2023) [35]	Meteorological / Southwest China	Multisource remote sensing + station drought indices	RF and XGBoost	5-fold CV scores 0.955 and 0.931; XGBoost drought-grade agreement with SPEI-1 >85%.
Docheshmeh Gorgij et al. (2022) [36]	Meteorological / semi-arid Iran	SPI-3, SPI-6, SPI-9, SPI-12	LSTM vs ET, VAR, MARS	LSTM consistently superior; RMSE improvements roughly 4%-27% depending on comparator and time scale.
Malik et al. (2020) [37]	Meteorological / Kumaon, India	SPI, multiple time scales	CANFIS vs MLP/MLR	Neuro-fuzzy prediction across six SPI scales; benchmarked against neural and regression models.
Yaseen et al. (2021) [38]	Meteorological / Bangladesh	SPI	ELM, RF and other data-intelligence models	ELM led most SPI scales; RF strongest for SPI-1 in reported comparison.
Li et al. (2020) [39]	Meteorological / Northeast China	SPEI at multiple scales	Penalized regression and tree ensembles	Compared 118 stations under a common test period.
Mei et al. (2022) [40]	Meteorological / Yunnan, China	Monthly MCI drought index	Deep-learning model	Extended deep learning to monthly regional MCI prediction.
Eini et al. (2023) [41]	Hydrological / Poland and projections	SRI/streamflow	Supervised ANN; direct vs indirect	Compared direct drought-index simulation with indirect streamflow-based approach.
Shamshirband et al. (2020) [42]	Hydrological	SSI/SPI/SPEI	SVR, GEP, M5 tree	M5 tree reported CC=0.8195 and RMSE=0.8186 for SSI.
c. Verified primary studies in the comparative evidence matrix (part 3).				
Almikaeel et al. (2022) [43]	Hydrological / Gidra River	Hydrological drought classes	ANN/SVM-based framework	Regional machine-learning drought classification and forecasting.

Continued on next page

Table 2 continued

Tareke & Awoke (2023) [44]	Hydrological / Ethiopia	Streamflow and drought	ANN	Used observed data plus future downscaled precipitation for long-term forecasting.
Aghelpour et al. (2020) [45]	Agricultural / Iran	PDSI, SPI, MSPI	ANFIS + ACO/GA/PSO	ACO and GA improved ANFIS performance by 45.9% and 43.2% in reported tests.
Mohammed et al. (2022) [46]	Agricultural/hydrological / Eastern Mediterranean	SPI-3/SPI-12	BG, RSS, RT, RF	Random tree training result RMSE=0.3, r=0.97.
Nafii et al. (2023) [47]	Agricultural/hydrological / Morocco	Drought indices	Machine-learning early forecasting	Integrated watershed-scale agricultural and reservoir hydrological drought forecasting.
Zhang et al. (2021) [48]	Agricultural / China	Soil-moisture drought	Meta-Gaussian prediction	Combined drought propagation, antecedent SPI and large-scale climate drivers.
Raza et al. (2022) [49]	Meteorological / Pakistan	SPTI 1-12 months	ELM vs MLP/ARIMA	ELM generally produced best reported multi-scale forecast metrics.
Poudel et al. (2024) [50]	Meteorological	SPI/SPEI	ANN, SVM, RF	Compared drought categorization and forecasting from precipitation/temperature.
Basak et al. (2022) [51]	Meteorological / Western India	SPI	Prophet vs SVR/MLR	Prophet reported more favorable short-term drought forecasting.
Akter et al. (2025) [52]	Meteorological / Bangladesh	SPI/SPEI	ANN, SVM, RF	Combined trend analysis with machine-learning prediction in Rangpur.
Acharki et al. (2023) [53]	Agricultural/meteorological / Morocco	VCI/TCI/VHI/SPI	RSS, RF, M5P, REPTree	GIS and ML comparison across a Mediterranean coastal watershed.
Al Moteri et al. (2024) [54]	Meteorological / coastal arid region	SPEI + evaporation-related inputs	Deep-learning drought-forecasting framework	Coastal arid-region forecasting using an evaporation-related drought index and climatic inputs.
Prodhan et al. (2021) [55]	Agricultural / South Asia	Remote-sensing precipitation/vegetation/soil	Deep feed-forward neural network	Regional agricultural drought monitoring from 2001-2016 satellite variables.
Li et al. (2024) [56]	Meteorological / Southwest China	SPEI at 144 stations	LSTM vs RF	LSTM outperformed the RF benchmark in the reported prediction setting.

Continued on next page

Table 2 continued

Brust et al. (2021) [57]	Categorical drought / United States	USDM D0-D4	Recurrent neural network	Forecast USDM 1-12 weeks ahead; evaluated 2017 Northern Plains flash drought.
Masroor et al. (2021) [58]	Groundwater/drought / Godavari, India	Groundwater potential + drought	AHP + Random Forest	RF regression reported R^2=0.858 for drought-groundwater relationship.

2.5 AI-Assisted Preparation Disclosure

OpenAI ChatGPT (GPT-5.6 Sol) assisted with language editing and manuscript organization

3 Overview

This section synthesizes the 48 primary studies thematically, from the choice of drought target (3.1) through conventional and deep-learning model families (3.2-3.3), remote sensing and explainability (3.4-3.5), transferability (3.6), and regional/high-resolution evidence (3.7-3.8), before drawing the findings together into comparative tables and explicit research priorities (3.9). [Table 2a], [Table 2b] and [Table 2c] give the full comparative evidence matrix referenced throughout; [Table 3] compares model families directly, and [Table 4] lists the resulting research-gap priorities.

3.1 Drought Targets Shape the Modelling Problem

"Drought" covers more than one prediction target, and that alone explains a lot of the scatter in reported accuracy. Tufaner and Ozbeyaz showed that the Palmer Drought Severity Index can be estimated directly from meteorological inputs using machine learning [11]. Mokhtar et al. obtained a comparable result for SPEI with several nonlinear algorithms applied to climatic predictors [12]. Hydrological drought is harder to predict because runoff and storage introduce delayed, basin-specific memory that meteorological indices do not have. In the Wadi Ouahrane Basin, average SVM performance reached RMSE 0.28, MAE 0.19, NSE 0.86 and R^2 0.90, with R^2 rising to 0.95 for the 12-month SRI case specifically [13]. Agricultural drought adds vegetation and soil-process complexity on top of this. In Sweden, soil moisture proved considerably more predictable than PDSI: R^2=0.79 for a temporal soil-moisture model, compared with R^2=0.59 for a national multi-feature Random Forest [14]. These differences are not really contradictions. They reflect the fact that each target has its own persistence and its own sources of noise, and a model tuned for one will not automatically transfer to another.

