Natural Resources for Human Health
eISSN: 2583-1194

Original Research

Received 19 July 2026
Revised 25 Aug 2026
Accepted 13 Sep 2026

Natural Resources for Human Health 2026, Vol. 6 (20s): 150–190

KEYWORDS:
natural language processing, crop growth estimation, crop yield forecasting, agricultural commodity prices, harvest readiness, multimodal fusion, PRISMA

Natural Language Processing-Enabled Crop Growth Estimation and Market-Readiness Forecasting: A PRISMA-Guided Review of Models, Data Fusion Strategies, Validation Practices, and Research Gaps

Mr. Ashish S. Kale[1]*, Mr. Siddharth S. Mandwade[2], Dr. Aruna Varanasi[3]

[1]*Department of Computer Science & Engineering,Vikrant University, Gwalior, Madhya Pradesh, India ORCID: 0009-0009-7475-4108 ashishsk.8872@gmail.com
[2]Department of Computer Science & Engineering, Vikrant University, Gwalior, Madhya Pradesh, India ORCID: 0009-0000-9807-5913 Siddharthraje1988@gmail.com
[3]Department of Computer Science & Engineering, Sreenidhi Institute of Science & Technology, Hyderabad ORCID: 0000-0002-2943-8861 varanasi.aruna2002@gmail.com

ABSTRACT: The estimation of crop growth and the determination of when a crop will be available for the market are typically considered to be two distinct categories of prediction tasks. The purpose of this review is to meticulously compile the findings of fifty studies that have been conducted on agricultural natural language processing (NLP), the prediction of crop growth and yield, the prediction of commodity prices, and the identification of crops that are mature or ready to be harvested. The end evidence base consisted of ten studies on natural language processing and knowing intelligence, twenty research on growth and yield, ten studies on price forecasting, and ten studies on maturity and readiness. Traditional ensembles such as Random Forest and XGBoost maintained their competitiveness when applied to structured regional datasets & samples. The utilization of optimized recurrent networks and transformer versions became increasingly prevalent in the course of price investigations. Object recognition, semantic segmentation, and lightweight vision models were the most commonly utilized in studies pertaining to maturity. The research conducted on natural language processing (NLP) led to improvements in the categorization of agricultural texts, the answering of queries, the extraction of information, the construction of ontologies, and the support of large language models. In contrast, it only occasionally contributed characteristics directly to models that predict the statuses of crops or the timing of market events. The architecture, not the algorithms, is the primary area where there is a lack of evidence: none of the studies that were examined generated a fully tested, natural language processing (NLP)-enabled system that could predict crop growth, harvest readiness, market potential, operational risk, and uncertainty. For the purpose of bridging this gap, a multimodal architecture with five layers is proposed. Among its features are natural language processing that takes into account provenance, calibrated prediction services, decision-level fusion, and feedback from actual farm and market outcomes that are updated over temporal instance sets.

1. INTRODUCTION

A number of technologies, including machine learning, deep learning, remote sensing, sensing from the Internet of Things, and high-frequency market data, have all contributed to the tremendous acceleration in agricultural forecasting. However, a decision regarding how to manage a farm almost never merely poses a single question. The farmer is responsible for determining how the crop is developing, the amount of marketable output that is likely to be produced, whether or not the crop has reached a level of maturity that is commercially viable, what pricing and demand conditions are anticipated, and whether or not the risks associated with postharvest or logistics make it more advantageous to sell the crop now than to wait. The literature that we analyzed provides responses to these issues in a variety of ways that are largely distinct from one another. The weather, the soil, the vegetation, and satellite or unmanned aerial vehicle (UAV) photos are the primary focus of yield studies [11–30]. Previous prices and price fluctuations are taken into consideration in studies that focus on commodities [31–40]. Studies that investigate maturity make use of field photos to locate areas that are ripe and to determine when they are ready to be harvested [41–50].

New ways of acquiring knowledge are made possible via NLP. There are signals that are difficult to express as regular numerical variables that can be found in agricultural advisories, scientific articles, farmer notes, extension reports, market announcements, policy releases, social media chatter, and the news. The classification of crops across languages [1, 2], the answering of questions based on evidence [3, 4], the discovery of regional knowledge [3, 4], the analysis of policy discourse [4, 5], the construction of ontologies [5, 6], the extraction of information [7], and the utilization of language models in larger large-language applications [9, 10] are all examples of areas in which language models have been utilized in recent research. The fact that a lot of the factors that make the market ready, such as pest warnings, procurement announcements, export limits, changes in local demand, transportation challenges, and quality requirements, first appear as text is one of the reasons why these abilities are so crucial.

