

Hybrid Residual SARIMAX-XGBoost for Provincial Rice Production Forecasting in Indonesia: A Systematic Literature Review and Conceptual Framework for Food Policy Decision Support

Imannudin Akbar^{1*}, Tonny Hidayat², Alva Hendi Muhammad³, Emha Taufiq Luthfi⁴

Abstract—This study contributes a systematic literature review and conceptual framework for provincial rice production forecasting in Indonesia by integrating residual-based hybrid modelling, explainability, and food policy decision support. A total of 60 articles published between 2023 and 2026 were selected from 124 initial records identified from six academic search sources: ScienceDirect/Elsevier, IEEE Xplore, SpringerLink, MDPI journals, Scopus/Crossref tracing, and Google Scholar/manual tracking, using a PRISMA-style screening process. The review synthesizes evidence across rice production forecasting, crop-yield forecasting, hybrid machine learning, explainable artificial intelligence, and decision-support-oriented forecasting. The literature findings show that recent studies increasingly use agricultural statistics, climate variables, remote sensing, UAV imagery, crop-model outputs, machine learning, and deep learning. However, the reviewed studies rarely integrate province-aware determinants, residual-based model decomposition, time-aware validation, residual-level explainability, and DSS-ready policy outputs within a single framework. Based on this gap, the study proposes a conceptual Hybrid Residual SARIMAX-XGBoost framework. SARIMAX is positioned to model temporal, seasonal, autoregressive, and exogenous baseline patterns, while XGBoost learns nonlinear residual structures. SHAP is positioned as an explanation layer for interpreting forecast adjustments rather than as causal evidence. The proposed framework remains conceptual and requires future empirical validation using Indonesian provincial rice production data. This review provides theoretical contributions to hybrid forecasting literature and practical implications for information systems, decision support, and public-sector food policy planning.

Index Terms—Systematic literature review, rice production forecasting, SARIMAX, XGBoost, residual learning, SHAP, food policy decision support system.

I. INTRODUCTION

Rice is a strategic staple in Indonesia because its production affects household food security, government procurement, market stability, and inflation control. At the provincial level, production emerges from interactions among harvested area, productivity, historical output, seasonal cycles, and climate variability. Rice production forecasting should therefore be treated as a multivariate time-series problem rather than a simple extrapolation of past output [1]–[2].

Provincial differences further complicate forecasting. Provinces vary in agro-climatic conditions, irrigation infrastructure, cropping intensity, land-use dynamics, and exposure to climate anomalies. Evidence from rice-yield studies shows that spatial heterogeneity can influence model accuracy and variable importance across field, district, county, and regional scales [3]–[8].

Recent studies combine agricultural statistics, weather data, climate reanalysis, satellite imagery, UAV imagery, crop-growth information, and machine learning algorithms [9]–[24]. These multi-source approaches improve information coverage, but they do not automatically solve the modelling problem. A forecasting model must still distinguish predictable linear-seasonal behaviour from nonlinear deviations produced by climate, production shocks, or regional interactions.

Previous reviews on crop-yield forecasting have mainly emphasized general machine learning models, remote-sensing-based yield estimation, in-season forecasting, or broad methodological trends. These reviews are valuable for mapping the development of agricultural forecasting, but they have not specifically examined how linear-seasonal time-series structures, nonlinear residual patterns, explainable residual correction, and DSS-oriented policy outputs can be integrated for provincial rice production forecasting. This creates an important gap for Indonesia, where rice production is not only a technical forecasting problem but also a public-sector decision-support issue related to food security, market stabilization, procurement planning, and regional policy intervention.

Statistical time-series models, including ARIMA, SARIMA, and SARIMAX, remain relevant because they explicitly model temporal dependence, trend, seasonality, and exogenous variables [25]. In this context, SARIMAX can incorporate harvested area, productivity, and climate variables as exogenous predictors. Its limitation is that it assumes relatively stable linear relationships after differencing and seasonal

Received: 21 May 2026; Revised: 17 June 2026; Accepted: 5 July 2026.

*Corresponding author

¹Imannudin Akbar, Department of Information Systems, Faculty of Information Technology, Universitas Informatika dan Bisnis Indonesia, Bandung, Indonesia (e-mail: imannudin@unibi.ac.id).

²Tonny Hidayat, Department of Informatics, Faculty of Computer Science, Universitas Amikom Yogyakarta, Yogyakarta, Indonesia (e-mail: tonny@amikom.ac.id).

³Alva Hendi Muhammad, Department of Informatics, Faculty of Computer Science, Universitas Amikom Yogyakarta, Yogyakarta, Indonesia (e-mail: alva@amikom.ac.id).

⁴Emha Taufiq Luthfi, Department of Informatics, Faculty of Computer Science, Universitas Amikom Yogyakarta, Yogyakarta, Indonesia (e-mail: emhataufiqluthfi@amikom.ac.id).

specification. When production is affected by nonlinear climate response, abrupt land-use change, irrigation differences, policy intervention, or extreme-weather anomalies, SARIMAX residuals may still contain systematic information.

Ensemble machine learning models such as XGBoost are widely used because they capture nonlinear relationships and interaction effects [17], [26]–[29]. However, standalone XGBoost is not inherently time-series aware. Without lag design, seasonal features, and chronological validation, it may learn associations that do not generalize across seasons or provinces. This makes integration more useful than simple model comparison: SARIMAX can model the structured temporal baseline, while XGBoost can focus on nonlinear residual correction.

SARIMAX-XGBoost is selected because the two models address complementary forecasting limitations. SARIMAX is suitable for representing temporal dependence, seasonality, autoregressive patterns, and exogenous production-related variables, but it may not fully capture nonlinear climate effects, province-specific deviations, or abrupt production shocks. XGBoost, in contrast, is effective for modelling nonlinear interactions in tabular and multi-source agricultural data, but it requires explicit temporal feature engineering and careful validation. Therefore, a residual-based hybrid design is more appropriate than a simple standalone model comparison because SARIMAX can first isolate the interpretable temporal-exogenous baseline, while XGBoost can focus on the remaining nonlinear residual structures.

The hybrid residual SARIMAX–XGBoost model is therefore conceptualized as a two-stage architecture. SARIMAX first captures the linear, seasonal, autoregressive, and exogenous structure of rice production. XGBoost then learns the residuals left by SARIMAX, treating them as potential signals of nonlinear climate effects, production shocks, regional heterogeneity, and predictor interactions. SHAP is used to interpret this residual-correction layer and identify variables that drive forecast adjustment [30]–[32].

The framework is proposed as a conceptual synthesis derived from the reviewed literature, not as an empirically validated superior model at this stage. Its empirical value must be examined in future work using Indonesian provincial rice production data, time-aware validation, baseline comparisons, and province-level error analysis.

This paper therefore has two connected purposes. First, it synthesizes recent evidence on contextual determinants, data characteristics, forecasting approaches, evaluation strategies, explainability, and DSS relevance in rice and crop production forecasting. Second, it translates the identified literature gap into a conceptual Hybrid Residual SARIMAX-XGBoost framework for food policy decision support. The updated research questions and their related objectives are presented in Table 1. The explicit gap identified from the 60 reviewed studies is the absence of an integrated SLR framework that jointly links province-aware contextual determinants, residual-based hybrid modelling, time-aware evaluation, residual-level explainability, and DSS-ready food policy outputs for provincial rice production forecasting.

Code	Updated Research Question	Related Research Objectives And Analytical Focus
RQ1	What temporal, provincial, production-related, and climate-related factors influence regional or provincial rice production forecasting?	RO1–RO2; contextual determinants and data characteristics.
RQ2	How does recent literature support an interpretable residual-based hybrid forecasting framework that integrates residual separation, model evaluation, explainability, and food policy DSS relevance?	RO3–RO7; hybrid residual modelling, methodological limitations and gaps, evaluation, explainability, and DSS relevance.

The two research questions are operationalized through seven research objectives (ROs):

- RO1. To identify temporal, provincial, production-related, and climate-related determinants relevant to regional or provincial rice production forecasting.
- RO2. To synthesize the data characteristics and predictor sources used in recent rice and crop-yield forecasting studies, including agricultural statistics, climate variables, remote sensing, UAV imagery, and crop-model outputs.
- RO3. To classify and map recent forecasting approaches, including statistical models, machine learning, deep learning, hybrid models, and explainable AI-based approaches.
- RO4. To analyze the methodological limitations and gaps of SARIMAX, XGBoost, SHAP, and existing hybrid forecasting approaches in the context of rice or crop production forecasting.
- RO5. To formulate the evidence-based logic of an interpretable Hybrid Residual SARIMAX-XGBoost framework for separating temporal-exogenous baseline patterns from nonlinear residual structures.
- RO6. To define an evaluation orientation that includes predictive accuracy, baseline comparison, time-aware validation, province-level performance, and forecasting-horizon analysis.
- RO7. To synthesize how residual-based explanations and DSS-oriented outputs can support food policy analysis through risk signals, provincial rankings, influential-variable summaries, and scenario insights.

II. RELATED WORK

A. Rice Production and Rice Yield Forecasting

Rice-specific forecasting studies provide the empirical core of this SLR. Recent works examine Indonesian rice production using advanced machine learning [1], district-level rice yield with climate reanalysis and remote sensing in India [2], field- and county-level rice yield using SAR, optical, and meteorological data [3], and Australian rice yield forecasting with satellite and weather variables [4]. Additional studies integrate crop-growth models, UAV imagery, PlanetScope data, hyperspectral imagery, mobile/edge deep learning, vegetation indices, and multi-season satellite sequences [5]–[24], indicating that rice production is shaped by temporal dynamics,

spatial scale, environmental variables, data resolution, and sensor availability.