3.2 Conventional Machine Learning Remains a Necessary Baseline

Support vector machines and tree ensembles are sometimes described as outdated relative to deep learning, but this is not supported by the regional evidence. Pande et al. found that SVM performance for multiscale drought indices depended heavily on kernel choice alone [15]. In Inner Mongolia, Random Forest performed competitively for grassland drought monitoring against newer alternatives [16]. A Queensland agricultural-drought mapping study reported Random Forest results of 97.7% AUC-ROC, TSS 0.873, efficiency 0.929 and F-score 0.898, ahead of every other model tested in that study [17]. A six-model comparison for seasonal SRI forecasting across four Han River basin watersheds found decision tree most accurate on average, with MAE 0.26, RMSE 0.34, NSE 0.87 and R^2 0.89 [18]. What this suggests is fairly simple: a newer architecture should be compared with a properly tuned conventional baseline on the same split, target and metric before any claim of superiority is made. Comparing it instead to numbers reported for a different basin is not a valid test.

3.3 Deep, Hybrid and Ensemble Models: Gains with Added Conditions

Deep networks are generally expected to perform well when long temporal dependencies are present, and the evidence mostly supports this, although usually with additional requirements attached. Oruç et al. found that LSTM, SVM, Gaussian-process regression and Random Forest all benefited from wavelet decomposition in Norway [19]. This is worth noting because decomposition, if not confined strictly to the training data, can leak future information into the predictors without the analyst realizing it. A more elaborate pipeline combining variational mode decomposition with JAYA optimization and an Informer architecture produced one-month NSE values of 0.8663, 0.8765, 0.8822 and 0.8416 across four validation stations [20]. A related hybrid approach, combining TVF-EMD, LASSO feature selection and an ensemble deep RVFL model, targeted multistep SPEI

forecasting in eastern Canada [21]. Not all hybrid methods avoid the lead-time penalty. In Iran, correlation at one station fell from 0.9961 at a one-month horizon to 0.7840 at twelve months [22]. This kind of degradation curve is informative in a way that a single reported "best" score is not, because it shows how quickly forecast skill declines as the horizon lengthens.

3.4 Remote Sensing and Multisource Fusion

Remote sensing extends drought monitoring into areas where ground station coverage is sparse. Xu et al. tested 19 individual learners before building a CatBoost-based stacking model, which achieved R^2 of 0.9065 and 0.8218 in the eastern and western study regions and maintained correlations above 0.8 with independent drought-map products across several seasonal comparisons [23]. Ferchichi et al. combined a CNN-LSTM-based generative adversarial framework with multivariate remote-sensing data over Africa, extending the modelling into a climate-change-sensitive setting [24]. A related fusion study built an interpretable agricultural drought index directly from combined remote-sensing products [59]; this is notable because feature-level explanation was reported alongside accuracy rather than as an afterthought. The broader spatial coverage that fusion provides comes at a cost, however. Satellite, reanalysis and gauge-based products are not interchangeable, since they differ in bias, spatial resolution and retrieval physics, and a fusion study is generally only as reliable as its harmonization procedure. Ideally, conclusions from such a study should be checked against a different data product before being generalized. Figure 2 summarizes the conceptual links among drought targets, data sources, model families and reliability evaluation.

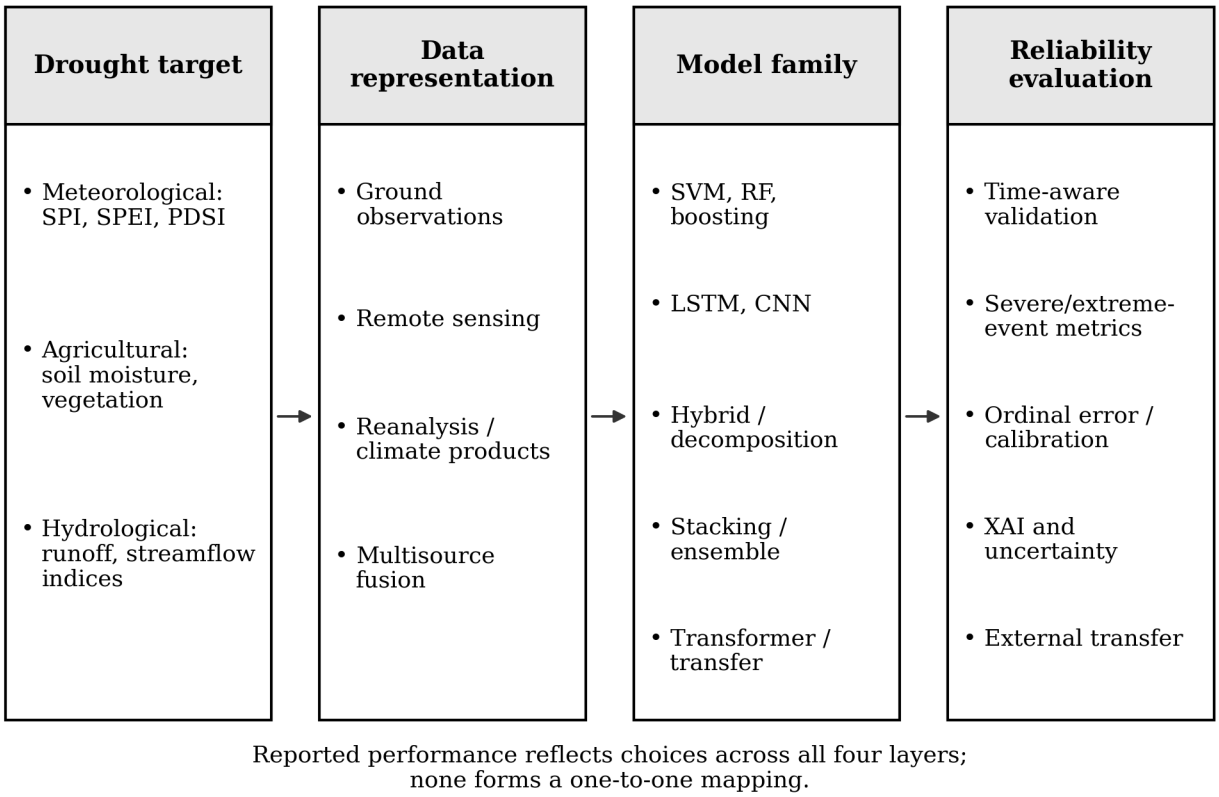


Fig 2. Conceptual links among drought targets, data sources, model families and reliability evaluation.