Therefore, the phrase "NLP-based growth estimation" should not be interpreted to suggest that text is employed in place of physical sensing. Because of their biophysical nature, the greatest techniques to observe how crops grow are by the use of measurements that involve the weather, soil, spectral, phenological, and image measurements. The utility of natural language processing (NLP) is demonstrated when it transforms unstructured contextual evidence into structured, time-stamped features and knowledge constraints that are associated with physical observations. In a market-readiness system that is extremely stringent, this distinction must be maintained: numerical and visual models provide educated guesses about the state of the crop, whereas natural language processing (NLP) draws in information from the outside, explains projections, and assists with decision fusion that can be confirmed.

In this review, there are four different sections. To begin, it provides a PRISMA-guided overview of fifty new studies that were published by the publishing families that were identified. The second part of the analysis examines the model families, data kinds, validation approaches, and restrictions that have been reported from four different study streams that are all connected to one another. In the third place, it identifies the gaps in the links between farming natural language processing (NLP), calculating growth, evaluating maturity, and forecasting the market. The fourth point is that it proposes an integrated design and evaluation plan for a crop market-readiness system. This plan would not only indicate when it is necessary to take action, but it would also indicate the degree of uncertainty, the origin of the information, and the degree to which certain options are likely to be effective.

These inputs determine the sequence in which the narrative of the review is presented. The first step in answering the five research questions is to determine what aspects of studies that are otherwise extremely diverse need to be compared among themselves. Next, the methodology provides an explanation of the selection and evaluation of the evidence. Following the descriptive writing and stream-specific synthesis that are included in the Results section, the evidence is transformed into an integrated market-readiness design by means of the comparative analysis and the proposed framework.

2. REVIEW QUESTIONS AND SCOPE

These five research questions (RQs) served as the basis for the construction of the study. The inquiries progress from a list of models to a determination of how they are intended to collaborate with one another. This is due to the fact that the objective of the scientific endeavor is not simply to identify the learner that is the most accurate within a single dataset; rather, it is to determine whether or not many models can provide support for an operational decision that is logical and can be supported.

Table 1. Research questions guiding the systematic review.

Question	Review focus
RQ1	Which NLP, machine-learning, deep-learning and hybrid model families were used for agricultural knowledge extraction, crop growth/yield estimation, market-price forecasting and harvest-readiness assessment during 2023–2026?
RQ2	Which input modalities, datasets, spatial scales, forecast horizons and evaluation metrics were used?
RQ3	What comparative performance patterns and methodological limitations recur across the evidence base?
RQ4	How has NLP been connected—directly or indirectly—to crop-state and market forecasting?
RQ5	What architecture and validation protocol are required for an integrated crop market-readiness system?

A clear illustration of this analytical development can be seen in Table 1. RQ1 and RQ2 provide a list of the model families, data, and evaluation settings; RQ3 investigates the performance and validity patterns that continue to occur; RQ4 investigates whether natural language processing is truly connected to crop-state and market forecasting; and RQ5 transforms the unresolved integration problem into an architectural and validation question. Considering that these questions have been established, the review process can now be evaluated in relation to a specific evidence purpose.

3. REVIEW METHODOLOGY

3.1 Reporting standard and review design

During the review, the reporting requirements from PRISMA 2020 [51] and the search-transparency guidelines from PRISMA-S [52] were utilized. Because the evidence that was included came from a variety of research that were conducted using a variety of methodologies, such as classification, regression, time-series forecasting, object detection, semantic segmentation, reviews, and applied decision systems, it was not necessary to do a statistical meta-analysis. An structured qualitative synthesis and a descriptive evidence mapping were utilized instead of the traditional method.

The selection process can be repeated when PRISMA is used as the reporting framework, and structured qualitative synthesis keeps classification, regression, detection, and forecasting metrics from being squeezed into a single pooled effect that doesn't work with them. Before deciding whether or not an individual is eligible, the following step in the technique is to ensure that the information sources, the amount of time covered, and the search logic are all entirely clear.