B. Crop Yield Forecasting Beyond Rice

Studies beyond rice add methodological evidence. Reviews and mapping papers show that crop-yield prediction increasingly uses machine learning, deep learning, Earth observation, temporal modelling, field-scale forecasting, data assimilation, and policy-oriented decision support [33]–[40]. These findings are relevant because wheat, maize, potato, alfalfa, cotton, and rice forecasting share challenges related to temporal transferability, feature selection, explainability, and multi-source data integration.

Although the target context of this SLR is provincial rice production, rice-yield forecasting studies were included because production is closely linked to harvested area and productivity or yield. Yield studies provide transferable evidence on temporal features, climate sensitivity, remote-sensing predictors, nonlinear modelling, and explainability. Therefore, they are used to support methodological synthesis, while the proposed empirical direction remains focused on provincial rice production.

C. Statistical and Time-Series Modelling

SARIMAX extends seasonal time-series modelling by allowing exogenous variables, which is important because harvested area, productivity, rainfall, temperature, humidity, and climate indices may influence rice production. In the proposed framework, SARIMAX is used as the first-stage model that extracts interpretable temporal, seasonal, and exogenous structure. Its weakness is the limited capacity to represent nonlinear interactions, threshold effects, and changing production responses across provinces; its value lies in producing both a base forecast and a diagnostically meaningful residual series.

D. Machine Learning, XGBoost, and Nonlinear Forecasting

Machine learning models are widely used because agricultural responses are often nonlinear. XGBoost is relevant because gradient-boosted decision trees and related ensemble methods can capture complex feature interactions in tabular, remote-sensing, weather, soil, and agronomic crop-yield forecasting data [17], [26]–[29], [31]. Its limitations are black-box behaviour, overfitting risk in small or biased datasets, and weak temporal awareness unless lags, seasonal variables, or rolling features are explicitly constructed.

E. Hybrid, Process-Aware, and Knowledge-Informed Approaches

The literature shows a movement from single-model forecasting toward hybrid, process-aware, and advanced machine learning approaches. Existing studies combine crop models with machine learning, data assimilation with forecasting algorithms, transfer learning with Sentinel data, knowledge-informed fusion networks, neural architectures, federated learning, and UAV-based prediction pipelines [26]–[28], [30], [41]–[59]. However, many hybrids combine

outputs without clearly separating component roles. A residual-based SARIMAX-XGBoost architecture provides a sharper division: SARIMAX estimates linear-seasonal-exogenous behaviour, and XGBoost models nonlinear residuals.

F. Explainability and Decision Support Systems

Food policy analysis requires more than a forecast value. Decision makers need to understand why production may change, which provinces are exposed to risk, and which variables drive forecast adjustment. SHAP is useful for interpreting variable contributions in agricultural prediction [30]–[32], but it explains model behaviour rather than causality and may be unstable when predictors are highly correlated. In this SLR, SHAP is positioned as an interpretation layer for residual correction, not as standalone causal evidence. Table 2 summarizes the literature streams, key references, and their specific contributions to this SLR.

Table 2.
Literature Streams and Citation Coverage

Literature Stream	Key References	Contribution
Rice production/yield forecasting	[1]–[24]	Supports RQ1 and RO1–RO2 by identifying contextual determinants, data sources, spatial scales, and predictor relationships.
Review and methodological mapping	[33]–[40]	Supports RO3 by positioning the review within broader trends in crop-yield forecasting, remote-sensing-based forecasting, and methodological mapping.
Hybrid and advanced ML/DL forecasting	[26]–[28], [30], [41]–[59]	Supports RQ2 and RO3–RO5 by informing hybrid statistical–machine learning, residual separation, data fusion, and methodological limitation analysis.
Explainability, climate, and RS support	[30]–[32]	Supports RQ2 and RO6–RO7 by informing residual explanation, climate interpretation, evaluation design, and DSS relevance.
Methodological foundations	[25], [29], [32], [60]–[61]	Supports PRISMA reporting, SLR procedure, XGBoost, SHAP, and time-series forecasting principles.

Across these streams, the literature provides strong evidence for data-rich and AI-based agricultural forecasting, but the streams remain only partially integrated. Rice-specific studies clarify contextual determinants, statistical models offer interpretable temporal baselines, machine learning models support nonlinear prediction, hybrid approaches connect domain structure with data-driven learning, and XAI/DSS studies support interpretation and policy use. However, most studies do not clearly define how each component should function within a residual-based forecasting architecture. Table

3 summarizes this critical comparison and explains how each literature stream supports the proposed conceptual framework.

Table 3.

Critical Comparison of Literature Streams and Support for The Proposed Framework

Literature stream	Critical synthesis and framework relevance
Rice production/yield forecasting	Identifies temporal, production, climate, and remote-sensing determinants, but many studies are not provincial-DSS oriented; therefore, the framework maps these determinants into province-aware variables.
Statistical time-series/SARIMAX	Interpretable for temporal-exogenous baselines but limited for nonlinear residual patterns; therefore, SARIMAX is used as the first-stage baseline.
Machine learning/XGBoost	Captures nonlinear interactions but is not automatically time-aware; therefore, XGBoost is restricted to residual correction with engineered temporal and provincial features.
Hybrid/process-aware forecasting	Shows the value of combining domain structure and machine learning, but component roles are often unclear; therefore, residual decomposition defines each component function.
XAI/DSS-oriented studies	Supports interpretation and policy outputs but remains sparse; therefore, SHAP is positioned for residual explanation and DSS-ready outputs.

III. RESEARCH METHOD

A. Review Protocol

This study follows an SLR design informed by PRISMA 2020 [60] and software-engineering-oriented review guidance [61]. The protocol was established before synthesis to ensure consistent identification, screening, quality assessment, extraction, and interpretation. The corpus was selected to address two main research questions and seven operational objectives covering provincial rice-production dynamics, data characteristics, hybrid residual model design, methodological limitations, evaluation, explainability, and DSS relevance.

B. Search strategy

The search strategy combined keywords related to rice production forecasting, rice yield forecasting, crop-yield forecasting, SARIMA, SARIMAX, XGBoost, machine learning, deep learning, hybrid forecasting, residual learning, climate variables, remote sensing, SHAP, and decision support system (DSS). Searches were conducted across publisher databases, indexing sources, and article-discovery platforms relevant to agriculture, remote sensing, environmental modelling, and information systems. To improve reproducibility, the core Boolean structure combined three search blocks: (1) prediction object, including "rice production", "rice yield", and "crop yield"; (2) modelling approach, including "machine learning", "deep learning", "hybrid model", "SARIMAX", "XGBoost", and "residual learning"; and (3) DSS and explanation terms, including "SHAP", "explainable AI", and "decision support system".

The review period of 2023–2026 was selected to capture the most recent development of AI-based crop forecasting, particularly the rapid growth of hybrid machine learning, remote-sensing-based forecasting, explainable AI, and DSS-oriented agricultural prediction. Earlier foundational works were not used as part of the main review corpus but were retained as methodological foundations where necessary,

including PRISMA, XGBoost, SHAP, and time-series forecasting principles.

The identification stage recorded the following source distribution and query focus to improve reproducibility:

Table 4 presents the source distribution and detailed search strings used during the identification stage.

Table 4.

Sources and Query Focus

Source	Records	Detailed search string / query focus
ScienceDirect / Elsevier	47	("rice production" OR "rice yield" OR "crop yield") AND ("forecasting" OR "prediction") AND ("machine learning" OR "deep learning" OR "hybrid model" OR "remote sensing")
IEEE Xplore	20	("rice production prediction" OR "crop-yield prediction") AND ("machine learning" OR "XGBoost" OR "explainable AI" OR "decision support system")
SpringerLink	17	("crop-yield forecasting" OR "rice yield prediction") AND ("machine learning" OR "hybrid forecasting" OR "climate variables" OR "remote sensing")
MDPI journals	15	("rice yield estimation" OR "crop-yield prediction") AND ("remote sensing" OR "satellite imagery" OR "climate data" OR "SHAP")
Scopus / Crossref tracing	10	("SARIMAX" OR "SARIMA" OR "ARIMA" OR "statistical forecasting") AND ("XGBoost" OR "residual learning" OR "hybrid forecasting")
Google Scholar / manual tracking	15	("hybrid SARIMAX XGBoost" OR "residual learning forecasting" OR "explainable crop-yield prediction" OR "food policy decision support")

C. Inclusion and Exclusion Criteria

Studies were included when they were recent international journal articles, addressed production or yield forecasting, used statistical, machine learning, deep learning, hybrid, or explainable modelling techniques, and reported adequate methodological information. Studies were excluded when they were duplicates, unrelated to forecasting, focused only on plant disease or image classification without yield/production prediction, lacked clear evaluation metrics, or were too general to support the updated research questions and objectives.

D. Study Selection Process

The selection process began with 124 records. After four duplicates were removed, 120 records were screened by title and abstract. Forty-five records were excluded at this stage. The remaining 75 reports were assessed through metadata, abstract, and full-text information. Fifteen reports were excluded because they were too general, insufficiently focused on rice or staple crop-yield forecasting, unresolved in metadata or publication status, or below the quality threshold. The final synthesis retained 60 studies.

Screening and quality assessment were conducted by the first author and then checked by the co-authors. Initial screening was performed based on title, abstract, publication year, and topic relevance. Potentially eligible articles were then examined using metadata, abstract, and full-text information. Any uncertainty regarding inclusion, exclusion, or quality score

was discussed among the authors until consensus was reached. When a study was relevant to crop-yield forecasting but not directly focused on rice production, it was retained only if it contributed to methodological synthesis, such as hybrid modelling, remote sensing, explainability, temporal validation, or DSS relevance.