3.5 Explainability, Rare Events and Ordered Severity

When model output is used for water allocation or agricultural planning, some ability to explain why the model produced a given prediction becomes necessary rather than optional. Dikshit and Pradhan incorporated SHAP directly into an LSTM drought-forecasting model, treating attribution as part of the modelling pipeline rather than something added afterward [25]. A larger comparison involving six ensemble learners, four stations and 768 SPI/SPEI forecasting scenarios, using both SHAP and

LIME, reported R^2 ranging from 0.513 to 0.864, with CatBoost performing best under SPEI-12; antecedent drought indices were consistently more important than temperature or wind speed in these explanations[26]. None of this establishes causality. It does, however, allow a reader to check whether the model relies on physically reasonable variables rather than an artifact of how the data happened to be split.

Rare event categories present a separate difficulty. When normal or mildly dry conditions dominate the historical record, a model can achieve a high overall accuracy score while still failing to detect most of the severe or extreme events that actually matter for early-warning purposes. There is also the issue of ordering, which standard multiclass accuracy does not capture: mistaking an extreme event for a severe one is a smaller error than mistaking it for no drought at all, yet conventional accuracy metrics treat both mistakes equally. One study addressed the rare-event problem directly, training a stacked LSTM model with an extreme-aware loss function on data from June 1981 to May 2015 and testing it chronologically on June 2015 to May 2022. In the strongest-performing regions, one-month probability-of-detection and percentage-correct values ranged from 0.6 to 1, and false-alarm rates stayed mostly below 0.3[27]. This kind of chronological, extreme-aware evaluation is much closer to how an early-warning system would actually be used than a randomly split evaluation scored only on average regression error.

Several of the studies above address class imbalance directly rather than leaving it implicit. The extreme-aware loss function used for stacked LSTM training assigns greater weight to severe and extreme drought cases during optimization, which is one reason its probability-of-detection scores for rare events (0.6 to 1) are noticeably higher than would be expected from a model trained with a standard loss on the same imbalanced record[27]. Other studies in this evidence base rely on ensemble methods such as Random Forest and boosting, which tend to be less sensitive to class imbalance than single classifiers because they aggregate predictions across many weak learners rather than optimizing a single decision boundary; this partly explains why tree-ensemble methods report comparatively strong TSS and F-score values even for skewed agricultural-drought classes[17]. Explicit resampling of minority drought classes, class-weighted objective functions, and cost-sensitive or focal-loss variants appear only occasionally across the 48 primary studies, and are rarely benchmarked against one another within the same paper. Based on the evidence assembled here, extreme-aware and cost-sensitive objectives appear more effective than resampling alone for detecting rare severe events, mainly because resampling changes the training distribution without necessarily changing what the loss function penalizes, whereas extreme-aware losses directly target the errors that matter for early warning. Even so, this remains an inconsistently reported area, and Table 4 lists it as a priority for more systematic treatment in future work.

3.6 Transferability and Validation Reliability

Performing well within familiar conditions is not the same as building a model that generalizes to a genuinely new setting. Pathania and Gupta tested this directly using a pre-trained transformer, with IMDAA as the source domain, 11 alternative target precipitation products across six climate zones, and a period (2017-2019) that had not been seen during training[28]. Performance was maintained across most of the target products, although season and climate still had an influence, and the authors themselves noted that fine-tuning remained useful rather than unnecessary. The broader point is that similar spatial resolution does not guarantee similar climate statistics, so domain shift needs to be tested explicitly rather than assumed not to matter. Two validation practices follow from this observation. For time series data, chronological holdout or rolling-origin evaluation is preferable to unrestricted random splitting whenever the claim concerns future drought conditions. For gridded or multi-station data, leave-region-out testing against genuinely external data products provides a stronger test than withholding nearby points from within the same climatic domain. Figure 3 summarizes the reliability dimensions that should accompany predictive accuracy in this kind of work.

3.7 High-Resolution and Multiscale Evidence

Matching model architecture to spatial and temporal scale appears to help consistently, across quite different climates. In New South Wales, short-lead SPEI intensity prediction reached R^2 above 0.99, with multiclass ROC-AUC values of 0.83 and 0.82 for SPEI-1 and SPEI-3 respectively[29]. In central India, CHIRPS rainfall was downscaled from 5 km to approximately 1 km using Random Forest and multisensor covariates, and the result was checked against 65 rain gauges before SPI-based drought characteristics were derived[30]. In Iran, SPEI-based Random Forest predictions became increasingly advantageous relative to SPI-based predictions as the accumulation period lengthened[31]. In southeastern Australia, 30 TRMM/MODIS drought factors were fused using a bias-corrected Random Forest model, which reached approximately R^2=0.9 with RMSE below 0.28, outperforming both SVM and MLP in reproducing station-based SPEI[32].

Several further examples of spatially explicit modelling come from China. One study combined multiple precipitation products, station observations and topographic information using Gaussian-process regression to produce high-resolution, multiscale SPI[33]. Another built a ConvLSTM-based drought-monitoring framework for Xinjiang using MODIS and CHIRPS