3.2 Information sources and search period

During this search, both general and specific web searches were conducted. These searches looked thru journals such as IEEE Xplore, SpringerLink and Springer Nature, ScienceDirect/Elsevier, and Hindawi journal records, all of which are now maintained by Wiley publications. Therefore, the evidence from 2026 is merely a snapshot of a portion of the year, and it should not be interpreted as a complete representation of the work that was done throughout the entire year in process.

Table 2. Core search blocks used across publisher platforms; syntax was adapted to each interface.

Search block	Representative Boolean structure
NLP and knowledge	(agricultur* OR crop OR farm*) AND (natural language processing OR NLP OR language model OR ChatGPT OR question answering OR named entity recognition OR ontology OR knowledge graph)
Growth and yield	(crop growth OR crop yield OR production) AND (prediction OR estimation OR forecasting) AND (machine learning OR deep learning OR transformer OR remote sensing OR UAV)
Price and market	(agricultural commodity OR crop price OR market price OR demand) AND (forecast* OR predict*) AND (LSTM OR GRU OR transformer OR machine learning)

Search block	Representative Boolean structure
Maturity and readiness	(maturity OR ripeness OR harvest ready OR harvest planning OR quality grading) AND (crop OR fruit OR vegetable) AND (deep learning OR computer vision OR UAV OR hyperspectral)

It is clear from looking at Table 2 that the search method is identical to the four different streams of evidence that were utilized in the synthesis. An agricultural theme phrase is included in each block, in addition to descriptors that are specific to the task and the model. Research on semantics, biology, markets, and maturity can all be located inside the same review boundary if it is conducted in this manner. In light of the fact that this all-encompassing method is intended to prioritize sensitivity during retrieval, it is necessary to establish explicit qualifying criteria in order to restore specificity during screening.

3.3 Eligibility criteria

Table 3. Inclusion and exclusion criteria.

Criterion	Operational definition
Inclusion	Peer-reviewed article or substantive review; publication year 2023–2026; hosted by IEEE, Springer Nature, Elsevier or Hindawi/Wiley; agricultural NLP, growth/yield prediction, price/market forecasting, maturity or harvest readiness; sufficient description of method and results; English-language record.
Exclusion	Published before 2023; non-agricultural application; editorial, patent, thesis, unreviewed manuscript or inaccessible record with inadequate methodological detail; duplicate version; retracted article; crop recommendation without growth, market or language Intelligence relevance.

An article or substantial review that has been reviewed by industry professionals and published between the years 2023 and 2026 and hosted by the IEEE, Springer Nature, Elsevier, or Hindawi/Wiley. In addition to providing sufficient information on the method and the outcomes, the article or review must be about agricultural natural language processing (NLP), growth/yield prediction, price/market forecasting, maturity or harvest readiness information, and more in process.

It is not permitted to publish anything that was published prior to the year 2023, is not associated with agriculture, is an editorial, patent, thesis, unreviewed manuscript, or record that cannot be located or does not contain sufficient methodological information, is a duplicate, was retracted, or makes a crop suggestion that has nothing to do with growth, market, or language intelligence.

This border is defined by Table 3, which retains current studies that have been subjected to peer review and include sufficient methodological and outcome detail. Additionally, records that are not connected to agriculture, are out of date, cannot be accessed, are duplicates, or are methodologically on the edge are removed. These guidelines prevent the text from gravitating toward broad ideas or artificial intelligence research that do not contribute to the development of linguistic intelligence, the market, or growth. A screening sequence similar to the one depicted in Figure 1 is produced whenever the criteria are utilized.

The search audit turned up 144 records, which was the result of the screening and selection process for the study. Prior to the screening of the titles and abstracts, 53 records were removed because they were duplicates, out-of-window, publication-type, or were obviously not relevant. A total of ninety-one records were examined, however only 29 were eliminated since they did not correspond to the selected theme. There were 62 reports that could be located and read with the use of full text or publisher records that were sufficiently detailed. At this point, twelve were rejected: five because the scope did not match, three because the source or publication type did not match, two because there was not enough methodological or outcome detail, and two because the publications had been retracted. They were rejected for those reasons. Within the final synthesis, fifty different studies were included.

