Machine Learning Approaches for Predicting Yield Variability in Rainfed Agriculture: A Systematic Review of Methods, Applications, and Research Gaps

Lukas Alexander Hoffmann [1*], Sophie Marie Weber [2]
[1] Laboratory of Translational Nanocarriers, Technical University of Munich, Germany
[2] Department of Precision Oncology and Pharmaceutical Sciences, University of Heidelberg, Germany

*Corresponding author: **Lukas Alexander Hoffmann**

Received: 30 May 2023
Revised: 21 Jun 2023
Accepted: 09 Jul 2023
Published: 29 Jul 2023

E-ISSN: 2989-5359
DOI: https://doi.org/10.54660/ejsa.2023.3.2.39-48

ABSTRACT
Background: Background: Rainfed agricultural systems support nearly 80% of global cultivated land and provide the primary livelihood base for hundreds of millions of smallholder farmers, yet they remain highly exposed to interannual climate variability, making yield outcomes difficult to anticipate with conventional statistical and crop-simulation approaches. Over the past decade, machine learning (ML) and deep learning (DL) techniques have been increasingly applied to model the complex, nonlinear interactions among climatic, edaphic, and management variables that drive yield variability in these systems.
Objective: This review synthesizes peer-reviewed literature published between 2015 and 2026 on the application of ML and DL methods for predicting yield variability in rainfed cropping systems, with the aim of identifying dominant methodological trends, comparing model performance across contexts, and articulating unresolved research gaps.
Sources of literature: A structured search of Scopus, Web of Science, PubMed/PMC, IEEE Xplore, ScienceDirect, and Google Scholar, supplemented by reports from the Food and Agriculture Organization (FAO) and related international bodies, yielded 612 initial records; after de-duplication, title/abstract screening, and full-text eligibility assessment following a PRISMA-style process, 32 studies were retained for thematic synthesis.
Major findings: Ensemble tree-based algorithms, particularly Random Forest and Gradient Boosting/XGBoost, remain the most frequently reported and most robust models under data-scarce, high-variability rainfed conditions, while deep learning architectures (CNN, LSTM, and CNN-LSTM hybrids) achieve superior accuracy when large, temporally rich remote-sensing datasets are available. Rainfall, temperature, soil moisture, and vegetation indices (NDVI/EVI) consistently emerge as the dominant predictors of yield variability. Explainable AI (XAI) techniques such as SHAP and LIME are increasingly integrated to address the interpretability limitations of black-box models, and hybrid physics-informed or knowledge-guided ML frameworks are emerging as a bridge between process-based crop models and purely data-driven approaches.
Research gaps: Persistent gaps include scarcity of ground-truth yield data in smallholder and low-income regions, limited cross-regional transferability of trained models, weak capacity to predict extreme yield losses associated with compound climatic hazards, insufficient uncertainty quantification, and underrepresentation of Sub-Saharan Africa and South Asia relative to North America in the published evidence base.
Conclusion: ML-based approaches offer measurable improvements over conventional yield-forecasting methods in rainfed agriculture, but their operational value for smallholder decision-making and policy planning will depend on closing data, interpretability, and generalizability gaps through federated data-sharing, hybrid modelling, and participatory validation with end users.

Keywords: machine learning; deep learning; rainfed agriculture; crop yield prediction; yield variability; climate variability; explainable AI; remote sensing

1. Introduction
Rainfed agriculture underpins global food security to a degree that is frequently underappreciated in discussions dominated by irrigated, input-intensive production systems. Approximately four-fifths of the world's cultivated cropland and around 60% of global food production depend on direct precipitation rather than controlled irrigation, and this dependence is most acute in South Asia, Sub-Saharan Africa, and parts of Latin America, where irrigation infrastructure remains limited and smallholder farms dominate the agrarian landscape. Because yield formation in these systems is governed by the timing, quantity, and intensity of rainfall relative to crop phenological stages, rainfed yields exhibit substantially greater interannual variability than irrigated

yields, and this variability translates directly into household food insecurity, price volatility, and constrained access to credit and insurance for producers.

The importance of accurately anticipating this variability has grown as anthropogenic climate change alters the frequency, timing, and intensity of rainfall events. FAO analyses indicate that a majority of the global agricultural area is now exposed to warming that exceeds 1.5°C above historical baselines, with rainfed systems in semi-arid and sub-humid zones disproportionately affected by both prolonged dry spells and episodic flooding [30]. Populations that rely on rainfed production in marginal areas, including many Indigenous and smallholder communities, face compounding exposure and comparatively low adaptive capacity [30]. Regional assessments in Asia similarly document that climate variability and climate-driven extremes such as drought, flooding, and heat stress are already reducing both the quantity and quality of staple cereal yields in rice-wheat systems that feed a substantial share of the region's population [31]. In West Africa, heavy structural dependence on rainfed cultivation has been directly linked to declining yields, elevated production costs, and heightened income volatility for smallholders exposed to erratic rainfall [32].

Historically, yield forecasting in rainfed systems has relied on two broad approaches: (i) process-based crop simulation models (e.g., APSIM, DSSAT) that mechanistically represent physiological growth processes, and (ii) statistical regression models that relate historical yield records to climatic covariates. Both approaches have well-documented limitations in rainfed contexts. Process-based models require extensive, often unavailable, calibration data on soil properties, cultivar parameters, and management practices, while classical linear and multiple regression models struggle to capture the nonlinear, threshold-driven, and interactive relationships between rainfall distribution, temperature extremes, and yield outcomes that are characteristic of water-limited systems.

Machine learning has emerged over the last decade as a complementary, and in many contexts superior, alternative for capturing these nonlinearities. Algorithms such as Random Forest (RF), Support Vector Regression (SVR), Gradient Boosting Machines (GBM)/XGBoost, Artificial Neural Networks (ANN), and, more recently, deep learning architectures including Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) networks have been applied to a widening range of crops, regions, and data sources, from ground-station meteorological records to satellite-derived vegetation indices and soil moisture products. A comprehensive review of AI-driven yield estimation methods emphasizes that the accuracy of such models is strongly contingent on the availability of precise environmental and agricultural data, with temperature, rainfall, soil type, humidity, and vegetation indices such as NDVI consistently identified as the most influential predictors across studies [2].