To strengthen reproducibility, the operational search strategy was defined using three Boolean search blocks: S1 = ("rice production" OR "rice yield" OR "crop yield") AND ("forecasting" OR "prediction") AND ("machine learning" OR "deep learning" OR "hybrid model"); S2 = ("SARIMAX" OR "SARIMA" OR "ARIMA") AND ("XGBoost" OR "residual learning" OR "hybrid forecasting") AND ("agriculture" OR "crop yield" OR "rice"); and S3 = ("explainable AI" OR "SHAP" OR "decision support system") AND ("crop yield" OR "rice production" OR "food security"). These strings were adapted to the syntax of each database while preserving the same conceptual blocks: forecasting object, modelling approach, and policy/explainability relevance.

The screening procedure was implemented in four sequential stages: duplicate removal, title-abstract screening, eligibility assessment using metadata and full-text information, and final quality-based inclusion. A screening log was maintained to record the reason for exclusion, especially for studies that were general agricultural AI papers, disease-only image classification papers, or studies without clear forecasting evaluation.

Bias mitigation was addressed through predefined inclusion and exclusion criteria, multi-source database searching, backward and manual tracing for relevant studies, standardized extraction fields, co-author checking, and consensus-based resolution of uncertain cases. To reduce confirmation bias toward the proposed hybrid framework, the synthesis explicitly reports the limitations of SARIMAX, XGBoost, SHAP, and hybrid models, and avoids claiming empirical superiority before future validation.

E. Quality Assessment

Quality assessment used six criteria: clarity of forecasting problem, clarity of dataset and variables, clarity of modelling method, availability of evaluation metrics, comparison with baselines or alternatives, and discussion of limitations. Each criterion was scored using a binary scale: 1 if the criterion was clearly satisfied and 0 if it was not clearly satisfied, producing a total score between 0 and 6. Studies scoring 4 or above were retained because they provided sufficient methodological clarity, evaluation evidence, and relevance to the research questions, whereas studies below this threshold were excluded. The quality threshold was applied as a transparency filter, not as a statistical weighting mechanism.

F. Data Extraction and Synthesis Strategy

For each selected study, data were extracted on author and year, journal, country or region, prediction object, dataset source, variables, modelling method, model type, evaluation metrics, key findings, limitations, and relevance to the

proposed SARIMAX-XGBoost framework. The synthesis combined descriptive analysis, thematic grouping, and RQ/RO-based interpretation across contextual determinants, predictor sources, forecasting approaches, methodological limitations, hybrid residual learning, evaluation design, explainability, and DSS relevance.

G. Critical Synthesis of Methodological Limitations

SARIMAX is limited by linear and parametric assumptions. It can represent trend, seasonality, autoregressive structure, and exogenous predictors, but may underfit nonlinear interactions among climate, productivity, and harvested area. XGBoost has the opposite profile: it can capture nonlinear relationships but requires careful feature engineering, chronological validation, and interpretability support. SHAP helps explain model contributions, but it should not be treated as causal evidence. These limitations support a residual-based hybrid design that combines interpretable statistical structure, nonlinear correction, and explanation-oriented analysis.

The included studies also reveal broader methodological risks: data bias, spatial-scale mismatch, limited temporal generalization, dependence on remote-sensing availability, weak comparison against time-aware baselines, and limited translation of forecasts into decision-support outputs. These issues are important because high accuracy in one region, season, or dataset does not guarantee stable performance for Indonesian provincial rice production. The inclusion and exclusion criteria are summarized in Table 5, while the quality assessment criteria are detailed in Table 6.

Table 5.
Inclusion and Exclusion Criteria

Aspect	Inclusion Criteria	Exclusion Criteria
Publication period	Articles published from 2023 to 2026	Articles outside the review period
Publication type	International journal articles and relevant review articles	Books, blogs, theses, non-peer-reviewed documents
Topic	Rice/crop production or yield forecasting, hybrid ML, XAI, or DSS relevance	Disease-only classification or unrelated agricultural studies
Method	Clear statistical, ML, DL, hybrid, or explainable method	No clear model or no evaluation metric
Relevance	Useful for RQ1 or RQ2 and the related RO1–RO7	Too general or not connected to the review objective

Table 6.
Quality Assessment Criteria

Code	Criterion	Purpose
QA1	Forecasting problem clearly defined	Ensures the study addresses prediction or forecasting.
QA2	Dataset and variables clearly described	Ensures reproducibility of the evidence.
QA3	Model or method clearly explained	Ensures methodological transparency.

Code	Criterion	Purpose
QA4	Evaluation metrics reported	Ensures accuracy evidence is available.
QA5	Compared with baselines or alternatives	Ensures relative model performance can be interpreted.
QA6	Limitations or future work discussed	Ensures critical reflection and gap identification.

Figure 1 illustrates the PRISMA-style flow of study identification, screening, eligibility assessment, and final inclusion.

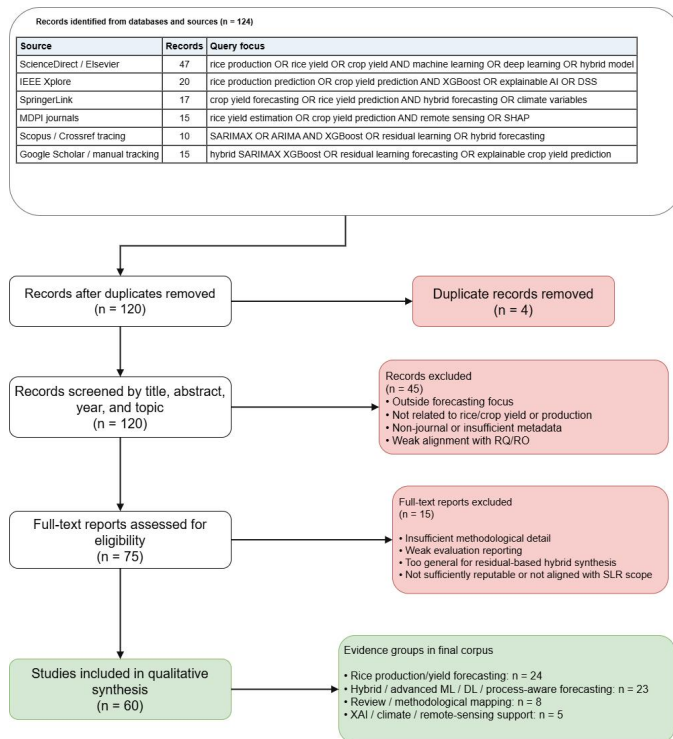


Fig. 1. PRISMA-style flow diagram of study selection.

IV. RESULTS

To address the reviewer concern regarding separation between review findings and framework development, the results are organized into three layers: descriptive findings, thematic synthesis, and conceptual framework development. Descriptive findings report the corpus characteristics and methodological patterns, thematic synthesis answers RQ1 and RQ2, and the conceptual framework section translates the identified gaps into a proposed SARIMAX-XGBoost residual architecture.

A. Study Selection and Final Corpus Characteristics

The selection process produced 60 included articles from 124 initial records. After duplicate removal, 120 records were screened by title and abstract; 45 were excluded, and 75 reports were assessed for eligibility. Fifteen reports were removed after full-text and quality checks. The final corpus therefore contains studies that are recent, methodologically relevant, and aligned with at least one updated research question or related research objective.

The corpus was organized into evidence groups covering rice production or rice yield forecasting studies [1]–[24], nonlinear forecasting and explainability support studies [26]–[32], review or methodological mapping studies [33]–[40], hybrid or advanced ML/DL crop-forecasting studies [41]–[59], and methodological or DSS-supporting studies [60]–[65]. This structure grounds the synthesis in rice-specific evidence while drawing methodological support from broader crop-yield research. Table 7 reports the detailed study selection outcomes across the PRISMA-style screening stages.

Table 7.
Detailed study selection outcomes

Selection stage	Number	Interpretation for the review
Initial records identified	124	Starting pool gathered from database and source searches.
Duplicates removed	4	Duplicate records were removed before title and abstract screening.
Records screened	120	Records were screened based on title, abstract, year, topic, and journal relevance.
Records excluded	45	Excluded mostly because they were too general, unrelated to forecasting, or outside the methodological focus.
Reports assessed for eligibility	75	Reports were checked through abstract, metadata, full-paper relevance, and quality criteria.
Reports excluded after full-paper review	15	Excluded because of weak focus, unclear metadata, or insufficient methodological detail.
Final studies included	60	Studies retained for descriptive, thematic, and RQ/RO-based synthesis.

B. Publication Trend and Source Distribution

Publication trends show growing attention to machine learning, hybrid modelling, and multi-source data in crop-yield forecasting. The final corpus includes four papers from 2023, fourteen from 2024, twenty-one from 2025, and twenty-one from 2026. Although the distribution is affected by the review period and article-in-press status, it confirms that the topic has become increasingly active in recent years.

The publication outlets indicate an intersection of agricultural informatics, remote sensing, environmental modelling, and applied machine learning. Smart Agricultural Technology, Computers and Electronics in Agriculture, Agricultural and Forest Meteorology, Field Crops Research, Remote Sensing, and the International Journal of Applied Earth Observation and Geoinformation appear repeatedly in the corpus, supporting the positioning of rice forecasting as both an agricultural and data-driven decision-support problem. Figure 2 shows the distribution of included papers by publication year.

Table 8 summarizes the four evidence groups in the final corpus and their contribution to this SLR. Quantitatively, the 60-study corpus indicates that rice-specific forecasting (24 studies, 40.0%) and hybrid/advanced ML/DL/process-aware forecasting (23 studies, 38.3%) dominate the evidence base. Because methodological coding is non-mutually exclusive, Table 9 summarizes the most frequent methodological patterns. The synthesis does not support a universal claim that one

algorithm is always superior; rather, it indicates that ML/ensemble models, remote-sensing-based data fusion, and hybrid/process-aware models are the most dominant and consistently promising patterns across the reviewed literature.