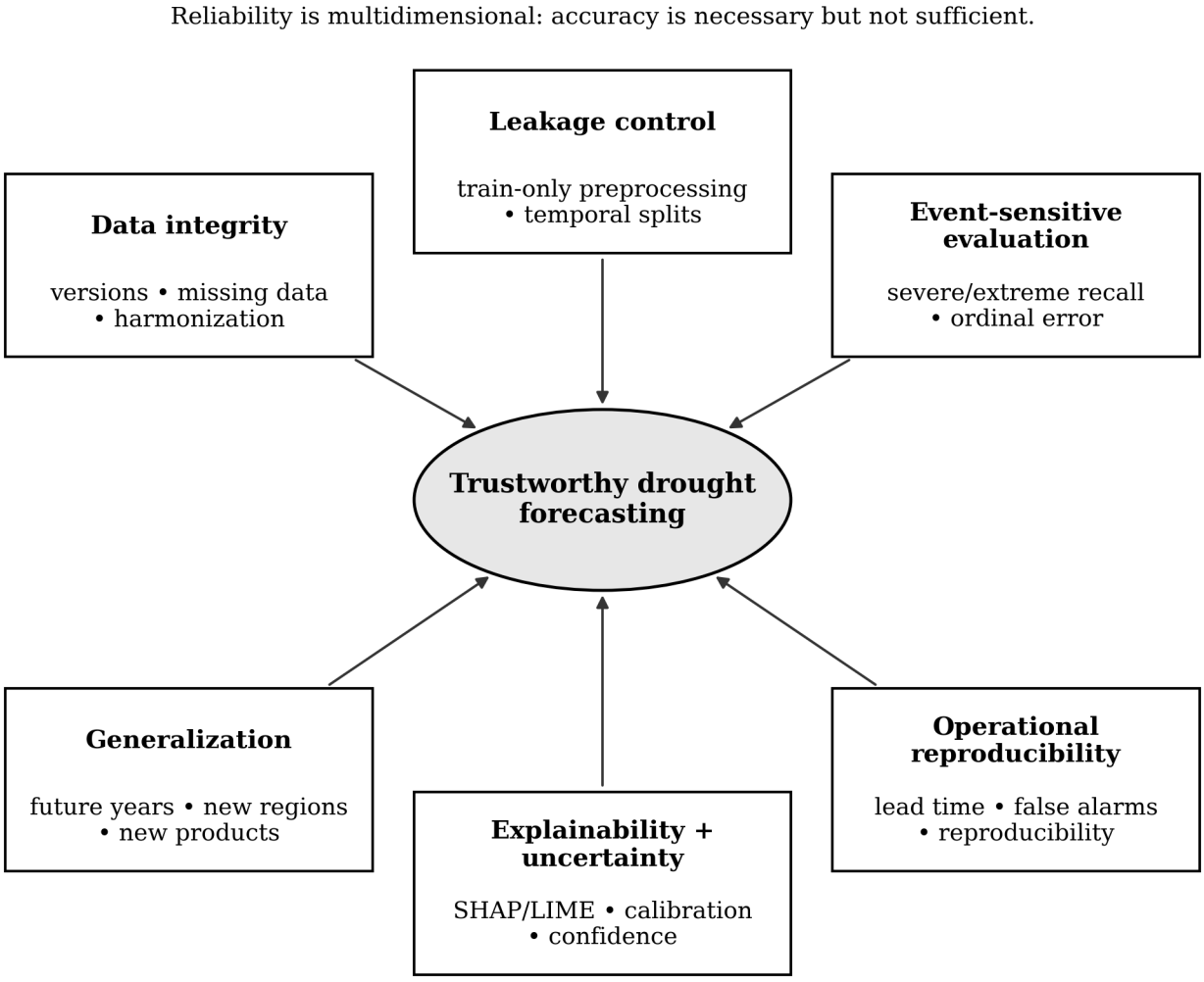


Fig 3. Six-part reliability framework for machine-learning drought forecasting and operational evaluation.

variables; the resulting composite index correlated above 0.5 with multiscale SPI and above 0.45 with 10-cm soil relative humidity[34]. A separate hydrological analysis linking deep-learning attribution methods to the climate drivers behind drought onset supports the same general point, that explicit driver attribution is preferable to relying on correlation alone[60]. In a comparison of Random Forest and XGBoost in southwest China, average five-fold cross-validation scores were 0.955 and 0.931 respectively, and XGBoost drought grades agreed with station-derived SPEI-1 more than 85% of the time[35]. Across SPI-3 through SPI-12, LSTM outperformed Extra Trees, VAR and MARS consistently, with RMSE improvements of roughly 4% to 27% depending on the comparator model and the time scale[36].

3.8 Broader Regional and Methodological Evidence

Beyond the studies that tend to dominate recent reviews, the wider literature presents a less tidy picture. A scientometric analysis tracked the growing use of artificial neural networks in drought prediction and identified deep learning and interpretability as important directions for future work[10]. In Kumaon, a co-active neuro-fuzzy inference system was compared with multilayer perceptron and multiple linear regression across several SPI scales[37]. In Bangladesh, the Extreme Learning Machine performed best at most SPI scales, although Random Forest was stronger specifically for the one-month case[38]. A comparison in Northeast China set penalized regression against tree ensembles across multiple SPEI scales[39], and a study in Yunnan extended deep learning to monthly MCI drought prediction[40]. Taken together, these four studies support a fairly straightforward conclusion: no single architecture performs best across every index and accumulation period, even within meteorological drought alone.

Hydrological studies reach a similar conclusion through different evidence. One comparison evaluated direct SRI prediction against an indirect, simulated-streamflow-based approach, under both historical and projected conditions[41]. Another modelled the standardized streamflow index using SVR, gene-expression programming and M5 model trees; the M5 tree produced CC=0.8195 and RMSE=0.8186 for SSI[42]. At the Gidra River, a single framework was used both to classify and to forecast hydrological drought[43]. In Ethiopia, an ANN-based system combined long historical records with downscaled future climate inputs for streamflow and drought forecasting[44]. What these studies have in common is that hydrological drought does not reduce to precipitation alone; catchment memory and the rainfall-to-runoff pathway shape both the choice of model and how far ahead a forecast can reasonably extend.

Agricultural drought studies increasingly combine optimization procedures, spatial data and multiple indicators within a single modelling pipeline. Bio-inspired optimization improved ANFIS performance for agricultural drought prediction, with ACO and GA producing reported gains of 45.9% and 43.2% respectively[45]. A comparison of bagging, random subspace, random tree and Random Forest in the eastern Mediterranean found random tree leading the training-stage results, with RMSE=0.3 and r=0.97[46]. In Morocco's Bouregreg Basin, an early-forecasting approach was developed covering both hydrological and agricultural drought[47]. A more process-oriented study in China predicted agricultural drought directly from antecedent meteorological drought conditions, initial soil moisture and large-scale climate drivers[48]. Across these studies, the choice of input variables appears to matter as much as the choice of learning algorithm.

Once further regional comparisons are included, the case for any single best model weakens further. The Extreme Learning Machine generally outperformed MLP and ARIMA for SPTI forecasting across nine Pakistani stations and several time scales[49]. A separate three-way comparison of ANN, SVM and Random Forest addressed both SPI/SPEI forecasting and drought categorization[50]. At 38 stations in semi-arid western India, the Prophet model outperformed SVR and multiple linear regression for short-term forecasts[51]. Two studies published in this journal extend the agricultural angle further. An empirical-mode-decomposition analysis of semi-arid agricultural productivity found that separating climatic volatility from long-term productivity trends can itself serve as an early indicator of agricultural stress, ahead of and in addition to direct drought-index forecasting[61]. A socio-economic extra-tree and boosting framework for predicting agricultural productivity offers a related, non-meteorological route toward the same early-warning objective[62]. In Rangpur, SPI and SPEI trend analysis was combined with ANN, SVM and Random Forest prediction[52]. A study of a Mediterranean coastal watershed compared several GIS-based, tree-driven methods for agricultural and meteorological drought[53]. A separate framework proposed how these model families could be combined for real-time monitoring and early warning[63]. Because these studies involve different indices and different climates, their individual results are best treated as local evidence rather than as directly comparable entries in a single ranking.