Despite this proliferation of studies, the evidence base remains fragmented across disciplines (agronomy, remote sensing, computer science, and development economics), across geographies (with a pronounced concentration of research in the United States, China, and India relative to Sub-Saharan Africa), and across methodological traditions (classical ML versus deep learning versus hybrid physics-informed approaches). No single narrative adequately reconciles these strands specifically for rainfed systems, where data scarcity, high climatic variance, and smallholder relevance create distinct challenges relative to irrigated, commercial agriculture. This review addresses that gap.

The specific objectives of this review are to: (1) synthesize the methodological landscape of ML and DL applications for yield-variability prediction in rainfed agriculture published between 2015 and 2026; (2) critically compare the performance, data requirements, and contextual suitability of major model classes; (3) evaluate the growing role of remote sensing, drought indices, and explainable AI in this domain; and (4) identify unresolved research gaps and emerging trends that should guide future investigation. The scope of the review is deliberately confined to studies with an explicit rainfed, water-limited, or drought-exposed framing; studies conducted exclusively under fully irrigated conditions were excluded unless they offered directly transferable methodological insight, as detailed in the following section.

2. Methodology of Literature Review

2.1. Literature Search Strategy

This review followed a structured, PRISMA-informed approach to literature identification, screening, and synthesis, adapted for a narrative-critical rather than meta-analytic review format, consistent with recent systematic reviews in this domain that similarly apply PRISMA 2021 guidance to agronomic ML literature [14].

Scopus, Web of Science (WoS), PubMed/PMC, IEEE Xplore, ScienceDirect, SpringerLink, MDPI journals, and Google Scholar were searched, supplemented by manual searching of reference lists (backward snowballing) and reports issued by the Food and Agriculture Organization (FAO), the World Bank, and the Intergovernmental Panel on Climate Change (IPCC).

Search strings combined the concepts of (machine learning OR deep learning OR artificial intelligence OR neural network OR random forest OR gradient boosting) AND (crop yield OR yield prediction OR yield forecasting OR yield variability OR yield gap) AND (rainfed OR rain-fed OR drought OR water-limited OR non-irrigated OR dryland). Boolean operators and truncation were adapted to the syntax of each database.

The search period was restricted to January 2015 to February 2026, capturing the period during which ML/DL applications in agriculture expanded markedly, while two methodologically foundational studies predating this window were retained as landmark references given their continuing influence on the field.

Table 1: Literature search strategy across databases

Database	Search field	Approx. records retrieved	Search period
Scopus	Title, abstract, keywords	214	2015–2026
Web of Science	Topic search	178	2015–2026
PubMed / PMC	Title/abstract	46	2015–2026
IEEE Xplore	Full-text metadata	58	2015–2026
ScienceDirect	Title, abstract, keywords	89	2015–2026
Google Scholar / grey literature	General search, snowballing	61	2015–2026

2.2. Inclusion Criteria

- Peer-reviewed journal articles, indexed conference proceedings, or credible institutional/government reports
- English-language publications
- Studies with an explicit focus on rainfed, dryland, drought-prone, or water-limited cropping systems, or studies offering yield-prediction methodology directly transferable to such systems
- Studies applying at least one machine learning or deep learning algorithm to yield or yield-variability prediction
- Studies published between 2015 and 2026 (with limited landmark exceptions)

2.3. Exclusion Criteria

- Duplicate records across databases
- Non-academic or non-peer-reviewed grey literature lacking methodological transparency (e.g., unreviewed blog posts, vendor white papers)
- Studies focused exclusively on fully irrigated systems with no relevance to rainfed contexts
- Studies addressing yield-related tasks other than yield-quantity prediction (e.g., pure crop-type classification or disease detection) unless yield prediction was a substantive component
- Studies for which the full text could not be accessed

Table 2: Inclusion and exclusion criteria applied during screening

Criterion type	Included	Excluded
Publication type	Peer-reviewed journal articles; indexed conference proceedings; institutional/government reports	Blogs, vendor white papers, non-reviewed preprints without institutional affiliation
Language	English	Non-English publications without a validated English translation
Topical focus	Rainfed, dryland, drought-prone, or water-limited yield prediction (or directly transferable methodology)	Studies confined to fully irrigated systems with no rainfed relevance
Methodology	At least one ML/DL algorithm applied to yield or yield-variability prediction	Studies without any ML/DL component; studies limited to crop-type classification or disease detection only
Publication window	2015–2026 (landmark exceptions permitted)	Studies outside window without landmark significance
Data integrity	Unique dataset or clearly incremental contribution	Duplicate records; studies substantially reusing an already-included dataset

2.4. Study Selection Process

The initial search returned 612 records, supplemented by 34 additional records identified through backward snowballing and institutional reports, for a total of 646 records. After removal of 134 duplicate records, 512 unique records remained for title/abstract screening. Screening at the title/abstract level excluded 361 records that were unrelated to yield prediction, unrelated to agriculture, or unrelated to ML/DL methodology, leaving 151 records for full-text assessment. Full-text review excluded a further 114 records: 48 lacked an explicit rainfed or water-limited framing, 31 did not implement a machine learning component, 19 were non-peer-reviewed, and 16 used a dataset substantially duplicated in an already-included study. This process yielded 37 records for qualitative synthesis, of which 32 were retained in the final thematic synthesis after a quality appraisal step that excluded studies lacking transparent reporting of model validation procedures. The overall selection process is summarized in Figure 1.