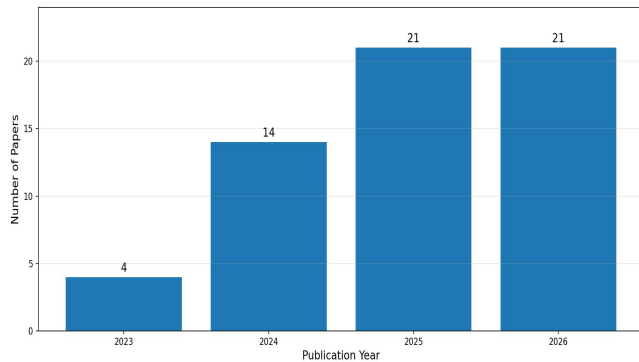


Fig. 2. Distribution of included papers by publication year.

Table 8.
Evidence Groups in The Final Corpus

Evidence group	Number of studies	Main contribution to this SLR
Rice production/yield forecasting	24	Supports RQ1 and RO1–RO2 by identifying contextual determinants and data-characteristic evidence.
Hybrid/advanced ML/DL/process-aware forecasting	23	Supports RQ2 and RO3–RO5 by informing hybrid residual modelling logic, advanced learning strategies, and methodological limitations.
Review/methodological mapping	8	Supports RO3 by providing broader methodological positioning, remote-sensing review evidence, and trend synthesis across crop-yield forecasting.
XAI/climate/remote-sensing support	5	Supports RQ2 and RO6–RO7 by informing evaluation, explanation, climate, and DSS-use dimensions.

Table 9.
Quantitative Synthesis of Dominant Methodological Patterns

Pattern coded in the 60 studies	Quantitative implication
Remote sensing/UAV/hyperspectral/data fusion ($n=25$)	Most frequent data strategy for spatial, early-season, and province-aware forecasting.
Machine learning/ensemble learning ($n=22$)	Dominant nonlinear modelling pattern for tabular and multi-source agricultural data.
Hybrid/process-aware/data-assimilation models ($n=18$)	Most promising direction for linking domain structure and data-driven learning.
Deep learning/temporal neural models ($n=12$)	Useful for image, sequence, and spatiotemporal data, but validation-sensitive.
Statistical/time-series baselines ($n=6$)	Important interpretable baselines and residual-diagnostic tools.

XAI/SHAP/DSS-oriented studies ($n=5$)	Still limited, indicating an underdeveloped policy-explanation layer.
---	---

Thus, the quantitative implication is that recent crop-forecasting literature has shifted from single-model comparison toward data-rich, hybrid, and validation-aware approaches, whereas residual-level explainability and DSS-ready policy translation remain comparatively sparse. The dominance of remote sensing, UAV imagery, hyperspectral data, and data-fusion strategies can be explained by the need to observe spatial variability, phenological stages, and early-season crop conditions that are not fully captured by historical statistics alone. Machine learning and ensemble methods dominate because agricultural production responses are often nonlinear and influenced by interactions among climate, land, productivity, and regional characteristics. Hybrid and process-aware models are increasingly used because they combine domain structure with data-driven learning, while the limited number of XAI/DSS-oriented studies indicates that forecast interpretation and policy translation remain underdeveloped compared with predictive modelling.

C. RQ1: Contextual Determinants and Data Characteristics

RQ1 examines temporal, provincial, production-related, and climate-related factors reported to influence rice production forecasting at regional or provincial scales. It directly supports RO1 and RO2 by identifying key determinants and synthesizing data characteristics used in recent rice and crop-yield forecasting studies. The evidence shows that rice forecasting is time-dependent, data-intensive, and region-specific. Studies at field, district, county, and national scales link production outcomes to crop calendars, phenological stages, rainfall distribution, temperature variation, water availability, local production systems, UAV imagery, hyperspectral data, and edge-based deep learning [2]–[24].

Temporal features are essential because rice production follows seasonal and phenological cycles. Multitemporal satellite data, phenology-based aggregation, early-season estimation, in-season forecasting, temporal fusion, vegetation-index monitoring, hyperspectral imaging, and lightweight deep learning are commonly used to capture planting, vegetative, reproductive, and harvest-stage signals across rice-specific and broader crop-yield forecasting studies [2]–[24], [26], [35], [39], [44], [65]. For Indonesian provincial forecasting, these findings support the use of lagged production, seasonal terms, and time-varying climate variables as core data characteristics.

Inter-provincial variation is also critical. Evidence from China, India, Sri Lanka, Peru, Tanzania, and Southeast Asia shows that model performance and variable importance differ by region, spatial scale, and agro-climatic context [2]–[3], [11]–[14], [18], [20]. Indonesian provinces therefore should not be treated as homogeneous units because they differ in irrigation infrastructure, cropping intensity, harvested area, climate exposure, and productivity level.

Production-related variables and climate variables form the main explanatory layer. Production is structurally linked to harvested area and productivity, while climate affects yield realization through rainfall distribution, temperature stress, drought, flooding, or delayed planting. Studies integrating meteorological variables, climate reanalysis, remote sensing, biophysical indicators, UAV data, and crop-model outputs consistently show stronger forecasting relevance than approaches based only on historical yield [2]–[7], [15], [17], [21]–[24], [27], [48], [50], [55], [59], [63]. This evidence establishes the contextual input space for residual-based modelling. Table 10 summarizes the contextual determinants, evidence from the reviewed studies, and implications for regional or provincial rice forecasting.

Table 10.

Contextual Determinants and Data Characteristics for RQ1

Analytical aspect	Evidence from reviewed studies	Implication for regional or provincial rice forecasting
Temporal patterns	Phenology-based aggregation, multitemporal remote sensing, early-season forecasting, in-season prediction, hyperspectral imaging, and mobile/edge yield forecasting are widely used across rice-specific and broader crop-yield forecasting studies [2]–[24], [26], [35], [39], [44], [65].	Forecasting should include lagged production, seasonal structure, and time-dependent exogenous variables.
Inter-provincial variation	Studies across different regions show that yield response and model performance vary by location and agro-climatic condition [2]–[3], [11]–[14], [18].	Province-aware modelling is needed because production dynamics and climate exposure differ across regions.
Production-harvested area-productivity relationship	Production is structurally related to land availability and yield per unit area; related variables repeatedly appear in forecasting studies [1], [15]–[16].	Harvested area and productivity should be treated as core production-related variables.
Climate influence	Meteorological, climate reanalysis, drought, and environmental factors improve forecasting and explain yield variability [2]–[7], [27], [48], [63].	Climate variables should be included to explain production fluctuations beyond historical output.
Remote sensing and biophysical indicators	SAR, optical, UAV, PlanetScope, Sentinel, hyperspectral, vegetation-index, and biophysical variables support rice and crop-yield estimation [3], [6]–[7], [10]–[13], [18], [21]–[24], [55], [59].	Remote sensing can complement official statistics when available, especially for spatial and early-season monitoring.

D. RQ2: Hybrid Residual Modelling Logic and Methodological Synthesis

RQ2 examines how recent literature supports an interpretable residual-based hybrid forecasting framework that integrates residual separation, model evaluation, explainability, and food policy DSS relevance. The first part of RQ2 supports RO3–RO5 by classifying forecasting approaches, analyzing methodological limitations and gaps, and formulating the

evidence-based logic of a Hybrid Residual SARIMAX-XGBoost framework. The literature supports this logic because SARIMAX and XGBoost address different modelling needs: SARIMAX represents autoregressive, seasonal, and exogenous structures, while XGBoost and related machine learning approaches handle nonlinear interactions, threshold effects, residual correction, data fusion, and advanced forecasting architectures [25], [27], [29], [41]–[42], [47], [52], [54]–[59].

The proposed Hybrid Residual SARIMAX-XGBoost framework should not be interpreted as an empirically validated superior model. Rather, it is a conceptual framework derived from the synthesis of methodological patterns, limitations, and gaps identified in the reviewed literature. Empirical superiority can only be claimed after implementation, benchmark comparison, residual diagnostics, time-aware validation, and province-level testing using Indonesian rice production data. The proposed conceptual architecture and DSS integration workflow are illustrated in Fig. 3.

The conceptual architecture in Fig. 3 clarifies the workflow requested by the reviewer. The process begins with province-period data preparation, followed by time-aware splitting and SARIMAX baseline estimation. The residual series is then diagnosed before XGBoost is trained as a residual learner. The hybrid forecast is obtained by adding the SARIMAX baseline forecast and the predicted residual correction. SHAP explains the residual-correction layer, and DSS integration converts the forecast and explanation outputs into risk signals, provincial rankings, influential-variable summaries, and scenario-based policy insights.

The reviewed studies indicate that residual separation requires well-structured province-period data. The target variable is provincial rice production, while core explanatory variables include harvested area, productivity, and selected climate indicators. Exploratory analysis should examine trend, seasonality, missing values, outliers, lag structure, and inter-provincial differences before residual modelling is justified.

The first evidence-based modelling implication is that SARIMAX should be used to estimate the linear, seasonal, autoregressive, and exogenous baseline. Residual diagnostics become a gatekeeping step: if residuals show autocorrelation, seasonal clustering, province-specific bias, or association with climate anomalies, they contain learnable information; if they resemble white noise, residual learning may add little value and risk overfitting.

The second implication is that XGBoost should operate as a residual learner rather than an unrestricted standalone forecaster. It can learn SARIMAX residuals using engineered features such as original exogenous variables, lags, province indicators, seasonal dummies, rolling statistics, climate anomalies, or transformed climate features. The final forecast combines the SARIMAX baseline and XGBoost residual correction, giving each component a clear function.

A deeper critical interpretation shows that the models should be viewed as complementary rather than competing alternatives. SARIMAX is transparent, parsimonious, and useful for modelling temporal, seasonal, autoregressive, and exogenous structures, but it may underfit threshold effects,

nonlinear climate responses, and province-specific shocks. XGBoost is stronger for nonlinear tabular interactions and heterogeneous predictor effects, but it is not automatically time-aware and may overfit if lags, seasonal indicators, province features, and chronological validation are not carefully designed.