3.4 Screening and study selection

The search audit identified 144 records. Fifty-three duplicate, out-of-window, publication-type or obviously irrelevant records were removed before title/abstract screening. Ninety-one records were screened; 29 were excluded for topical mismatch. Sixty-two reports were sought and assessed through full text or sufficiently detailed publisher records. Twelve were excluded

at this stage: five for scope mismatch, three for source/publication-type mismatch, two for insufficient methodological or result detail and two because they were retracted. The final synthesis included 50 studies.

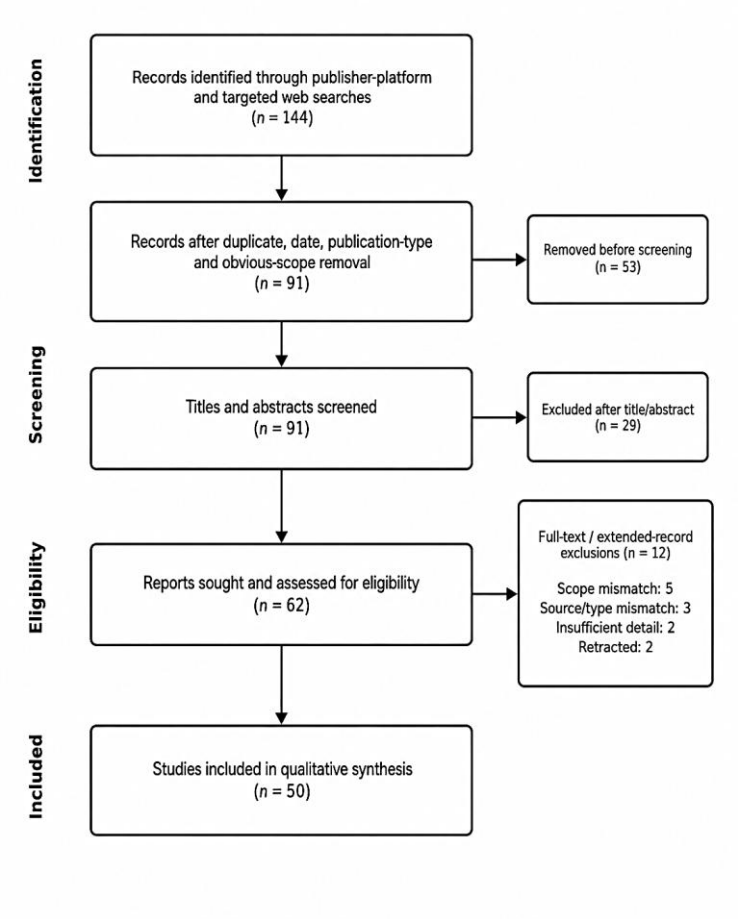


Figure 1. PRISMA-style flow of evidence identification, screening, eligibility assessment and inclusion.

Forty-four percent of the total 144 records that were discovered were retained by the screening procedure. Following the elimination of duplicates, dates, publishing kinds, and records that had a distinct scope, a total of 91 records were able to advance to the title/abstract screening step. Additionally, 62 recordings were able to advance to the eligibility assessment stage, and lastly, fifty records were successfully included. From the time of the eligibility review until the time of the synthesis, 80.6% of the papers were retained, which indicates that the majority of the articles were rejected at an earlier stage in the application process. It was during the full-text stage that the majority of cases involving retractions, inadequate information, scope, and source type were decided to be rejected in this.

As soon as the final corpus has been established, the review can proceed from the manner in which the studies were selected to the evaluation of the quality and congruence of the evidence that was retained. Therefore, in order to code the same fifty studies, a similar extraction strategy and reporting-quality checklist were utilized. This was done prior to the interpretation of cross-study trends.

3.5 Data extraction and quality appraisal

For the purpose of the review, the publication year of the study, the publisher platform, the research job, the crop or commodity, the data modality, the model family, the most important outcome recorded, and the primary limitation were removed. The quality was evaluated based on a list of eight factors, which included a clear statement of the goal, the source of the data, a description of the feature, a comparison to the baseline, the design of the validation, the utilization of the appropriate metrics, safeguards against leakage, and the reporting of reproducibility or limitations. The information that was provided was used to assign a score of either 0 or 1 to each item. A score of 7–8 points indicated that the reporting quality

was high, a score of 5–6 points indicated that it was moderate, and a score of 0–4 points indicated that it was limited. The purpose of the score was not to exclude a study that had already been received approval; rather, it was used to determine how strong the evidence was.