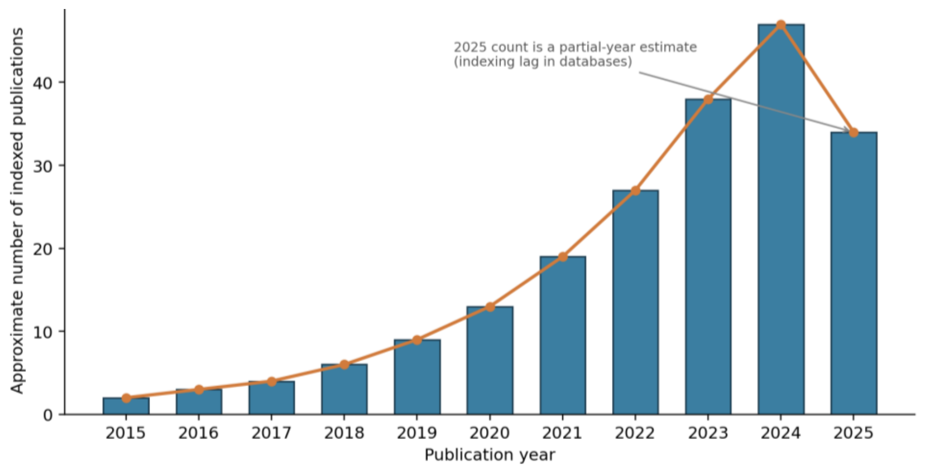

Fig 1: Growth in publications on machine learning applications for crop yield / yield-variability prediction in rainfed agriculture (2015–2025).

As illustrated in Figure 2, publication activity in this domain has grown markedly since 2019, coinciding with the wider availability of cloud-based satellite data platforms (e.g., Google Earth Engine) and pretrained deep-learning frameworks, before showing a partial-year dip in 2025 attributable to database indexing lag rather than a genuine decline in research activity.

3. Literature Review and Thematic Analysis

The 32 studies retained for synthesis were classified into six major themes based on their dominant methodological focus, summarized in Figure 3. The following subsections critically review each theme in turn.

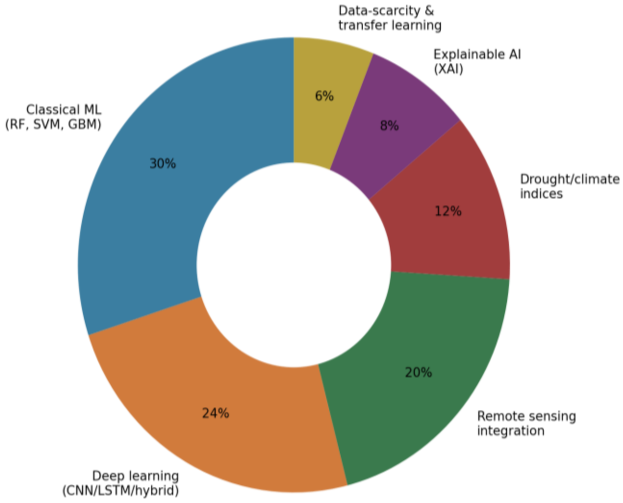

Figure 2: Thematic classification of the reviewed studies (n = 32) by dominant methodological focus.

3.1. Classical Machine Learning Approaches: Random Forest, SVR, and Gradient Boosting

Ensemble tree-based methods dominate the classical ML literature on rainfed yield prediction. A recent comparative assessment of ML baselines across eight rainfed administrative divisions in India (2000–2025) found that Random Forest achieved the lowest RMSE (0.268 t/ha) and highest R^2 (0.271) among Linear Regression, Random Forest, and Support Vector Regression under strict chronological (temporally held-out) validation, with temperature and precipitation identified as the dominant yield drivers via feature-importance analysis [1]. Notably, the authors emphasize that the modest explained variance is not a methodological failure but an honest reflection of the inherent unpredictability of rainfed systems once realistic, non-shuffled train–test splits are enforced — a methodological caution echoed throughout the broader literature but frequently overlooked in studies that report inflated performance using random k-fold splits on time-series data.

This pattern of RF robustness under data-scarce and noisy conditions recurs across geographically diverse studies. In an assessment of satellite-integrated yield prediction for Eastern Ethiopia, tree-based methods such as Random Forest and Gradient Boosting were found to offer greater robustness with limited data than deep-learning alternatives, which the authors argue are more appropriate for regions with scarce, high-quality datasets typical of smallholder farming systems [8]. A systematic review of maize yield prediction spanning 81 peer-reviewed studies (2014–2025) similarly found Random Forest to be the most frequently applied algorithm (49.4% of studies), followed by XGBoost (16.1%) and Support Vector Machines (12.4%), with hybrid deep-learning frameworks showing superior performance only where sufficiently large datasets were available [14]. A comparable pattern was reported for early-season drought-based yield prediction in the southeastern United States, where regional- and state-scale prediction accuracy reached 74% with Random Forest and 77% with k-nearest-neighbours models, with performance varying meaningfully by crop and spatial scale [4].

Support Vector Regression and Support Vector Machines appear consistently as a second-tier but reliable option, particularly when combined with variable-selection procedures. A comparative study on wheat yield prediction found that pairing SVR, Random Forest, and Extreme Gradient Boosting with structured variable-selection techniques improved predictive accuracy relative to using either the full feature set or a single algorithm alone, underscoring the value of feature engineering as a complement to algorithm choice [7]. Taken together, these findings support a broad consensus that ensemble tree-based methods represent the most operationally practical class of models for rainfed systems where labelled yield data are limited, heterogeneous in quality, and unevenly distributed across years and locations, though this comes at some cost to peak predictive accuracy relative to data-hungry deep-learning alternatives when abundant training data exist.

3.2. Deep Learning Architectures: CNN, LSTM, and Hybrid Models

Where sufficiently large and temporally continuous datasets are available — typically county-level or national datasets in the United States and China spanning 10–30 years of remote-sensing and meteorological records — deep learning architectures consistently outperform classical ML baselines. An early and influential contribution combined a Gaussian process with both LSTM and CNN architectures for county-level soybean yield prediction using MODIS imagery, outperforming official USDA yield forecasts by approximately 15% on average [28, 23]. Building on this line of work, deep LSTM architectures were subsequently applied to predict county-level corn yields directly from meteorological and remote-sensing time series [27, 22].