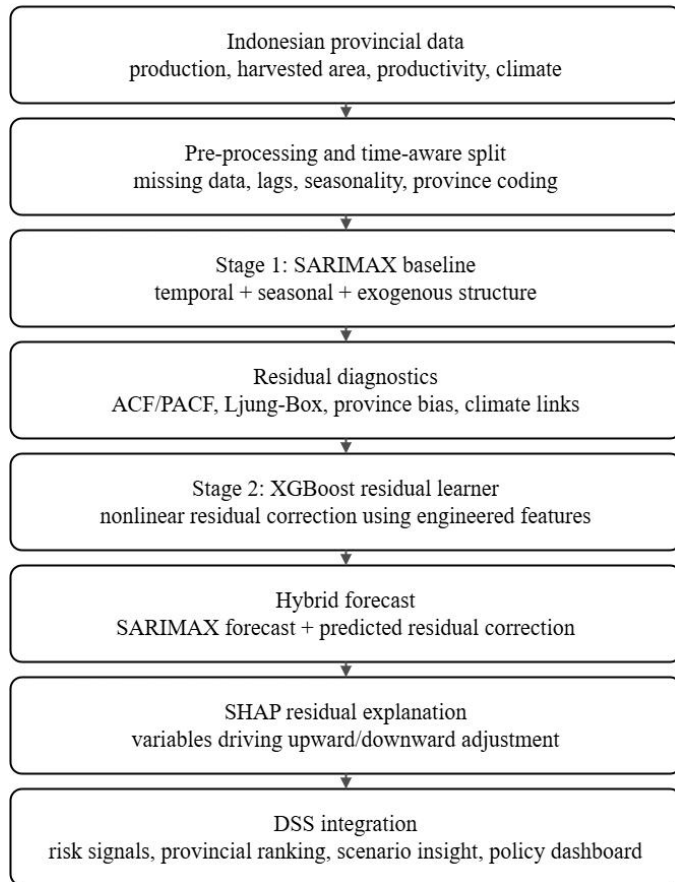


Fig. 3. Conceptual architecture of the hybrid residual SARIMAX-XGBoost-SHAP framework and DSS integration.

The main trade-off of a residual hybrid design is that it can improve functional role separation between the statistical baseline and nonlinear correction layer, but it also adds modelling complexity and may propagate baseline errors if residual diagnostics are ignored. SHAP improves interpretability by showing which variables drive residual correction, yet it should be interpreted as model explanation rather than causal evidence, especially when predictors are correlated or when spatial-temporal dependence remains unresolved. Table 11 translates this modelling logic into evidence-based operational implications for the proposed Hybrid Residual SARIMAX-XGBoost framework.

Table 11.

Evidence-based logic for Hybrid Residual SARIMAX-XGBoost		
Evidence-based modelling implication	Operationalization in the proposed framework	Expected analytical output
Province-period data are	Organize production,	Clean dataset for

Evidence-based modelling implication	Operationalization in the proposed framework	Expected analytical output
required	harvested area, productivity, and climate variables by province and period; handle missing data and outliers.	contextual and temporal modelling.
Temporal structure must be examined	Inspect trend, seasonality, lag structure, inter-provincial variation, and climate-production relationships.	Evidence for seasonal terms, lags, and exogenous variables.
Linear-seasonal baseline should be isolated	Use SARIMAX to estimate autoregressive, seasonal, and exogenous effects.	Interpretable baseline forecast and residual series.
Residuals must be diagnostically meaningful	Assess residual autocorrelation, seasonal clustering, province bias, and links with climate anomalies.	Decision on whether residual learning is justified.
Nonlinear residual patterns require a separate learner	Train XGBoost to predict residual patterns using engineered temporal, provincial, and climate features.	Nonlinear residual correction model.
Hybrid forecast should combine baseline and correction	Combine SARIMAX forecast and XGBoost residual prediction.	Final hybrid forecast for provincial rice production.
Evaluation and explanation must be policy-oriented	Evaluate against baselines and apply SHAP to the residual learner.	Accuracy metrics, residual explanations, and DSS-ready insights.

E. RQ2: Evaluation, Residual-Based Explainability, and Food Policy DSS Relevance

As the evaluation and decision-support dimension of RQ2, this subsection addresses RO6 and RO7 by defining how forecasting performance, residual-based explainability, and DSS-oriented outputs can be evaluated for food policy analysis. Performance evaluation should use multiple metrics, such as RMSE, MAE, MAPE, SMAPE, and R-squared, and compare the hybrid model against SARIMAX-only, XGBoost-only, naive seasonal, process-aware, deep learning, and other benchmark models [1]–[24], [26]–[28], [30]–[31], [39]–[40], [49], [50], [51], [52], [53], [54], [55], [56], [57], [58], [59], [62], [63], [64], [65].

The validation design must preserve temporal ordering. Random splitting may leak future information into training data, so rolling-origin validation, walk-forward validation, province-wise validation, or chronological train-test splits are more appropriate. Results should also be examined by forecasting horizon and province because national average error can hide weak performance in smaller or climate-sensitive regions.

Residual-based explainability should focus on the correction layer rather than claiming to explain the entire production process. In this role, SHAP identifies which variables drive upward or downward adjustment from the SARIMAX baseline, such as harvested-area contraction,

productivity anomaly, rainfall deviation, seasonal timing, or province-specific behaviour [30]–[32]. This framing makes interpretation more relevant to policy use while avoiding causal overclaiming.

Decision-support evaluation should assess whether forecasts can be translated into usable outputs: forecast intervals or risk categories, provincial rankings of expected production decline or increase, influential-variable summaries, and scenario-based insights. These outputs help decision makers identify whether a predicted decline is associated with harvested-area contraction, productivity decrease, rainfall anomaly, or residual nonlinear behaviour.

From an information systems perspective, the value of the proposed framework lies in converting forecasting outputs into decision-support artefacts. These artefacts include provincial risk alerts, production-change rankings, residual-driver summaries, forecast-horizon comparisons, and scenario insights that can be incorporated into a food policy DSS. Table 12 summarizes the recommended evaluation, residual-based explainability, and policy-use dimensions for RQ2.

Table 12.
Evaluation, explainability, and policy-use dimensions for RQ2

Evaluation dimension	Recommended assessment	Policy-use relevance
Predictive performance	Compare SARIMAX-XGBoost with SARIMAX-only, XGBoost-only, naive seasonal benchmark, process-aware models, deep learning models, and other ML baselines using RMSE, MAE, MAPE, SMAPE, and R-squared.	Checks whether forecasts are reliable enough for planning.
Temporal validation	Use time-aware validation such as rolling-origin, walk-forward, or chronological train-test split.	Prevents future-data leakage and supports realistic forecasting evaluation.
Residual-based explainability	Apply SHAP to the XGBoost residual learner to identify variables driving forecast adjustment from the SARIMAX baseline.	Explains why a forecast is adjusted upward or downward.
Province and horizon usefulness	Assess forecasts and explanations by province, season, and forecasting horizon.	Supports region-specific intervention and planning.
DSS output design	Translate forecasts into dashboards, risk signals, variable rankings, and scenario summaries.	Turns model output into actionable policy information.

F. Integrated Research Gap and Evidence Matrix

The updated RQ- and RO-based synthesis of 60 included studies reveals three connected gaps. First, contextual determinants and data characteristics are often reported, but not always translated into province-aware modelling requirements. Second, hybrid and advanced machine learning models are increasingly used, yet many do not explicitly separate linear-seasonal structures from nonlinear residual patterns. Third, evaluation frequently emphasizes accuracy while giving less attention to residual-based explanation and DSS-ready policy outputs. The central explicit gap from the 60-article

corpus is that province-aware determinants, residual-based hybrid modelling, time-aware evaluation, residual-level explainability, and DSS-ready food policy outputs have not yet been synthesized as one coherent review framework.

Table 13 presents the research gap matrix and proposed response, while Table 14 provides a compact included-study evidence matrix organized by synthesis stream rather than by individual study. This compact format reduces redundancy while showing how the 60 included studies support the main analytical dimensions of this SLR. The complete study-level extraction matrix can be provided as supplementary material.

Table 13.
Research Gap Matrix and Proposed Response

Gap	Evidence from literature	Proposed response
Contextual factors not translated into modelling requirements	Studies report temporal, spatial, production-related, and climate variables, but often treat them mainly as inputs [1]–[20].	Map contextual determinants into province-aware variables, lags, seasonal terms, and exogenous predictors.
Insufficient attention to province-level heterogeneity	Regional studies show spatial variability, but many models focus on field or national scales [2]–[3], [11]–[14], [20].	Evaluate forecasts and explanations at provincial level.
Limited separation of linear-seasonal and nonlinear residual structures	Hybrid studies exist, but many do not explicitly define residual learning as a second-stage mechanism [26]–[28], [30], [41]–[54].	Use SARIMAX for the linear-seasonal-exogenous baseline and XGBoost for nonlinear residual correction.
Explainability not focused on residual correction	XAI is increasingly used, but explanations often remain disconnected from the specific function of each model component [30]–[32].	Use SHAP to explain residual adjustment rather than claiming causal explanation of the whole production process.
Forecast outputs not translated into DSS-ready information	Many studies report error metrics without explaining how forecasts support policy decisions.	Design outputs such as risk categories, provincial ranking, influential variables, and scenario summaries.