During this stage of evaluation, the link between the mechanics of review and the substantive synthesis is established. Where the majority of the activity is located may be determined by looking at the number of publications, while the extracted modalities, models, metrics, and limits can be used to determine what that activity contributes. As a result, Section 4 begins with the manner in which the proof base is assembled, and then progressively moves on to the more challenging subject of the cross-stream integration process.

4. RESULTS

4.1 Publication profile

In the years following the year 2023, the fifty-study set expands at a rapid pace due to the rapid utilization of transformers, agricultural large language models, multimodal remote sensing, and optimal time-series structures. Due to the fact that the search was completed on August 9, 2026, rather than at the end of the year, it is considered to be less in 2026. It was Elsevier and Springer that provided the majority of the evidence. While Hindawi/Wiley supplied concentrated yield and maturity-classification applications, the Institute of Electrical and Electronics Engineers (IEEE) contributed a number of significant crop yield and price studies. Table 4 provides a breakdown of this profile according to the study stream, and Figures 2 and 3 illustrate the annual and publisher-level concentrations, respectively.

Table 4. Temporal distribution of the included evidence by research stream.

Year	NLP / knowledge	Growth / yield	Price / market	Maturity / readiness	Total
2023	2	6	0	0	8
2024	4	8	2	3	17
2025	4	6	4	5	19
2026*	0	0	4	2	6

It is made abundantly obvious in Table 4 that the increase in the number of publications is not the result of a single stream by itself. It is most typical to conduct growth and yield studies between the years 2023 and 2024, while the year 2025 has the most cross-stream coverage. The 2026 subset is intended to be incomplete, and its primary focus is on developing an understanding of market price and maturity.

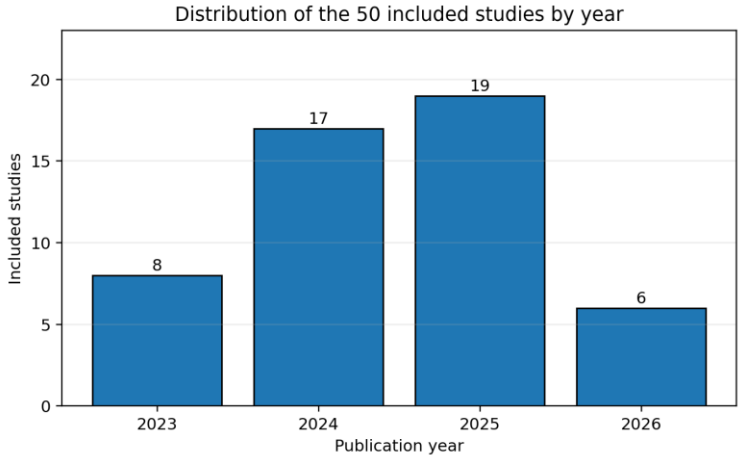


Figure 2. Publication-year distribution of the included studies.

The greatest number of studies were published during the years 2024 and 2025, with 17 thirds of all studies being published in 2024 and 19 thirds of all studies being published in 2025. In the year 2023, just eight (16%) emerged. Due to the fact that the

six studies that were recorded for the year 2026 only cover the time period up to August 9th, it is not appropriate to use them as proof of a decrease on an annual basis. The pattern, on the other hand, indicates that the body of evidence will expand rapidly after the year 2023, particularly in fields such as multimodal yield modeling, transformers, and market forecasting.

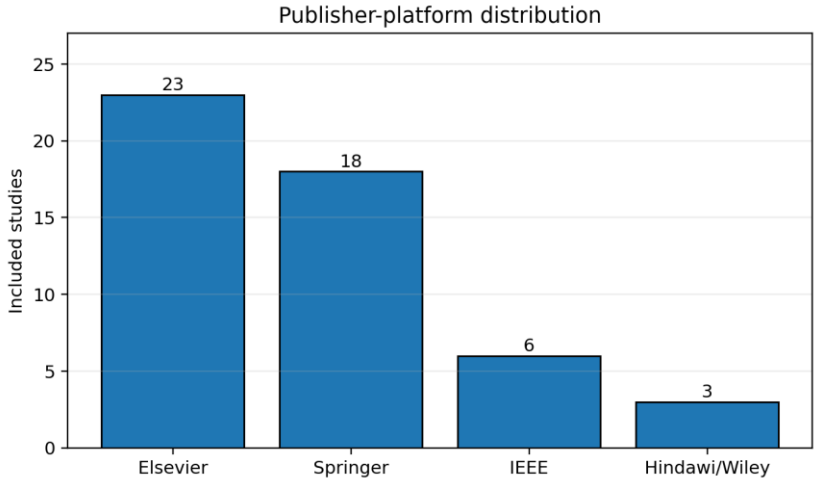


Figure 3. Distribution of included studies by publisher platform.