A subsequent methodological advance combined convolutional and recurrent architectures into unified CNN-LSTM models, on the reasoning that CNNs are better suited to extracting spatial features from imagery while LSTMs are better suited to capturing the temporal, phenological progression of crop development. A CNN-LSTM model developed for county-level soybean yield prediction across the contiguous United States, trained on Google Earth Engine-derived histogram tensors combining weather, land-surface temperature, and surface-reflectance data, outperformed both pure CNN and pure LSTM baselines for both end-of-season and in-season prediction tasks [22]. Related work applying a fully connected multi-layer perceptron to field-level corn yield estimation using soil and weather variables achieved an RMSE of 0.87 t/ha, illustrating that even comparatively simple deep architectures can achieve competitive accuracy when input features are well engineered [23].

More recent contributions have sought to explicitly quantify predictive uncertainty, moving beyond point estimates. A Bayesian neural network approach applied to remotely sensed corn-yield prediction incorporated uncertainty estimation alongside point prediction, addressing a long-standing criticism that deep-learning yield models report accuracy metrics without corresponding confidence intervals, which limits their usefulness for risk-sensitive applications such as crop insurance underwriting [23]. Separately, domain-adaptation techniques, including adversarial domain adaptation, have been proposed to address the substantial inter-year variability characteristic of rainfed corn production in the United States, framing yield prediction across years as a distribution-shift problem rather than a stationary regression problem [24]. Knowledge-guided machine learning frameworks that embed crop-physiological constraints directly into neural network architectures have also been proposed specifically for corn-yield prediction under drought stress, representing an emerging hybridization between mechanistic crop science and data-driven learning [25].

The overwhelming majority of this deep-learning evidence base originates from well-instrumented commercial agricultural regions in North America and China. Deep learning's data appetite is precisely what constrains its transferability to the smallholder rainfed contexts of Sub-Saharan Africa and South Asia, where multi-decadal, geo-referenced, quality-controlled yield records at fine spatial resolution are rare — a tension explored further in Section 3.3 and Section 5.

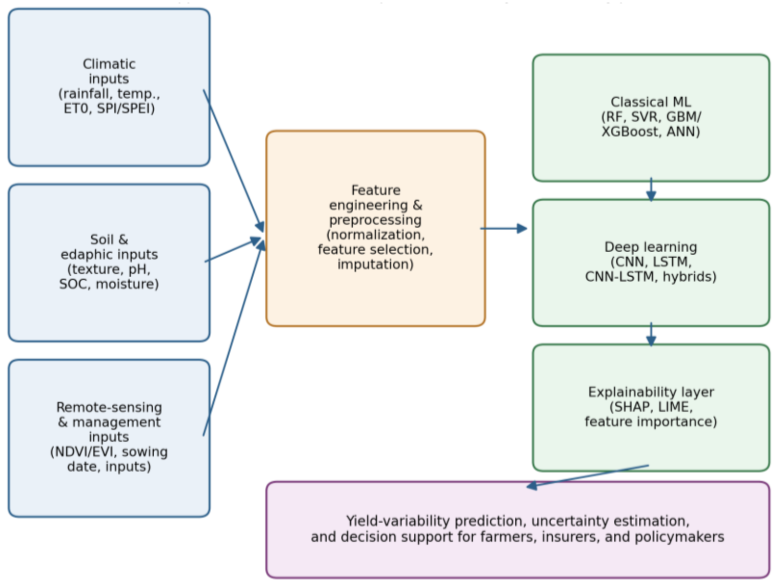

Fig 4: Conceptual framework linking data domains, modelling approaches, and decision outputs in ML-based yield-variability prediction.

Figure 4 synthesizes the modelling pipeline common to the studies reviewed in Sections 3.1–3.3: climatic, edaphic, and remote-sensing/management inputs are processed through feature engineering, routed through either classical ML, deep learning, or hybrid architectures, and increasingly passed through an explainability layer before informing decision support for farmers, insurers, and policymakers.

3.3. Remote Sensing Integration and Applicability to Smallholder Systems

The integration of satellite and UAV-derived remote sensing data has substantially expanded the feature space available to ML-based yield models, particularly in regions lacking dense networks of ground-based weather stations. A systematic review of remote-sensing-based crop mapping and yield prediction found that approximately 81% of reviewed studies relied on satellite-based sensors, with UAV platforms accounting for a further 12%, reflecting the dominance of freely available, globally consistent satellite products such as Sentinel-2 and MODIS over more logistically demanding UAV surveys [10]. Vegetation indices, particularly NDVI and the Enhanced Vegetation Index (EVI), together with land-surface temperature and soil-moisture proxies, are the remote-sensing features most consistently linked to yield outcomes across the reviewed literature [2, 10].

A pointed illustration of the transferability problem central to this review comes from a synthesis noting that a yield-prediction model trained on U.S. Corn Belt data can experience a 12–18% degradation in performance when applied directly, without local recalibration, to smallholder farming systems in East Africa [9]. This finding substantiates a broader concern raised across the reviewed literature: that models validated in data-rich, large-field, mechanized contexts cannot be assumed to generalize to the fragmented, heterogeneous, small-field conditions typical of rainfed smallholder agriculture. Consistent with this concern, a review of remote-sensing-based yield estimation methods explicitly identifies the difficulty of capturing intra-field heterogeneity in fragmented smallholder fields as a major driver of reduced model robustness and generalization capacity [11].