Two additional themes appear across the more recent deep-learning and operational literature. The first is representing spatial drought signals directly rather than through an intermediate index; the second is forecasting decision-oriented products rather than index values alone. A deep-learning framework for coastal arid regions incorporated an evaporation-related drought index together with climatic information[54]. A deep feed-forward network used remote-sensing precipitation, vegetation and soil data for agricultural drought monitoring across South Asia[55]. At 144 stations in southwest China, LSTM outperformed a Random Forest benchmark on the meteorological-drought task[56]. DroughtCast addressed a different kind of target altogether, using a recurrent model to forecast United States Drought Monitor categories one to twelve weeks ahead, with demonstrated skill during the 2017 Northern Plains flash drought[57]. Random Forest was also used to relate drought conditions to groundwater potential in the Godavari Middle Sub-Basin, reaching R^2=0.858 for that regression relationship[58]. This last group of studies is particularly relevant for deployment, since the model output is already close to the spatial or categorical product that a stakeholder would actually use in practice.

3.9 Comparative Synthesis and Research Directions

Across the 48 primary studies reviewed here, model performance appears to depend on the specific combination of target, data and validation choices rather than on any single algorithm being universally best. Tree ensembles perform well on structured meteorological and remote-sensing predictors. SVMs remain competitive once kernel and scale are properly tuned. LSTM-type networks are useful where persistence and long temporal memory dominate. Decomposition-based hybrid methods can help with non-stationary signals, although they carry an increased risk of information leakage if not handled carefully. Transformers offer long-range attention and transfer potential, at a higher computational cost than the alternatives. The most consistent weakness across this literature is not a shortage of new model architectures. It is the inconsistency with which those architectures are validated and compared against one another.

This raises a direct question: are the wide performance differences reported across the 48 studies mainly a property of the algorithm used, or do they mainly reflect the data, the type of drought problem, and the validation method? The evidence assembled in Sections 3.1 to 3.8 supports the second explanation. The clearest case is Section 3.1, where the same broad family of

models produces very different scores depending on whether the target is meteorological, hydrological or agricultural drought, because each target carries different memory and different noise, not because the underlying algorithm changed[13],[14]. Section 3.2 makes a related point directly: a newer architecture only earns a claim of superiority when it is compared with a well-tuned conventional baseline on the same split, target and metric; several of the highest reported scores in this review, including the 97.7% AUC-ROC result, come from Random Forest and other conventional methods rather than deep networks[17],[18]. Validation method matters just as much. The transfer-learning study in Section 3.6 shows the same model family losing and regaining accuracy purely as a function of which precipitation product and climate zone it was tested against, with no change to the algorithm itself[28]; the lead-time degradation reported in Section 3.3, where correlation fell from 0.9961 to 0.7840 for the same hybrid model, shows a comparable pattern tied to forecast horizon rather than architecture[22]. Taken together, these patterns suggest that algorithm choice is one factor among several, and typically not the dominant one, in explaining the spread of reported performance across this literature. This is also why Table 3 compares model families by suited data type and validation requirement rather than by a single ranked accuracy score.

Four priorities emerge fairly directly from this evidence. First, severe and extreme drought events should be reported separately, through class-wise recall, F1, probability of detection or similar event-level measures, rather than folded into an overall accuracy figure that can obscure them. Second, because drought categories are ordered rather than simply categorical, evaluation should use ordinal or distance-aware losses and metrics instead of purely nominal classification schemes. Third, preprocessing, decomposition, resampling and feature selection should be fitted within the training data only, with no exceptions. Fourth, interpretability should not be reported in isolation; it should be paired with uncertainty estimates and independent validation, since a plausible SHAP explanation on its own does not indicate whether a model will perform reliably in a new climate or in years it has not yet encountered. [Table 3] compares the principal model families directly, and [Table 4] summarizes the main research gaps and recommended methodological directions.

Table 3. Comparative suitability of major drought prediction model families.

Model family	Best suited for	Strengths	Limitations	Preferred validation
SVM/SVR	Small-medium tabular data	Strong nonlinear kernels; efficient	Kernel/hyperparameter sensitivity	Chronological holdout; kernel sensitivity
Random Forest	Mixed meteorological/remote-sensing predictors	Robust nonlinear fit; feature importance	Limited native long-memory modelling	External-region test; class-wise metrics
Boosting	Structured multivariate predictors	Captures interactions; strong tabular performance	Can fit regional idiosyncrasies	Blocked validation; calibration
LSTM/RNN	Long sequential histories	Learns persistence and delayed response	Large data need; overfitting risk	Rolling-origin testing; lead-time analysis
CNN	Gridded or local spatiotemporal patterns	Efficient local feature extraction	Less natural for very long dependencies	Spatial holdout + independent years
Transformer	Long-range multivariate sequences	Attention and transfer potential	High data/compute demand; domain shift	Unseen-period + cross-product/climate tests
Hybrid/decomposition	Nonstationary multiscale signals	Can isolate components and reduce noise	Complexity; leakage risk	Fit transformations within training only
Stacking/ensemble	Heterogeneous learner strengths	Combines complementary errors	Lower transparency; meta-model leakage risk	Strict out-of-fold meta-model training

4 Conclusion

This review assembled 55 publisher- and DOI-verified journal articles from the 2018-August 2026 period, comprising 48 primary studies and 7 reviews or methodological syntheses. The evidence does not support a single best drought algorithm. Random Forest reached 97.7% AUC-ROC in one agricultural-drought mapping study[17], SVM achieved average R^2=0.90 in an Algerian hydrological setting[13], and a stacking model reached R^2=0.9065 in one remote-sensing region[23]. Elsewhere, performance declined with forecast lead, climate or target scale. These values should not be compared as if they came from the same experiment.

Table 4. Research gaps and recommended methodological directions.

Gap	Why it matters	Recommended direction
Severe/extreme class imbalance	Aggregate metrics can hide missed high-impact events.	Report class-wise/event metrics; consider cost-sensitive or extreme-aware objectives.
Ordinal drought severity	Ordered categories have unequal error distances.	Use ordinal regression/classification and distance-aware agreement/error measures.
Temporal leakage	Adjacent drought observations are strongly autocorrelated.	Use blocked/rolling/future holdout and train-only preprocessing.
Cross-region transfer	Local calibration may fail in another climate.	Use leave-region-out, cross-climate and external validation.
Cross-product domain shift	Satellite/reanalysis products differ in bias and resolution.	Test alternative products and document harmonization sensitivity.
Explainability stability	One SHAP/LIME plot may not be stable across seasons/sites.	Compare explanations across season, lead time and region; check physical plausibility.
Uncertainty	Point predictions do not communicate confidence.	Add calibrated intervals/probabilities or ensemble/Bayesian uncertainty.
Reproducibility	Incomplete splits and hyperparameters hinder replication.	Report exact dates, data versions, seeds, hyperparameters and code availability.