Table 14.
Compact Included-Study Evidence Matrix by Synthesis Stream

Synthesis stream and references	Dominant evidence and implication
Rice production and rice-yield forecasting [1]–[24]	Agricultural statistics, climate variables, remote sensing, UAV, hyperspectral imagery, and crop-model outputs support province-aware, multivariate, and time-aware forecasting.
Review and methodological mapping [33]–[40]	Reviews show a shift toward data-rich, AI-based, remote-sensing-based, and multi-source prediction frameworks.
Hybrid, advanced ML/DL, and process-aware forecasting [26]–[28], [30], [41]–[59]	Hybrid learning, data fusion, process-aware modelling, deep learning, neural architectures, and crop-model integration support combining domain structure with nonlinear learning.
XGBoost and ensemble-based nonlinear prediction [17], [27]–[29], [42],	Ensemble learning and gradient boosting support XGBoost as a residual learner for nonlinear patterns not fully represented by statistical baselines.

[56]–[59]			
Explainability and residual interpretation [30]–[32]	SHAP and explainable ML support residual-level interpretation while avoiding causal overclaiming.	[43]–[47]	crop-yield prediction; benchmark error metrics. Supports hybrid logic; residual diagnostics less explicit. Attention, spatiotemporal, transfer-learning, multi-task, and knowledge-informed models; RS/crop-model data. Strong modelling evidence; interpretability concerns.
Evaluation and operational relevance [35], [38]–[39], [62], [65]	In-season forecasting, temporal validation, trend treatment, climate-based forecasting, and operational monitoring support realistic evaluation.	[26], [48]	Drought-related AI and spectral-temporal crop prediction; climate/spectral data. Supports climate-sensitive forecasting; policy translation limited.
Food policy DSS relevance [1], [33], [39]–[40], [48], [64]	Food security and sustainability-oriented studies support DSS-ready outputs: risk signals, provincial rankings, influential-variable summaries, and scenario insights.	[27], [49]–[50]	Hybrid potato, XGBoost wheat, and Sentinel-meteorological yield studies; seasonal/field scale. Supports XGBoost/hybrid ML; crop transferability must be tested.
		[28], [30], [51]	Explainable ML and DL using environmental/RS variables; SHAP/uncertainty/error metrics. Supports XAI; explanations are model-based, not causal.
		[52]–[54]	Data assimilation, crop-model integration, pixel-level rice prediction, and knowledge-guided DL; seasonal/large-area scale. Supports process-aware hybrid forecasting; DSS artefacts limited.
		[30]–[32]	XAI, climate-based forecasting, multisource ML, sustainability review, and in-season forecasting; UAV/RS/weather data. Supports operational monitoring; residual-level SHAP sparse.
		[21]–[24]	Rice-yield studies using mobile CNN-LSTM, UAV, hyperspectral, agronomic, VI, and meteorological data. Strong rice-specific evidence; data availability challenging.
		[39], [55]–[59]	Robust wheat prediction, RS review, neural networks, meteorological partitioning, federated learning, and Bayesian neural networks. Advanced/uncertainty-aware evidence; DSS integration limited.
		[40]	Maize-yield SLR; crop-specific review evidence. Useful SLR comparison; not focused on rice, residual learning, or DSS.

To expand the evidence synthesis, Table 15 provides a more detailed study-level matrix for selected reviewed studies across rice forecasting, hybrid modelling, XGBoost/ensemble learning, explainability, and review-level evidence. The matrix explicitly summarizes dataset or data-source characteristics, evaluation metrics or evaluation orientation, forecasting horizon or spatial scale, strengths, and limitations. Because the full corpus contains 60 studies, the complete study-level extraction matrix can be provided as supplementary material, while the table below presents the main evidence most directly connected to the proposed SARIMAX-XGBoost-SHAP framework.

Table 15.

Detailed Study-Level Evidence Matrix for Selected Reviewed Studies	
Study evidence	Dataset, metrics/horizon, strengths, and limitations
[1]	Indonesia rice-production data; accuracy/error evaluation. Strong Indonesia relevance; residual learning, SHAP, and DSS translation remain limited.
[2]	India district rice yield; climate reanalysis and remote sensing; RMSE/MAE/R2-type metrics. Strong climate-RS evidence; provincial transfer must be validated.
[3]	SAR, optical, and meteorological rice-yield data; field/county scale; error metrics. Strong multi-source evidence; no residual SARIMAX-XGBoost design.
[4]	Australia rice yield; phenology-based satellite/weather aggregation; seasonal horizon. Strong temporal aggregation; different policy context.
[5]	Remote sensing + crop model + deep learning for rice yield; RMSE/R2-type evaluation. Supports process-aware DL; data and computation intensive.
[6]–[8]	SAR/optical, meteo, and spatial heterogeneity rice studies. Confirms spatial heterogeneity and predictors; DSS and residual interpretation limited.
[9]–[13]	Satellite, PlanetScope, Sentinel, climate, and field-scale rice-yield studies; pre-harvest/field horizons. Strong early-season evidence; provincial policy outputs limited.
[14]–[16]	Farm-level or experimental yield datasets for maize, rice, beans, or rice gaps; ML/crop-modelling evaluation. Supports multi-crop learning; transfer must be cautious.
[17]–[18]	Multi-source rice-yield and UAV imagery; ML metrics such as RMSE/R2; field/regional scale. Supports nonlinear ML; residual-DSS role not explicit.
[19]–[20]	Early forecasting, spatial typology, and phenology-aligned temporal fusion for rice. Strong temporal evidence; not SARIMAX + XGBoost + SHAP.
[33]–[38]	Review evidence on ML crop-yield prediction, in-season forecasting, Earth observation, and detrending. Useful positioning; no provincial residual-DSS framework.
[41]–[42]	Knowledge-informed hybrid ML and ensemble

G. Specific Contribution of the Present SLR

The present SLR contributes a structured synthesis of rice and crop-yield forecasting approaches by connecting three evidence streams that are often discussed separately: temporal and exogenous production modelling, machine learning-based nonlinear prediction, and explainable decision-support use. Through two main research questions and seven operational objectives, this synthesis positions provincial rice production forecasting as a multivariate temporal problem shaped by seasonality, harvested area, productivity, climate variability, and inter-provincial heterogeneity. The quantitative distribution of the corpus further shows that rice-specific forecasting and hybrid/advanced ML-DL studies dominate the evidence base, while XAI and DSS-oriented studies remain limited, reinforcing the need for an integrated residual-based and policy-oriented synthesis.

To refine the novelty claim, Table 16 compares this SLR with existing review studies. The comparison shows that prior reviews provide valuable coverage of machine learning, remote sensing, in-season forecasting, detrending, drought-related AI, sustainability, and crop-specific prediction. However, they do not simultaneously synthesize provincial rice production forecasting, SARIMAX-based temporal-exogenous baselines, XGBoost residual learning, SHAP-based residual explanation, and DSS-oriented food policy outputs.

Table 16.
Systematic Novelty Comparison Against Existing Review Studies

Existing review stream	Gap relative to this manuscript and unique contribution
Broad ML crop-yield reviews [33]–[34]	They map general AI/ML trends, while this SLR focuses on provincial rice production and residual-based SARIMAX-XGBoost-SHAP design.
Remote-sensing and in-season reviews [35]–[36], [39]	They emphasize EO/RS forecasting, while this SLR translates such evidence into temporal-exogenous baseline modelling, residual diagnostics, and DSS outputs.
Detrending and methodological reviews [37]–[38]	They discuss modelling and trend treatment, while this SLR links time-series baseline, nonlinear residual correction, SHAP interpretation, and policy use.
Food security and sustainability reviews [40], [48], [64]	They establish policy relevance, while this SLR specifies DSS-ready artefacts such as risk signals, provincial rankings, driver summaries, and scenario insights.
Present SLR	The novelty is conceptual integration, not empirical superiority: provincial determinants + SARIMAX baseline + XGBoost residual learner + SHAP residual explanation + DSS integration.

Its main methodological contribution is the conceptualization of an interpretable Hybrid Residual SARIMAX–XGBoost framework. SARIMAX is positioned to estimate the linear, seasonal, autoregressive, and exogenous baseline, while XGBoost learns the remaining nonlinear residual signals. SHAP then explains the residual adjustment layer, making the hybrid structure more transparent than a loose ensemble or a simple comparison of forecasting algorithms.

The present SLR also bridges forecasting models and food policy decision support. It converts the limitations of SARIMAX, XGBoost, and SHAP into design requirements: residual diagnostics before nonlinear correction, time-aware validation, residual-based explanation, and DSS-oriented outputs such as provincial risk signals, influential-variable summaries, and scenario-based policy insights.

H. Preliminary Proof-of-Concept Protocol

In response to the recommendation for a preliminary proof-of-concept, this SLR adds an implementation protocol rather than an empirical performance claim. A full pilot experiment is beyond the scope of the current review because the manuscript is designed as an SLR and the empirical Indonesian provincial dataset has not yet been assembled in this paper. However, the following protocol specifies how the proposed framework can be operationalized and tested in future work without changing the conceptual status of the present manuscript.

The preliminary protocol consists of six steps: (1) construct a province-period panel containing rice production, harvested area, productivity, climate variables, lags, seasonal indicators, and province identifiers; (2) fit SARIMAX as the temporal-exogenous baseline using chronological training data; (3) diagnose residuals using residual plots, ACF/PACF, Ljung-Box, province-level bias, and residual-climate association; (4) train XGBoost to predict residual correction using engineered temporal, provincial, production, and climate

features; (5) compare SARIMAX-only, XGBoost-only, naive seasonal, and hybrid models using RMSE, MAE, MAPE, SMAPE, and province/horizon analysis; and (6) apply SHAP to the residual learner and translate the outputs into DSS artefacts such as risk categories, provincial rankings, residual-driver summaries, and scenario-oriented policy insights.

V. CONCLUSION

This SLR synthesized 60 recent studies on rice production forecasting, crop-yield forecasting, hybrid machine learning, explainable AI, and decision-support-oriented forecasting. The review shows that rice production forecasting requires a multivariate and time-aware perspective because production is influenced by temporal dependence, harvested area, productivity, climate variability, and regional heterogeneity. However, the reviewed literature still rarely integrates residual-based hybrid modelling, time-aware validation, residual-level explainability, and DSS-ready policy outputs into a single framework.