Encouragingly, several studies demonstrate that remote sensing can still deliver meaningful value in smallholder rainfed contexts when combined with algorithms suited to smaller sample sizes. Time-series Sentinel-2 imagery processed through an LSTM network achieved 93.77% accuracy in classifying staple crops (rice, wheat, sugarcane) as early as four weeks after sowing in smallholder farming systems, demonstrating that early-season monitoring, a precursor to yield forecasting, is achievable even in fragmented-field contexts when appropriate temporal deep-learning architectures are used [13]. A machine-learning-based crop-yield model developed specifically for Eastern Ethiopia, integrating satellite imagery with rainfall, temperature, and soil-property data, was explicitly designed to improve prediction accuracy while remaining feasible for a smallholder, data-scarce operating environment [8]. Similarly, a continental-scale effort to map maize yield across Africa at high spatial resolution combined earth-observation data with a tabular foundation model (TabPFN), representing what its authors describe as the first empirical benchmarking of foundation-model approaches for African crop-yield prediction, a promising direction for data-efficient ML given the continent's persistent ground-truth scarcity [12].

Nonetheless, infrastructure and cost barriers continue to constrain adoption. UAV-based data acquisition, while offering higher spatial resolution than most satellite products, has been estimated to cost between USD 500 and USD 2,000 per square kilometre, a cost structure that is prohibitive for smallholder farmers and limits UAV-ML integration to research contexts or well-resourced commercial operations [9]. Sensor interoperability issues further compound this barrier, restricting the practical accessibility of UAV-based precision agriculture tools in low-resource rainfed settings [9].

3.4. Drought Indices, Climatic Extremes, and Yield-Gap Analysis

A distinct but closely related thread of literature focuses less on point-yield prediction and more on characterizing the yield gap attributable to water limitation, using ML in combination with frontier and boundary-function analyses. A ten-year (2009–2018) county-level analysis of rainfed alfalfa production across 393 counties in 12 major U.S. rainfed-producing states used frontier-function analysis to estimate attainable yield and boundary-function analysis to estimate water-limited potential yield, finding a mean attainable-yield gap of 34% and a mean water-limited yield gap of 58%, with conditional inference trees identifying growing-season rainfall and minimum temperature as the principal weather-related yield-limiting variables [33]. This kind of hybrid statistical/ML yield-gap framework offers a complementary lens to the point-prediction models reviewed in Sections 3.1–3.3, shifting the analytical question from 'what will the yield be' to 'how far is the yield from its water-limited potential, and why.'

The relationship between drought indices and ML-based yield prediction remains comparatively underdeveloped relative to the volume of work using raw precipitation and temperature as inputs. Several reviewed studies incorporate standardized indices such as the Standardized Precipitation Index (SPI) or Standardized Precipitation Evapotranspiration Index (SPEI) as engineered features, on the reasoning that these indices normalize rainfall deficits relative to local climatology and are therefore more portable across regions than raw rainfall totals; however, few studies systematically compare model performance with versus without such engineered drought indices, representing a methodological gap addressed further in Section 5.

Broader climate-agriculture assessments provide important contextual grounding for this thread of the literature. FAO analysis indicates that climate change is contributing to land degradation through increased rainfall intensity, flooding, and drought frequency and severity, with rainfed systems in marginal areas bearing disproportionate exposure and comparatively low adaptive capacity [30]. Regional analysis for Asia documents that climate-driven extremes, including drought, flooding, heat stress, and cold waves, are already reducing both the quantity and quality of rice-wheat cropping-system yields that underpin regional food security [31]. In West Africa, empirical work using multivariate probit and propensity-score-matching methods has linked erratic rainfall and temperature fluctuations under heavy rainfed dependence to declining yields, rising production costs, and greater income volatility among smallholders adopting (or failing to adopt) climate-adaptation strategies [32]. While this last study is not itself an ML-prediction study, it illustrates the socioeconomic stakes that make accurate ML-based yield-variability forecasting practically consequential rather than a purely academic exercise.

3.5. Explainable AI (XAI) and the Interpretability–Accuracy Trade-off

A recurring critique of both classical ensemble and deep-learning yield models is their limited interpretability, which constrains trust and adoption among farmers, agronomists, and policymakers who must act on model outputs. This concern has catalyzed a rapidly growing sub-literature applying Explainable AI (XAI) techniques, particularly SHAP (SHapley Additive exPlanations) and LIME (Local Interpretable Model-agnostic Explanations), to yield-prediction models. A hybrid model combining Random Forest, LSTM, and XGBoost with SHAP, LIME, and counterfactual analysis, trained on a large multi-year Indian agricultural dataset spanning over 246,000 records across 33 states, achieved high predictive accuracy ($R^2 = 0.9827$ for aggregate crop yield and 0.9721 for rice yield specifically) while explicitly aiming to make the resulting predictions transparent enough for integration into a farmer-facing decision-support interface [15]. Comparable work applying Decision Tree, Random Forest, and LightGBM regressors alongside SHAP and LIME reported R^2 values as high as 0.92 for climate-influenced yield prediction, explicitly framing XAI integration as necessary infrastructure for translating accurate predictions into actionable guidance on climatic stressors [18].

Beyond model-agnostic post hoc explanation methods, some studies have explored inherently interpretable architectures as an alternative to the 'explain the black box' paradigm. A comparison of a Kolmogorov-Arnold Network (KAN) against multilayer perceptron and Random Forest baselines for soybean-yield forecasting across major producing countries found that while KAN performed comparably to MLP and RF in predictive accuracy despite a small sample size, it substantially outperformed both in interpretability, a property the authors argue is essential for models intended for real-world agricultural decision-making [19]. This reflects a broader methodological debate in the interpretable-ML literature, captured in the influential argument that reliance on post hoc explanation of inherently opaque models can itself be a source of risk in high-stakes domains, and that inherently interpretable models should be preferred wherever they achieve comparable accuracy [21].