The more consequential progress is methodological. Studies are beginning to test models on genuinely future periods, unfamiliar precipitation products and different climate zones; explainability is being used to inspect the variables behind predictions and extreme-aware losses are beginning to focus training on events that average error tends to underweight. The next generation of drought early-warning systems should build on these advances by combining event-sensitive evaluation, ordered severity, leakage-resistant validation, cross-domain transfer tests, uncertainty quantification and reproducible reporting. In that setting, predictive accuracy remains essential, but it is no longer sufficient evidence of reliability.

This review has two main limitations. First, because it draws on a supplied discovery export rather than an independently re-executed multi-database search, some eligible articles outside that export may have been missed, although citation chaining and DOI verification were used to reduce this risk. Second, because primary studies report heterogeneous metrics for different targets, the comparative synthesis is necessarily qualitative in places rather than a pooled quantitative meta-analysis. Future work extending this evidence base with a registered, multi-database systematic search and a formal meta-analysis of comparable metric families (for example, standardized RMSE or class-wise F1 for a single target) would further strengthen the conclusions drawn here.

5 Data Availability Statement

No new experimental dataset was generated for this review. The evidence base consists of the journal articles listed in the References. The study-selection and citation-verification records can be retained as supplementary material for editorial checking.

6 Details of Ethical Clearance

Ethical clearance was not required because the work is a literature review and did not involve human participants, animals, clinical intervention or collection of personally identifiable information.

References

1) Osman AIA, AlDahoul N, Chong KL, Huang YF, Ng JL, Elshafie A, et al. A review on machine learning models for drought monitoring and forecasting. *Climate Risk Management*. 2025;50:100758. https://doi.org/10.1016/j.crm.2025.100758.
2) de Oliveira Luz AE, Negri RG, Santos LBL. Deep learning approaches for drought forecasting, monitoring and impact assessment: A comprehensive systematic review. *Applied Soft Computing*. 2026;197:115196. https://doi.org/10.1016/j.asoc.2026.115196.
3) Prodhan FA, Zhang J, Hasan SS, Sharma TPP, Mohana HP. A review of machine learning methods for drought hazard monitoring and forecasting: Current research trends, challenges, and future research directions. *Environmental Modelling & Software*. 2022;149:105327. https://doi.org/10.1016/j.envsoft.2022.105327.
4) Gyaneshwar A, Mishra A, Chadha U, Vincent PMDR, Rajinikanth V, Ganapathy GP, et al. A contemporary review on deep learning models for drought prediction. *Sustainability*. 2023;15(7):6160. https://doi.org/10.3390/su15076160.