In this framework, SARIMAX is positioned as an interpretable temporal-exogenous baseline model, XGBoost as a nonlinear residual correction model, and SHAP as an explanation layer for interpreting forecast adjustments. This framework contributes to hybrid forecasting literature by clarifying the functional role of each model component and contributes to information systems research by translating forecasting outputs into DSS elements.

Practically, the proposed framework can support public-sector food policy through provincial risk signals, production-change rankings, influential-variable summaries, and scenario-based policy insights. However, the framework remains conceptual and requires empirical validation using Indonesian provincial rice production data. Future research should evaluate the framework against SARIMAX-only, XGBoost-only, naive seasonal, machine learning, and deep learning baselines using time-aware validation, province-level error analysis, forecasting-horizon evaluation, and expert validation of residual-based explanations. The revised manuscript also clarifies the implementation pathway through a preliminary proof-of-concept protocol. This protocol does not replace empirical validation, but it specifies the data structure, residual diagnostics, benchmarking, SHAP interpretation, and DSS artefacts required for future testing.

This SLR has several limitations. First, the synthesis is qualitative and evidence-mapping oriented; it does not perform statistical meta-analysis. Second, the included studies vary in crop type, spatial scale, data source, and modelling design, so their findings are used to support conceptual framework development rather than to claim direct empirical superiority of a specific algorithm.

REFERENCES

- [1] E. Erlin, A. Yudianta, L. A. Wulandhari, Y. Desnelita, N. Nasution, and Junadhi, “Enhancing rice production prediction in Indonesia using advanced machine learning models,” *IEEE Access*, vol. 12, pp. 151161–151177, 2024, doi: 10.1109/ACCESS.2024.3478738.
- [2] J. De Clercq and M. Mahdi, “Feasibility of machine-learning-based rice yield prediction in India at the district level using climate reanalysis and remote sensing data,” *Agricultural Systems*, vol. 220, Art. no. 104099, 2024, doi: 10.1016/j.agsy.2024.104099.

- [3] Y. Yu *et al.*, “Improved prediction of rice yield at field and county levels by synergistic use of SAR, optical and meteorological data,” *Agricultural and Forest Meteorology*, vol. 342, Art. no. 109729, 2023, doi: 10.1016/j.agrformet.2023.109729.
- [4] J. Brinkhoff *et al.*, “Analysis and forecasting of Australian rice yield using phenology-based aggregation of satellite and weather data,” *Agricultural and Forest Meteorology*, vol. 353, Art. no. 110055, 2024, doi: 10.1016/j.agrformet.2024.110055.
- [5] Y. Lu *et al.*, “Estimation of rice yield using multi-source remote sensing data combined with crop growth model and deep learning algorithm,” *Agricultural and Forest Meteorology*, vol. 370, Art. no. 110600, 2025, doi: 10.1016/j.agrformet.2025.110600.
- [6] S. Sah, D. Haldar, R. N. Singh, B. Das, and A. S. Nain, “Rice yield prediction through integration of biophysical parameters with SAR and optical remote sensing data using machine learning models,” *Scientific Reports*, vol. 14, no. 1, Art. no. 21674, 2024, doi: 10.1038/s41598-024-72624-4.
- [7] M. M. Islam *et al.*, “Rapid rice yield estimation using integrated remote sensing and meteorological data and machine learning,” *Remote Sensing*, vol. 15, no. 9, Art. no. 2374, 2023, doi: 10.3390/rs15092374.
- [8] Y. Zhou *et al.*, “Rice yield prediction in Hubei Province based on deep learning and the effect of spatial heterogeneity,” *Remote Sensing*, vol. 15, no. 5, Art. no. 1361, 2023, doi: 10.3390/rs15051361.
- [9] J. Clarke *et al.*, “Integrating climate and satellite data for multi-temporal pre-harvest prediction of head rice yield in Australia,” *Remote Sensing*, vol. 16, no. 10, Art. no. 1815, 2024, doi: 10.3390/rs16101815.
- [10] S. Liu *et al.*, “Rice yield estimation using multi-temporal remote sensing data and machine learning: A case study of Jiangsu, China,” *Agriculture*, vol. 14, no. 4, Art. no. 638, 2024, doi: 10.3390/agriculture14040638.
- [11] A. Jarro-Espinal *et al.*, “Field-scale rice yield prediction in northern coastal Peru using Sentinel-2 vegetation indices and machine learning models,” *Agriculture*, vol. 15, no. 19, Art. no. 2054, 2025, doi: 10.3390/agriculture15192054.
- [12] M. Ozdoğan *et al.*, “Field-scale rice area and yield mapping in Sri Lanka with optical remote sensing and limited training data,” *Remote Sensing*, vol. 17, no. 17, Art. no. 3065, 2025, doi: 10.3390/rs17173065.
- [13] S. Ibrahim, H. Balzter, and M. S. Hadi, “Estimating paddy rice yield using PlanetScope imagery and machine learning,” *Smart Agricultural Technology*, vol. 12, Art. no. 101447, 2025, doi: 10.1016/j.atech.2025.101447.
- [14] R. A. Omary *et al.*, “Farm-level yield prediction for maize, rice, and beans in Tanzania using machine learning and multi-source agricultural data,” *Smart Agricultural Technology*, vol. 13, Art. no. 101904, 2026, doi: 10.1016/j.atech.2026.101904.
- [15] S. Sahoo *et al.*, “Advanced prediction of rice yield gaps under climate uncertainty using machine learning techniques in Eastern India,” *Journal of Agriculture and Food Research*, vol. 18, Art. no. 101424, 2024, doi: 10.1016/j.jafr.2024.101424.
- [16] T. Yamaguchi *et al.*, “Understanding rice yield gaps with crop modeling and machine learning in a long-term continuous cropping experiment,” *Field Crops Research*, vol. 342, Art. no. 110477, 2026, doi: 10.1016/j.fcr.2026.110477.
- [17] S. Saha and S. Nath, “Harnessing multi-source data and machine learning for enhanced rice yield estimation,” *Computers and Electronics in Agriculture*, vol. 237, Art. no. 110633, 2025, doi: 10.1016/j.compag.2025.110633.
- [18] X. Liang *et al.*, “Rice yield prediction base on UAV multispectral imagery using machine learning methods,” *Smart Agricultural Technology*, vol. 12, Art. no. 101549, 2025, doi: 10.1016/j.atech.2025.101549.
- [19] Y. Wang *et al.*, “Early forecasting and spatial typologies indicate rice yield growth potential in Yangtze River Delta,” *Ecological Indicators*, vol. 179, Art. no. 114200, 2025, doi: 10.1016/j.ecolind.2025.114200.
- [20] Y. Lin *et al.*, “Phenology-aligned multi-task temporal fusion framework for satellite-based triple-seasonal rice yield estimation in Southeast Asia,” *International Journal of Applied Earth Observation and Geoinformation*, vol. 149, Art. no. 105231, 2026, doi: 10.1016/j.jag.2026.105231.
- [21] S. Gandotra, R. Chhikara, and A. Dhull, “Smart farming: Real-time rice yield forecasting on mobile devices using lightweight CNN-LSTM,” *Smart Agricultural Technology*, vol. 13, Art. no. 101664, 2026, doi: 10.1016/j.atech.2025.101664.
- [22] S. Rana *et al.*, “Field-scale rice yield prediction using UAV imagery and machine learning in a developing country context,” *Smart Agricultural Technology*, vol. 13, Art. no. 102114, 2026, doi: 10.1016/j.atech.2026.102114.
- [23] T. Hu *et al.*, “Yield prediction method for regenerated rice based on hyperspectral image and attention mechanisms,” *Smart Agricultural Technology*, vol. 10, Art. no. 100804, 2025, doi: 10.1016/j.atech.2025.100804.
- [24] J. A. Fernandez-Jibaja *et al.*, “Integration of agronomic information, vegetation indices (VIs), and meteorological data for phenological monitoring and yield estimation of rice (*Oryza sativa* L.),” *Smart Agricultural Technology*, vol. 12, Art. no. 101203, 2025, doi: 10.1016/j.atech.2025.101203.
- [25] R. J. Hyndman and Y. Khandakar, “Automatic time series forecasting: The forecast package for R,” *Journal of Statistical Software*, vol. 27, no. 3, pp. 1–22, 2008, doi: 10.18637/jss.v027.i03.
- [26] S. Fatholouloumi *et al.*, “Enhancing crop yield predictions through spectral-temporal features and machine learning,” *Computers and Electronics in Agriculture*, vol. 244, Art. no. 111527, 2026, doi: 10.1016/j.compag.2026.111527.
- [27] A. Amazirh *et al.*, “Seasonal field-scale wheat yield forecasting using XGBoost with radar, optical, and weather data in Morocco,” *International Journal of Applied Earth Observation and Geoinformation*, vol. 149, Art. no. 105242, 2026, doi: 10.1016/j.jag.2026.105242.
- [28] S. Hashemi *et al.*, “Yield estimation from SAR data using patch-based deep learning and machine learning techniques,” *Computers and Electronics in Agriculture*, vol. 226, Art. no. 109340, 2024, doi: 10.1016/j.compag.2024.109340.
- [29] T. Chen and C. Guestrin, “XGBoost: A scalable tree boosting system,” *Proc. 22nd ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining*, pp. 785–794, 2016, doi: 10.1145/2939672.2939785.
- [30] Y. Zhou *et al.*, “Unraveling the impact of environmental factors on wheat yield across the European Union via explainable machine learning,” *Computers and Electronics in Agriculture*, vol. 241, Art. no. 111268, 2026, doi: 10.1016/j.compag.2025.111268.
- [31] C. Kumar, J. Dhillon, Y. Huang, and K. N. Reddy, “Explainable machine learning models for corn yield prediction using UAV multispectral data,” *Computers and Electronics in Agriculture*, vol. 231, Art. no. 109990, 2025, doi: 10.1016/j.compag.2025.109990.
- [32] S. M. Lundberg and S.-I. Lee, “A unified approach to interpreting model predictions,” *Advances in Neural Information Processing Systems*, vol. 30, pp. 4765–4774, 2017, doi: 10.48550/arXiv.1705.07874.
- [33] M. M. R. Shawon *et al.*, “Crop yield prediction using machine learning: An extensive and systematic literature review,” *Smart Agricultural Technology*, vol. 10, Art. no. 100718, 2025, doi: 10.1016/j.atech.2024.100718.
- [34] M. A. Betew *et al.*, “Historical and contemporary crop yield prediction models: Key lessons and innovations,” *Smart Agricultural Technology*, vol. 13, Art. no. 101672, 2026, doi: 10.1016/j.atech.2025.101672.
- [35] J. Liu *et al.*, “In-season crop yield prediction: State of the art and future research direction,” *International Journal of Applied Earth Observation and Geoinformation*, vol. 146, Art. no. 105129, 2026, doi: 10.1016/j.jag.2026.105129.
- [36] F. Li *et al.*, “Earth observation derived yield forecasting and estimation in low- and lower-middle-income countries dominated by smallholder agriculture: A review,” *International Journal of Applied Earth Observation and Geoinformation*, vol. 148, Art. no. 105216, 2026, doi: 10.1016/j.jag.2026.105216.
- [37] N. Morales and F. J. Villalobos, “Using machine learning for crop yield prediction in the past or the future,” *Frontiers in Plant Science*, vol. 14, Art. no. 1128388, 2023, doi: 10.3389/fpls.2023.1128388.
- [38] Y. He *et al.*, “Comprehensive review of detrending methods for crop yields: Approaches, applications, and future directions,” *Agricultural and Forest Meteorology*, vol. 378, Art. no. 111017, 2026, doi: 10.1016/j.agrformet.2026.111017.
- [39] G. Xiao *et al.*, “Progress and perspectives of crop yield forecasting with remote sensing: A review,” *IEEE Geoscience and Remote Sensing*