A critical perspective raised within this sub-literature warns against uncritical adoption of black-box models for climate-impact evaluation specifically, arguing that opaque models risk obscuring spurious correlations between climatic covariates and yield outcomes that would not survive mechanistic scrutiny, a risk with direct policy relevance given that yield-prediction outputs increasingly inform insurance underwriting and adaptation planning [16]. A broader review of ensemble learning and XAI techniques for evaluating crop yields under abnormal climate conditions similarly emphasizes that ensemble methods (bagging, boosting, and stacking) improve robustness by compensating for individual model weaknesses, but that this robustness gain must be paired with XAI-based transparency to be operationally trustworthy, particularly in high-dimensional, climate-stressed prediction contexts [20]. Collectively, this theme indicates a maturing recognition within the field that predictive accuracy alone is an insufficient criterion for evaluating ML yield models intended for real-world deployment; interpretability, trust, and actionability are increasingly treated as co-equal evaluation criteria.

3.6. Data Scarcity, Transfer Learning, and Foundation Models

A cross-cutting theme spanning virtually all of the preceding sections is the constraint imposed by data scarcity in the very regions where rainfed agriculture is most consequential for food security. The systematic review of maize-yield ML studies found that Africa contributed less than 25% of the reviewed evidence base despite maize being a staple crop central to numerous African economies and livelihoods, a geographic imbalance the authors attribute to persistent challenges of data scarcity, limited interpretability of complex models, and uneven research capacity across regions [14]. This imbalance is consequential: models, features, and hyperparameters optimized on data-rich U.S. or Chinese datasets are not automatically valid elsewhere, and the 12–18% cross-regional performance degradation documented for Corn Belt-to-East-Africa model transfer is a direct symptom of this problem [9].

In response, several methodological strategies are emerging to reduce dependence on large labelled datasets. Transfer learning, in which a model pretrained on a data-rich region or crop is fine-tuned using a smaller labelled dataset from the target region, has been explored specifically for smallholder soybean-yield prediction across distinct U.S. ecoregions, demonstrating that even modest amounts of target-region data can meaningfully improve transferred-model performance relative to naive direct application [24]. Tabular foundation models, pretrained on large heterogeneous tabular datasets and capable of few-shot adaptation to new prediction tasks, have recently been benchmarked for continental-scale African maize-yield mapping, representing an early but promising application of the broader foundation-model paradigm to a domain long constrained by ground-truth scarcity [12].

Existing continental and global yield datasets that partially substitute for dense ground-truth records — including M3Crops, GAEZ, SPAM, RAY2012, and the more recent GDHY hybrid dataset — remain constrained by coarse spatial resolution, discontinuous temporal coverage, and dependence on model-based assumptions rather than direct measurement, limiting their capacity to fully resolve the fine-grained spatial yield variability most relevant to smallholder decision-making [12]. Federated learning, in which models are trained collaboratively across distributed local datasets without centralizing raw farmer-level data, is beginning to be discussed as a mechanism to expand the effective training-data pool available for rainfed-system models while addressing legitimate data-privacy and data-sovereignty concerns among smallholder farmers and national agricultural agencies, though empirical applications specific to yield-variability prediction remain nascent within the reviewed evidence base and are flagged as a priority direction in Section 5.

4. Comparative Analysis of Previous Studies

Table 3 summarizes representative studies drawn from the thematic synthesis, compares major model classes across key evaluation dimensions relevant to rainfed, smallholder-relevant deployment.

Table 3: Comparative analysis of model classes across key evaluation dimensions

Model class	Typical data requirement	Reported accuracy range	Interpretability	Suitability for smallholder rainfed systems
Linear/multiple regression	Low	Low–moderate (often lowest R^2 among baselines)	High	Moderate (simple, but misses nonlinearity)
Random Forest / Gradient Boosting	Low–moderate	Moderate–high; most robust under scarce/noisy data	Moderate (feature importance available)	High
Support Vector Regression / SVM	Low–moderate	Moderate; improved with feature/variable selection	Low–moderate	Moderate–high
Artificial Neural Network (MLP)	Moderate	Moderate–high with adequate features	Low	Moderate
CNN / LSTM / CNN-LSTM hybrids	High (large, temporally rich datasets)	High, often best-in-class where data-rich	Low	Low–moderate (data-hungry)
XAI-augmented hybrid models (RF/LSTM/XGBoost + SHAP/LIME)	Moderate–high	High, with added transparency	High (post hoc)	Moderate–high
Foundation / tabular pretrained models (e.g., TabPFN)	Low at fine-tuning stage (leverages pretraining)	Emerging; promising in early benchmarks	Low–moderate	High potential (data-efficient)

5. Research Gaps and Emerging Trends

The thematic synthesis in Section 3, summarized comparatively in Tables 4 through 6, points toward several unresolved issues that constrain the practical value of ML-based yield-variability prediction specifically for rainfed systems. First, the geographic imbalance in the evidence base — with Sub-Saharan Africa contributing less than a quarter of the reviewed maize-yield literature despite the centrality of rainfed maize to regional food security — means that the methodological lessons drawn from data-rich contexts cannot be assumed to transfer without loss of accuracy, as directly demonstrated by the documented 12–18% performance degradation when Corn Belt-trained models are applied to East African smallholder systems [9, 14].

Second, methodological limitations in previous studies frequently stem from validation design rather than algorithm choice. Several widely cited high-accuracy results are obtained using random k-fold cross-validation on time-series agricultural data, which can leak future information into training folds and inflate apparent accuracy; the comparatively modest R^2 (0.271) reported under strict chronological validation for rainfed Indian cereals is instructive precisely because it reflects a more honest, operationally realistic estimate of model skill under genuine forecasting conditions [1]. Reviews that aggregate accuracy claims across studies without scrutinizing validation methodology risk overstating the field's genuine predictive progress.

Third, emerging technologies such as foundation models, knowledge-guided hybrid architectures, and federated learning remain in early stages of application to rainfed yield-variability prediction specifically, even though they are conceptually well suited to the field's core constraints of data scarcity and cross-regional heterogeneity. The single identified benchmarking of a tabular foundation model for African maize-yield mapping represents a promising but isolated data point rather than an established methodological consensus [12].