5) Nandgude N, Singh TP, Nandgude S, Tiwari M. Drought prediction: A comprehensive review of different drought prediction models and adopted technologies. *Sustainability*. 2023;15(15):11684. https://doi.org/10.3390/su151511684.
6) Alawsi MA, Zubaidi SL, Al-Bdairi NSS, Al-Ansari N, Hashim K. Drought forecasting: A review and assessment of the hybrid techniques and data pre-processing. *Hydrology*. 2022;9(7):115. https://doi.org/10.3390/hydrology9070115.
7) Ascenso G, Giuliani M, Perez-Aracil J, Salcedo-Sanz S, Bertini C, Bonetti P, et al. The maturation of AI in drought science: A review of trends, pitfalls and priorities. *Water Resources Research*. 2026;62:e2025WR041828. https://doi.org/10.1029/2025WR041828.
8) Naresh C, Mathew A. Machine and deep learning approaches for drought characterization and prediction: A comprehensive review. *Environmental Modeling & Assessment*. 2026. https://doi.org/10.1007/s10666-026-10131-8.
9) Oyarzabal RS, Santos LBL, Cunningham C, et al. Forecasting drought using machine learning: A systematic literature review. *Natural Hazards*. 2025;121:9823–9851. https://doi.org/10.1007/s11069-025-07195-2.
10) Dikshit A, Pradhan B, Santosh M. Artificial neural networks in drought prediction in the 21st century—A scientometric analysis. *Applied Soft Computing*. 2022;114:108080. https://doi.org/10.1016/j.asoc.2021.108080.
11) Tufaner F, Özbeyaz A. Estimation and easy calculation of the Palmer Drought Severity Index from the meteorological data by using the advanced machine learning algorithms. *Environmental Monitoring and Assessment*. 2020;192(9):576. https://doi.org/10.1007/s10661-020-08539-0.
12) Mokhtar A, Jalali M, He H, Al-Ansari N, Elbeltagi A, Alsafadi K, et al. Estimation of SPEI meteorological drought using machine learning algorithms. *IEEE Access*. 2021;9:65503–65523. https://doi.org/10.1109/ACCESS.2021.3074305.
13) Achite M, Jehanzaib M, Elshaboury N, Kim TW. Evaluation of machine learning techniques for hydrological drought modeling: A case study of the Wadi Ouahrane Basin in Algeria. *Water*. 2022;14(3):431. https://doi.org/10.3390/w14030431.
14) Kan JC, Ferreira CSS, Destouni G, Pan H, Passos MV, Barquet K, et al. Predicting agricultural drought indicators: ML approaches across wide-ranging climate and land use conditions. *Ecological Indicators*. 2023;154:110524. https://doi.org/10.1016/j.ecolind.2023.110524.
15) Pande CB, Kushwaha NL, Orimoloye IR, Kumar R, Abdo HG, Tolche AD, et al. Comparative assessment of improved SVM method under different kernel functions for predicting multi-scale drought index. *Water Resources Management*. 2023;37(3):1367–1399. https://doi.org/10.1007/s11269-023-03440-0.
16) Wang Y, Cui J, Miao B, Li Z, Wang Y, Jia C, et al. Evaluating performance of multiple machine learning models for drought monitoring: A case study of typical grassland in Inner Mongolia. *Land*. 2024;13(6):754. https://doi.org/10.3390/land13060754.
17) Rahmati O, Falah F, Dayal KS, Deo RC, Mohammadi F, Biggs T, et al. Machine learning approaches for spatial modeling of agricultural droughts in the south-east region of Queensland Australia. *Science of the Total Environment*. 2020;699:134230. https://doi.org/10.1016/j.scitotenv.2019.134230.
18) Jehanzaib M, Idrees MB, Kim D, Kim TW. Comprehensive evaluation of machine learning techniques for hydrological drought forecasting. *Journal of Irrigation and Drainage Engineering*. 2021;147(7):04021022. https://doi.org/10.1061/(ASCE)IR.1943-4774.0001575.
19) Oruç S, Hınıs MA, Tuğrul T. Evaluating performances of LSTM, SVM, GPR, and RF for drought prediction in Norway: A wavelet decomposition approach on regional forecasting. *Water*. 2024;16(23):3465. https://doi.org/10.3390/w16233465.
20) Wei J, Tang W, Ditthakit P, Shang J, Guo H, Zhao B, et al. Enhancing drought forecast accuracy through Informer model optimization. *Land*. 2025;14(1):126. https://doi.org/10.3390/land14010126.
21) Karbasi M, Ali M, Farooque AA, Jamei M, Khosravi K, Cheema SJ, et al. Robust drought forecasting in Eastern Canada: Leveraging EMD-TVF and ensemble deep RVFL for SPEI index forecasting. *Expert Systems with Applications*. 2024;256:124900. https://doi.org/10.1016/j.eswa.2024.124900.
22) Moghaddasi M, Moradi M, Ghaleni MM, Yaseen ZM. Reinforcing long lead time drought forecasting with a novel hybrid deep learning model: A case study in Iran. *Applied Water Science*. 2025;15:51. https://doi.org/10.1007/s13201-025-02377-6.
23) Xu X, Chen F, Wang B, Harrison MT, Chen Y, Liu K, et al. Unleashing the power of machine learning and remote sensing for robust seasonal drought monitoring: A stacking ensemble approach. *Journal of Hydrology*. 2024;634:131102. https://doi.org/10.1016/j.jhydrol.2024.131102.
24) Ferchichi A, Chihaoui M, Ferchichi A. Spatio-temporal modeling of climate change impacts on drought forecast using Generative Adversarial Network: A case study in Africa. *Expert Systems with Applications*. 2024;238:122211. https://doi.org/10.1016/j.eswa.2023.122211.
25) Dikshit A, Pradhan B. Explainable AI in drought forecasting. *Machine Learning with Applications*. 2021;6:100192. https://doi.org/10.1016/j.mlwa.2021.100192.
26) Ozupek E, Teke A, Celik N, Kavzoglu T. Explainable artificial intelligence to explore the intrinsic characteristics of climatic parameters governing meteorological drought forecasting: Opening the black box. *Stochastic Environmental Research and Risk Assessment*. 2025;39:3201–3222. https://doi.org/10.1007/s00477-025-03007-y.
27) Verma S, Karthikeyan L, Ghosh S. An explainable deep learning framework for sub-seasonal to seasonal agricultural drought prediction over India. *Agricultural Water Management*. 2026;333:110597. https://doi.org/10.1016/j.agwat.2026.110597.
28) Pathania A, Gupta V. Transfer learning for transformer-based drought forecasting across diverse precipitation products in India. *Journal of Geophysical Research: Machine Learning and Computation*. 2026;3:e2025JH001216. https://doi.org/10.1029/2025JH001216.
29) Dikshit A, Pradhan B, Huete A. An improved SPEI drought forecasting approach using the long short-term memory neural network. *Journal of Environmental Management*. 2021;283:111979. https://doi.org/10.1016/j.jenvman.2021.111979.
30) Neeti N, Murali CMA, Chowdary VM, Rao NH, Kesarwani M. Integrated meteorological drought monitoring framework using multi-sensor and multi-temporal earth observation datasets and machine learning algorithms: A case study of central India. *Journal of Hydrology*. 2021;601:126638. https://doi.org/10.1016/j.jhydrol.2021.126638.
31) Lotfirad M, Esmaeili-Gisavandani H, Adib A. Drought monitoring and prediction using SPI, SPEI, and random forest model in various climates of Iran. *Journal of Water and Climate Change*. 2022;13(2):383–406. https://doi.org/10.2166/wcc.2021.287.
32) Feng P, Wang B, Liu DL, Yu Q. Machine learning-based integration of remotely-sensed drought factors can improve the estimation of agricultural drought in South-Eastern Australia. *Agricultural Systems*. 2019;173:303–316. https://doi.org/10.1016/j.agsy.2019.03.015.
33) He Q, Wang M, Liu K, Li B, Jiang Z. Spatiotemporal analysis of meteorological drought across China based on the high-spatial-resolution multiscale SPI generated by machine learning. *Weather and Climate Extremes*. 2023;40:100567. https://doi.org/10.1016/j.wace.2023.100567.
34) Zhang Y, Xie D, Tian W, Zhao H, Geng S, Lu H, et al. Construction of an integrated drought monitoring model based on deep learning algorithms. *Remote Sensing*. 2023;15(3):667. https://doi.org/10.3390/rs15030667.
35) Li X, Jia H, Wang L. Remote sensing monitoring of drought in Southwest China using Random Forest and eXtreme Gradient Boosting methods. *Remote Sensing*. 2023;15(19):4840. https://doi.org/10.3390/rs15194840.
36) Gorgij AD, Alizamir M, Kisi O, Elshafie A. Drought modelling by standard precipitation index (SPI) in a semi-arid climate using deep learning method: Long short-term memory. *Neural Computing and Applications*. 2022;34:2425–2442. https://doi.org/10.1007/s00521-021-06505-6.