There is a significant concentration of publishers, however they are not exclusive. Elsevier is responsible for 23 of the total research, which accounts for 46% of the total, while Springer is responsible for 18 of the total studies. Individually, they are responsible for 82% of the evidence that was included in the analysis. On the other hand, Hindawi/Wiley provides six studies, which accounts for twelve percent of the total. The significance of this distribution lies in the fact that it is essential for comprehending generalizability. This is due to the fact that variations in publisher concentration can be connected to various kinds of applications, datasets, and reporting patterns. Figure 6 better illustrates this pattern than any other.

When taken together, Table 4 and Figures 2–3 explain when the body of literature expanded and where the majority of the evidence is focused. However, they have not yet demonstrated whether or not the four different lines of inquiry are genuinely coming together. We are going to investigate that issue in the following section by dividing the corpus into categories based on the task function and the model family.

4.2 Task distribution and model taxonomy

With ten studies on natural language processing and knowing intelligence, twenty studies on crop growth or yield, ten studies on commodity pricing, and ten studies on maturity or readiness, there are a total of twenty studies. The integration gap is visible due to the fact that the samples were selected to be distributed evenly over the board. However, natural language processing (NLP) research typically do not result in reliable agricultural forecasts, despite the fact that they have a lot to say about semantic representation and providing people with access to knowledge. Growth models are effective at making use of distant sensing and environmental aspects; nevertheless, they typically do not take into account how the market is performing by themselves. When it comes to maturity systems, the data typically concludes with a class name, although in pricing models, the crop is frequently thot of as an abstract item that has a time series. Figure 4 illustrates the organizational structure of the most prevalent ways that are used in various task families. The amount of each stream's contribution to operations is illustrated in Figure 5, along with the amount of integration that each stream is lacking.

Table 5. Functional contribution of the four evidence streams to crop market-readiness estimation.

Evidence stream	Studies	Dominant inputs	Primary analytical output	Role in market readiness	Main integration deficit
NLP / knowledge intelligence	10	Advisories, scientific text, farmer questions, policy and social	Entities, relations, topics, retrieved evidence and structured	Context, event detection, provenance and explanation	Rarely connected to calibrated crop-state or market-timing forecasts

		discourse	knowledge		
Crop growth / yield	20	Weather, soil, production records, satellite/UAV imagery and vegetation indices	Yield, biomass, phenological or spatial production estimates	Expected marketable supply and production confidence	Limited text-derived context; external transfer and leakage-resistant validation remain uneven
Price / market	10	Historical prices, arrivals, volatility and selected exogenous variables	Price level, volatility and multi-horizon market forecasts	Expected market opportunity and downside exposure	Policy, logistics, sentiment and crop-state information are weakly represented
Maturity / readiness	10	RGB, multispectral, Sentinel/UAV imagery and crop morphology	Ripeness class, harvestable area, quality grade or harvest plan	Biological harvestability and quality timing	Usually stops before price, logistics, uncertainty and selling-utility fusion

The reasons why the review cannot be reduced to a straightforward model rating exercise are presented in Table 5. Due to the fact that each stream estimates a separate decision variable, the evidence that they present cannot be substituted for one another but rather contributes to the information that is previously known. Because there is no standard interface that can align these outputs by crop, location, and time while maintaining uncertainty and provenance, the most significant challenge that research faces is the lack of such an interface for this process. The fact that a high level of accuracy in a single stream is not sufficient to demonstrate that the market is ready on its own is also explained by this.