Fourth, areas requiring further investigation include the systematic comparison of engineered drought indices (SPI, SPEI, soil-moisture anomalies) against raw climatic variables as model inputs; the development of standardized, openly available benchmark datasets specifically for rainfed smallholder systems analogous to those available for the U.S. Corn Belt; and the design of evaluation frameworks that explicitly weight extreme-event (tail-risk) prediction accuracy alongside average-case RMSE, given that it is precisely the extreme, low-probability yield failures that carry the greatest consequences for household food security and agricultural insurance solvency.

6. Discussion

The literature synthesized in this review reveals a field that has matured considerably in methodological sophistication while remaining geographically and infrastructurally uneven in its practical reach. Three cross-cutting patterns merit integrated discussion.

The first is a consistent, and largely reconcilable, apparent tension between the reported superiority of deep-learning architectures and the practical dominance of ensemble tree-based methods in the rainfed-specific literature. This is not a genuine contradiction but a reflection of differing data regimes: deep-learning superiority is demonstrated primarily in data-rich, multi-decadal, remote-sensing-dense contexts such as the U.S. Corn Belt [22, 23, 27, 28]], whereas Random Forest and Gradient Boosting/XGBoost robustness is demonstrated primarily in data-scarce, high-variance rainfed contexts typical of India, Ethiopia, and other smallholder-dominated regions [1, 8, 14]. Read together, these findings suggest that model selection in this field should be treated as a function of data regime rather than as a search for a single universally superior algorithm — a conclusion with direct implications for practitioners deciding which approach to deploy in a given operational context, discussed further below.

The second cross-cutting pattern concerns the growing but still incomplete integration of interpretability into model evaluation. Early literature in this domain, consistent with the broader trajectory of applied ML research, prioritized predictive accuracy metrics (RMSE, R^2, MAE) almost exclusively [2,3]. More recent work increasingly treats interpretability as a co-equal criterion, reflecting the practical reality that farmers, agricultural extension officers, and policymakers cannot act on a prediction they do not trust or understand [15, 18, 19, 21]. This shift parallels developments in other high-stakes ML application domains (e.g., clinical decision support) and appears likely to continue, particularly as ML-based yield predictions are increasingly proposed as inputs to index-based crop insurance products, where transparency has direct financial and legal implications for both insurers and policyholders.

The third pattern concerns the persistent gap between methodological sophistication and operational deployment, particularly in the smallholder rainfed contexts where accurate yield-variability prediction arguably carries the greatest humanitarian stakes. The evidence reviewed here documents substantial performance degradation when models trained in data-rich contexts are transferred without adaptation to smallholder settings [9], compounded by genuine infrastructure barriers including UAV and sensor costs that are prohibitive at smallholder scale [9] and structural underrepresentation of African and South Asian smallholder contexts in the published evidence base [14]. This suggests that the central bottleneck constraining the field's practical impact is not primarily algorithmic — the classical ML and deep-learning toolkits reviewed here are, on balance, mature and well validated — but rather infrastructural and data-related: the availability of ground-truth yield records, affordable remote-sensing access, and locally calibrated training data in precisely the regions where rainfed yield variability is most consequential.

These patterns collectively suggest that future progress in this field will depend less on incremental algorithmic refinement and more on structural investments: expanding ground-truth data collection infrastructure in underrepresented regions, developing standardized rainfed-specific benchmark datasets, and designing hybrid modelling frameworks (knowledge-guided ML, transfer learning, foundation models) explicitly engineered for the low-data, high-variance conditions that characterize the majority of the world's rainfed cropland. The theoretical implication is that yield-variability prediction in rainfed agriculture should be conceptualized less as a conventional supervised-learning problem to be solved through ever-larger models, and more as a data-scarce, distribution-shift problem requiring methodological strategies purpose-built for that regime.

7. Conclusion

This review has synthesized 32 peer-reviewed studies published between 2015 and 2026 on the application of machine learning and deep learning for predicting yield variability in rainfed agricultural systems. The evidence indicates that ensemble tree-based methods, particularly Random Forest and Gradient Boosting/XGBoost, currently represent the most operationally robust model class for the data-scarce, high-variability conditions typical of smallholder rainfed agriculture, while deep-learning architectures, including CNN, LSTM, and CNN-LSTM hybrids, achieve superior accuracy in data-rich contexts where multi-decadal remote-sensing and meteorological records are available. Rainfall, temperature, soil moisture, and vegetation indices consistently emerge as the dominant predictors across studies, and the field is increasingly integrating explainable AI techniques, uncertainty quantification, and emerging paradigms such as knowledge-guided hybrid models, transfer learning, and tabular foundation models to address longstanding limitations in interpretability, generalizability, and data efficiency.

The overall significance of this synthesis lies in clarifying that the central constraint on this field's practical impact is not principally a shortage of capable algorithms, but rather a shortage of representative training data, affordable remote-sensing access, and locally validated models in the smallholder rainfed regions where accurate yield-variability prediction is most consequential for food security. This reframing has direct implications for how research effort and investment should be allocated going forward.

For researchers, this review recommends: (i) adopting strict chronological or spatially held-out validation protocols rather than random k-fold splits when evaluating models on time-series agricultural data, to avoid inflated and operationally misleading accuracy claims; (ii) prioritizing the development and public release of standardized, quality-controlled rainfed-system benchmark datasets, particularly for Sub-Saharan Africa and South Asia; (iii) systematically comparing engineered drought indices against raw climatic variables as model inputs; and (iv) explicitly evaluating model performance on extreme, tail-risk yield outcomes rather than average-case error metrics alone.

For practitioners and policymakers, this review recommends: (i) treating model selection as contingent on locally available data volume and quality rather than defaulting to the most sophisticated available architecture; (ii) prioritizing investment in interpretable or XAI-augmented models where predictions will directly inform farmer decision-making, insurance underwriting, or public policy; and (iii) supporting infrastructure investments — including affordable satellite data access, ground-station networks, and data-sharing arrangements such as federated learning — that can narrow the demonstrated performance gap between data-rich and data-scarce rainfed regions.