- Magazine*, vol. 13, no. 3, pp. 338–368, 2025, doi: 10.1109/MGRS.2025.3571906.
- [40] J. N. Nyengere *et al.*, “Maize yield prediction using machine learning: A systematic literature review,” *Frontiers in Artificial Intelligence*, vol. 9, Art. no. 1735157, 2026, doi: 10.3389/frai.2026.1735157.
- [41] M. von Bloh *et al.*, “Knowledge informed hybrid machine learning in agricultural yield prediction,” *Computers and Electronics in Agriculture*, vol. 227, Art. no. 109606, 2024, doi: 10.1016/j.compag.2024.109606.
- [42] A. Ravikumar *et al.*, “A hybrid SERWI ensemble model for crop yield prediction using an inverse RMSE weighting strategy,” *Scientific Reports*, vol. 15, no. 1, Art. no. 31987, 2025, doi: 10.1038/s41598-025-31987-y.
- [43] W. Xiang *et al.*, “A crop model based on dual attention mechanism for large area adaptive yield prediction,” *Smart Agricultural Technology*, vol. 11, Art. no. 100957, 2025, doi: 10.1016/j.atech.2025.100957.
- [44] S. Ofori-Ampofo *et al.*, “On the strategy of exploring spatio-temporal information from Earth observation data for crop yield prediction,” *Smart Agricultural Technology*, vol. 12, Art. no. 101540, 2025, doi: 10.1016/j.atech.2025.101540.
- [45] A. Bueechi *et al.*, “Making optimal use of limited field-scale data for crop yield forecasting using transfer learning and Sentinel-1 and 2 data,” *Smart Agricultural Technology*, vol. 12, Art. no. 101567, 2025, doi: 10.1016/j.atech.2025.101567.
- [46] S. Liu, D. Wang, H. Guo, C. Han, and W. Zeng, “MT-CYP-Net: Multi-task network for pixel-level crop yield prediction under very few samples,” *International Journal of Applied Earth Observation and Geoinformation*, vol. 143, Art. no. 104748, 2025, doi: 10.1016/j.jag.2025.104748.
- [47] R. Fang *et al.*, “A knowledge-informed fusion network for process-aware crop yield prediction,” *International Journal of Applied Earth Observation and Geoinformation*, vol. 146, Art. no. 105154, 2026, doi: 10.1016/j.jag.2026.105154.
- [48] A. Fares *et al.*, “A review of artificial intelligence applications for predicting crop performance and enhancing food security under drought-induced water stress,” *Agricultural Water Management*, vol. 325, Art. no. 110212, 2026, doi: 10.1016/j.agwat.2026.110212.
- [49] M. Karbasi *et al.*, “Utilizing soil characteristics and hybrid machine learning for interpretable potato yield prediction: A study with satin-bowerbird optimization and deep neural network,” *Field Crops Research*, vol. 338, Art. no. 110311, 2026, doi: 10.1016/j.fcr.2025.110311.
- [50] P. Gamez *et al.*, “Alfalfa yield estimation using the combination of Sentinel-2 and meteorological data,” *Field Crops Research*, vol. 326, Art. no. 109857, 2025, doi: 10.1016/j.fcr.2025.109857.
- [51] H. Tian *et al.*, “Attention mechanism-based deep learning approach for wheat yield estimation and uncertainty analysis from remotely sensed variables,” *Agricultural and Forest Meteorology*, vol. 356, Art. no. 110183, 2024, doi: 10.1016/j.agrformet.2024.110183.
- [52] Z. Zhuang *et al.*, “Integrating data assimilation, crop model, and machine learning for winter wheat yield forecasting in the North China Plain,” *Agricultural and Forest Meteorology*, vol. 347, Art. no. 109909, 2024, doi: 10.1016/j.agrformet.2024.109909.
- [53] J. Simeon-Brocal *et al.*, “Rice yield prediction model at pixel level using machine learning and multi-temporal Sentinel-2 data in Valencia, Spain,” *Agriculture*, vol. 16, no. 2, Art. no. 201, 2026, doi: 10.3390/agriculture16020201.
- [54] Z. Cheng *et al.*, “Integrating multiple crop models and multi-source data in a knowledge-guided deep learning framework for wheat and maize yield forecasting in the Huang-Huai-Hai Plain, China,” *Field Crops Research*, vol. 340, Art. no. 110372, 2026, doi: 10.1016/j.fcr.2026.110372.
- [55] R. A. F. Ishaq *et al.*, “Geospatial robust wheat yield prediction using machine learning and integrated crop growth model and time-series satellite data,” *Remote Sensing*, vol. 17, no. 7, Art. no. 1140, 2025, doi: 10.3390/rs17071140.
- [56] N. A. K. Chang, S. Dey, and D. K. Das, “Nested sequential feed-forward neural network: A cumulative model for crop yield prediction,” *Computers and Electronics in Agriculture*, vol. 227, Art. no. 109562, 2024, doi: 10.1016/j.compag.2024.109562.
- [57] B. Petrovic, D. Paluba, A. Y. Nofrizal, and A. Kucera, “Analyzing meteorological effects on crop yield and yield prediction through machine learning with different data partitioning approaches: A case study from Czech Republic,” *Journal of Agriculture and Food Research*, vol. 26, Art. no. 102662, 2026, doi: 10.1016/j.jafr.2026.102662.
- [58] D. D. Koksai, K. E. Bennin, and B. Tekinerdogan, “Centralized and vertical federated learning for climate-based crop yield prediction,” *Computers and Electronics in Agriculture*, vol. 249, Art. no. 111851, 2026, doi: 10.1016/j.compag.2026.111851.
- [59] L. Zhao *et al.*, “Cotton yield prediction utilizing unmanned aerial vehicles (UAV) and Bayesian neural networks,” *Computers and Electronics in Agriculture*, vol. 226, Art. no. 109415, 2024, doi: 10.1016/j.compag.2024.109415.
- [60] M. J. Page *et al.*, “The PRISMA 2020 statement: An updated guideline for reporting systematic reviews,” *BMJ*, vol. 372, Art. no. n71, 2021, doi: 10.1136/bmj.n71.
- [61] B. Kitchenham, O. P. Brereton, D. Budgen, M. Turner, J. Bailey, and S. Linkman, “Systematic literature reviews in software engineering: A systematic literature review,” *Information and Software Technology*, vol. 51, no. 1, pp. 7–15, 2009, doi: 10.1016/j.infsof.2008.09.009.
- [62] M. Zachow, S. Ofori-Ampofo, H. Kunstmann, R. S. Kuzu, X. X. Zhu, and S. Asseng, “Wheat yield forecasts with seasonal climate models and long short-term memory networks,” *Computers and Electronics in Agriculture*, vol. 239, Art. no. 110965, 2025, doi: 10.1016/j.compag.2025.110965.
- [63] S. A. Khosravani Shariati and A. Abbasi, “Machine-learning-based winter wheat yield prediction using multisource data,” *Agricultural Water Management*, vol. 322, Art. no. 109951, 2025, doi: 10.1016/j.agwat.2025.109951.
- [64] M. N. Javed and S. A. Murad, “Crop yield prediction in agriculture: A comprehensive review of machine learning and deep learning approaches, with insights for future research and sustainability,” *Heliyon*, vol. 10, no. 24, Art. no. e40836, 2024, doi: 10.1016/j.heliyon.2024.e40836.
- [65] M. Carbajal, D. Jones, C. Williams, and N. G. Nelson, “In-season yield forecasting using multitemporal remote sensing environmental observations and machine learning: Applications for sweetpotato in North Carolina, USA,” *European Journal of Agronomy*, vol. 172, Art. no. 127818, 2026, doi: 10.1016/j.eja.2025.127818.