37) Malik A, Kumar A, Salih SQ, Kim S, Kim NW, Yaseen ZM, et al. Drought index prediction using advanced fuzzy logic model: Regional case study over Kumaon in India. *PLOS ONE*. 2020;15(5):e0233280. https://doi.org/10.1371/journal.pone.0233280.
38) Yaseen ZM, Ali M, Sharafati A, Al-Ansari N, Shahid S. Forecasting standardized precipitation index using data intelligence models: Regional investigation of Bangladesh. *Scientific Reports*. 2021;11:3435. https://doi.org/10.1038/s41598-021-82977-9.
39) Li Z, Chen T, Wu Q, Xia G, Chi D. Application of penalized linear regression and ensemble methods for drought forecasting in Northeast China. *Meteorology and Atmospheric Physics*. 2020;132(1):113–130. https://doi.org/10.1007/s00703-019-00675-8.
40) Mei P, Liu J, Liu C, Liu J. A deep learning model and its application to predict the monthly MCI drought index in the Yunnan Province of China. *Atmosphere*. 2022;13(12):1951. https://doi.org/10.3390/atmos13121951.
41) Eini MR, Najminejad F, Piniewski M. Direct and indirect simulating and projecting hydrological drought using a supervised machine learning method. *Science of the Total Environment*. 2023;898:165523. https://doi.org/10.1016/j.scitotenv.2023.165523.
42) Shamshirband S, Hashemi S, Salimi H, Samadianfard S, Asadi E, Shadkani S, et al. Predicting Standardized Streamflow Index for hydrological drought using machine learning models. *Engineering Applications of Computational Fluid Mechanics*. 2020;14(1):339–350. https://doi.org/10.1080/19942060.2020.1715844.
43) Almikaeel W, Čubanová L, Šoltész A. Hydrological drought forecasting using machine learning—Gidra River case study. *Water*. 2022;14(3):387. https://doi.org/10.3390/w14030387.
44) Tareke KA, Awoke AG. Hydrological drought forecasting and monitoring system development using artificial neural network (ANN) in Ethiopia. *Heliyon*. 2023;9(2):e13287. https://doi.org/10.1016/j.heliyon.2023.e13287.
45) Aghelpour P, Bahrami-Pichaghchi H, Kisi O. Comparison of three different bio-inspired algorithms to improve ability of neuro fuzzy approach in prediction of agricultural drought, based on three different indexes. *Computers and Electronics in Agriculture*. 2020;170:105279. https://doi.org/10.1016/j.compag.2020.105279.
46) Mohammed S, Elbeltagi A, Bashir B, Alsafadi K, Alsilibe F, Alsalman A, et al. A comparative analysis of data mining techniques for agricultural and hydrological drought prediction in the eastern Mediterranean. *Computers and Electronics in Agriculture*. 2022;197:106925. https://doi.org/10.1016/j.compag.2022.106925.
47) Nafii A, Taleb A, Mesbahi ME, Ezzaouini MA, Bilali AE. Early forecasting hydrological and agricultural droughts in the Bouregreg Basin using a machine learning approach. *Water*. 2023;15(1):122. https://doi.org/10.3390/w15010122.
48) Zhang Y, Hao Z, Feng S, Zhang X, Xu Y, Hao F. Agricultural drought prediction in China based on drought propagation and large-scale drivers. *Agricultural Water Management*. 2021;255:107028. https://doi.org/10.1016/j.agwat.2021.107028.
49) Raza MA, Almazah MMA, Ali Z, Hussain I, Al-Duais FS. Application of extreme learning machine algorithm for drought forecasting. *Complexity*. 2022;2022:4998200. https://doi.org/10.1155/2022/4998200.
50) Poudel B, Dahal D, Banjara M, Kalra A. Assessing meteorological drought patterns and forecasting accuracy with SPI and SPEI using machine learning models. *Forecasting*. 2024;6(4):1026–1044. https://doi.org/10.3390/forecast6040051.
51) Basak A, Rahman ATMS, Das J, Hosono T, Kisi O. Drought forecasting using the Prophet model in a semi-arid climate region of western India. *Hydrological Sciences Journal*. 2022;67(9):1397–1417. https://doi.org/10.1080/02626667.2022.2082876.
52) Akter ML, Rahman MN, Azim SA, Rony MRH, Sohel MS, Abdo HG. Estimation of drought trends and comparison between SPI and SPEI with prediction using machine learning models in Rangpur, Bangladesh. *Geology, Ecology, and Landscapes*. 2025;9(2):596–610. https://doi.org/10.1080/24749508.2023.2254003.
53) Acharki S, Singh SK, do Couto EV, Arjdal Y, Elbeltagi A. Spatio-temporal distribution and prediction of agricultural and meteorological drought in a Mediterranean coastal watershed via GIS and machine learning. *Physics and Chemistry of the Earth*. 2023;131:103425. https://doi.org/10.1016/j.pce.2023.103425.
54) Moteri MA, Alrowais F, Mtouaa W, Aljehane NO, Alotaibi SS, Marzouk R, et al. An enhanced drought forecasting in coastal arid regions using deep learning approach with evaporation index. *Environmental Research*. 2024;246:118171. https://doi.org/10.1016/j.envres.2024.118171.
55) Prodhan FA, Zhang J, Yao F, Shi L, Sharma TPP, Zhang D, et al. Deep learning for monitoring agricultural drought in South Asia using remote sensing data. *Remote Sensing*. 2021;13(9):1715. https://doi.org/10.3390/rs13091715.
56) Li X, Jia H, Wang L, Xiao T. Prediction and assessment of meteorological drought in southwest China using long short-term memory model. *Open Geosciences*. 2024;16(1):20220708. https://doi.org/10.1515/geo-2022-0708.
57) Brust C, Kimball JS, Maneta MP, Jencso K, Reichle RH. DroughtCast: A machine learning forecast of the United States Drought Monitor. *Frontiers in Big Data*. 2021;4:773478. https://doi.org/10.3389/fdata.2021.773478.
58) Masroor M, Rehman S, Sajjad H, Rahaman MH, Sahana M, Ahmed R, et al. Assessing the impact of drought conditions on groundwater potential in Godavari Middle Sub-Basin, India using analytical hierarchy process and random forest machine learning algorithm. *Groundwater for Sustainable Development*. 2021;13:100554. https://doi.org/10.1016/j.gsd.2021.100554.
59) Chen H, Yang N, Song X, Lu C, Lu M, Chen T, et al. A novel agricultural drought index based on multi-source remote sensing data and interpretable machine learning. *Agricultural Water Management*. 2025;308:109303. https://doi.org/10.1016/j.agwat.2025.109303.
60) Tripathy KP, Mishra A. Deep learning reveals the role of climate drivers in shaping hydrological droughts. *Water Resources Research*. 2026;62(4):e2025WR042855. https://doi.org/10.1029/2025WR042855.
61) Reddy RK, Shivaprakash. Disentangling climatic volatility from long-term productivity trends in semi-arid agriculture: An empirical mode decomposition approach. *Indian Journal of Science and Technology*. 2025;18(43):3445–3455. https://doi.org/10.17485/IJST/v18i43.1671.
62) Naik B, Dash PB, Nayak J. Extra-tree learning based socio-economic factor analysis and multi-class adaptive boosting meta-estimator for prediction of agricultural productivity. *Indian Journal of Science and Technology*. 2020;13(29):2981–3001. https://doi.org/10.17485/IJST/v13i29.839.
63) Niaz R, Ilyas M, Persio LD, Mehr AD, Ahmad N, Cheng H, et al. Machine learning integration for robust real-time drought monitoring and early warning systems. *Geomatics, Natural Hazards and Risk*. 2026;17(1):2709220. https://doi.org/10.1080/19475705.2026.2709220.