Future research should prioritize hybrid, knowledge-guided modelling frameworks that combine the mechanistic grounding of process-based crop models with the pattern-recognition strength of machine learning; expanded application of foundation models and transfer learning specifically calibrated for smallholder rainfed contexts; privacy-preserving federated learning architectures that expand usable training data without compromising farmer data sovereignty; and participatory validation studies conducted directly with farming communities to ensure that model outputs are not only statistically accurate but also actionable, trusted, and appropriately communicated within local decision-making contexts.

References

1. Author(s). Forecasting crop yields in rainfed India: a comparative assessment of machine learning baselines and implications for precision agribusiness. Agriculture. 2026;16(1):65.
2. Author(s). Crop yield prediction in agriculture: a comprehensive review of machine learning and deep learning approaches, with insights for future research and sustainability. Heliyon (ScienceDirect). 2024.
3. Author(s). Crop yield prediction using machine learning: an extensive and systematic literature review. Smart Agric Technol (ScienceDirect). 2024.
4. Author(s). A framework for early-season crop yield prediction using machine learning and drought conditions in the southeastern United States. Smart Agric Technol (ScienceDirect). 2026.
5. Author(s). A comparative study of machine learning models in predicting crop yield. Discov Agric. 2025.
6. Author(s). Machine learning approaches for crop yield prediction: a review. In: Proceedings of the IEEE Conference. IEEE; 2023.
7. Sánchez JCM, Mesa HGA, Espinosa AT, Castilla SR, Lamont FG. Improving wheat yield prediction through variable selection using support vector regression, random forest, and extreme gradient boosting. Smart Agric Technol. 2025;10:100791.
8. Author(s). Integration of satellite data for predicting crop yields in Eastern Ethiopia using machine learning. Sci Rep. 2025.
9. Author(s). Integrating UAVs, satellite remote sensing, and machine learning in precision agriculture: pathways to sustainable food production, resource efficiency, and scalable innovation. Front Agron. 2025.
10. Author(s). Remote-sensing data and deep-learning techniques in crop mapping and yield prediction: a systematic review. Remote Sens. 2023;15(8):2014.
11. Author(s). A review of remote sensing-based crop yield estimation: machine learning techniques and environmental, algorithmic, and hardware limitations. Front Plant Sci. 2026.
12. Author(s). High-resolution maize yield mapping across Africa using earth observation and machine learning, deep learning, and foundation model. Comput Electron Agric (ScienceDirect). 2025.
13. Khan HR, Gillani Z, Jamal MH, Athar A, Chaudhry MT, Chao H, *et al*. Early identification of crop type for smallholder farming systems using deep learning on time-series Sentinel-2 imagery. Sensors. 2023;23(4):1779.
14. Author(s). Maize yield prediction using machine learning: a systematic literature review. Decis Anal J. 2026.
15. Author(s). An explainable AI-based hybrid machine learning model for interpretability and enhanced crop yield prediction. Artif Intell Agric (ScienceDirect). 2025.
16. Author(s). Crop yield prediction via explainable AI and interpretable machine learning: dangers of black box models for evaluating climate change impacts on crop yield. Preprint/Review. 2023.
17. Author(s). AgriYieldAI: an AI-driven decision support system for crop yield prediction using YieldFusionNet. Discov Comput. 2026.
18. Author(s). Next-gen agriculture: integrating AI and XAI for precision crop yield predictions. Front Plant Sci. 2024/2025.
19. Author(s). From data to decisions: the use of explainable AI to forecast soybean yield in major producing countries. Sci Rep. 2026.
20. Author(s). Recent trends in machine learning, deep learning, ensemble learning, and explainable artificial intelligence techniques for evaluating crop yields under abnormal climate conditions. Front Plant Sci (PMC). 2025.
21. Author(s). Enhancing crop yield prediction through explainable AI for interpretable insights. In: Springer Conference Proceedings. Springer; 2024.

22. Sun J, Di L, Sun Z, Shen Y, Lai Z. County-level soybean yield prediction using deep CNN-LSTM model. Sensors. 2019;19(20):4363.
23. Ma Y, Zhang Z, Kang Y, Özdoğan M. Corn yield prediction and uncertainty analysis based on remotely sensed variables using a Bayesian neural network approach. Remote Sens Environ. 2021;259:112408.
24. Author(s). An adaptive adversarial domain adaptation approach for corn yield prediction. Report. US Department of Energy Office of Scientific and Technical Information; 2023.
25. Author(s). Knowledge-guided machine learning for county-level corn yield prediction under drought. arXiv preprint arXiv:2503.16328. 2025.
26. Crane-Droesch A. Machine learning methods for crop yield prediction and climate change impact assessment in agriculture. Environ Res Lett. 2018;13(11):114003.
27. Jiang Z, Liu C, Hendricks NP, Ganapathysubramanian B, Hayes DJ, Sarkar S. Predicting county level corn yields using deep long short term memory models. arXiv preprint arXiv:1805.12044. 2018.
28. You J, Li X, Low M, Lobell D, Ermon S. Deep Gaussian process for crop yield prediction based on remote sensing data. In: Proceedings of the AAAI Conference on Artificial Intelligence. 2017;31(1).
29. Environmental and Energy Study Institute (EESI). FAO report describes relationship between climate change and food security [Internet]. Washington (DC): EESI; 2019.
30. Food and Agriculture Organization of the United Nations (FAO). The state of the world's land and water resources for food and agriculture 2025: climate change – an existential threat to agrifood systems. Rome: FAO; 2025.
31. Author(s). Impact of climate change on agricultural production; issues, challenges, and opportunities in Asia. Front Plant Sci. 2022;13:925548.
32. Author(s). Assessing the impact of climate change adaptation strategies on food security and farm income: insights from Burkina Faso. Agric Food Secur. 2026.
33. Author(s). Yield gap analysis of rainfed alfalfa in the United States. PMC. 2022.
34. Breiman L. Statistical modeling: the two cultures. Stat Sci. 2001;16(3):199–231.