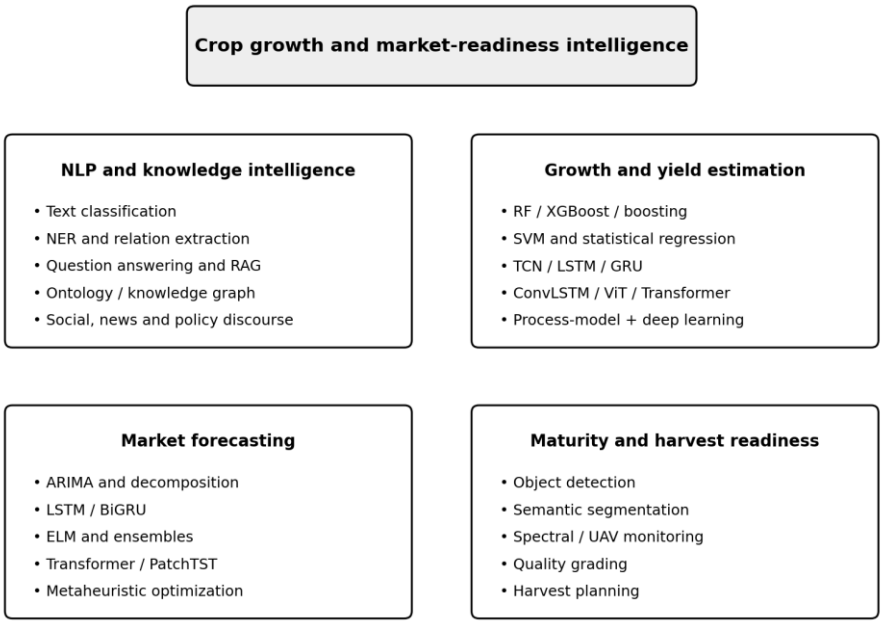


Figure 4. Taxonomy of model families and application tasks in the reviewed evidence.

Th There is a distinct division of functions that may be seen in the classification. The majority of the time, natural language

processing (NLP) techniques make changes to unstructured data, growth models speculate on the biophysical condition, market models speculate on how prices will change, and maturity models determine how to harvest visually. The literature study does not contain any arrows that would connect these results to a well Informed selling choice, despite the fact that each of the methodologies has been improved sufficiently to yield valuable intermediate outcomes. Based on this separation, the cross-stream analyzes that are shown in Figures 5–8 are performed.

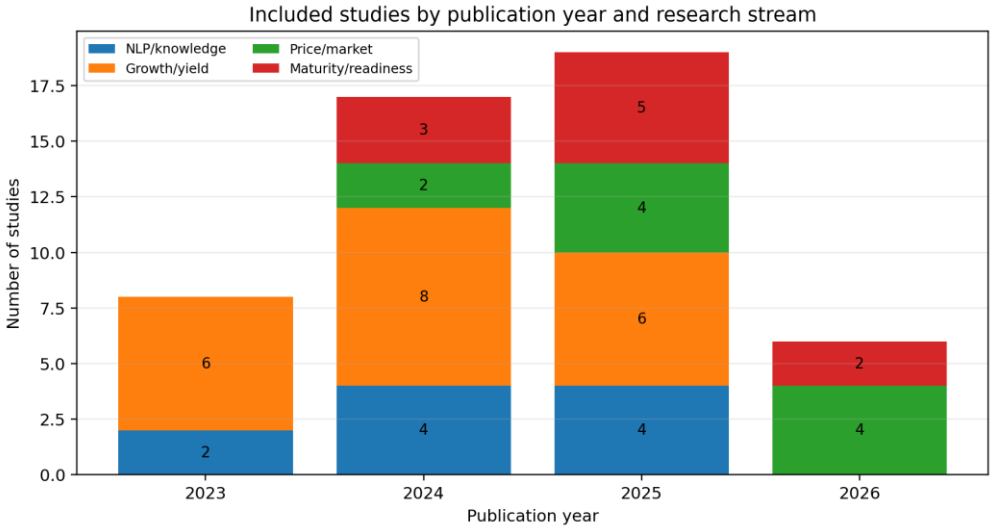


Figure 5. Year-by-task composition of the 50 included studies.

A further piece of information regarding the composition is included in Figure 5, which is a publication trend. The topics of growth or yield are the focus of six of the eight studies conducted in 2023 and eight of the seventeen research conducted in 2024. Thru the year 2025, the evidence is dispersed throughout all four streams in a more equitable manner. Because of this development, it appears that the literature is moving away from concentrating primarily on production forecasts and instead shifting its attention to timing and market considerations further downstream, despite the fact that the streams are still strictly distinct.

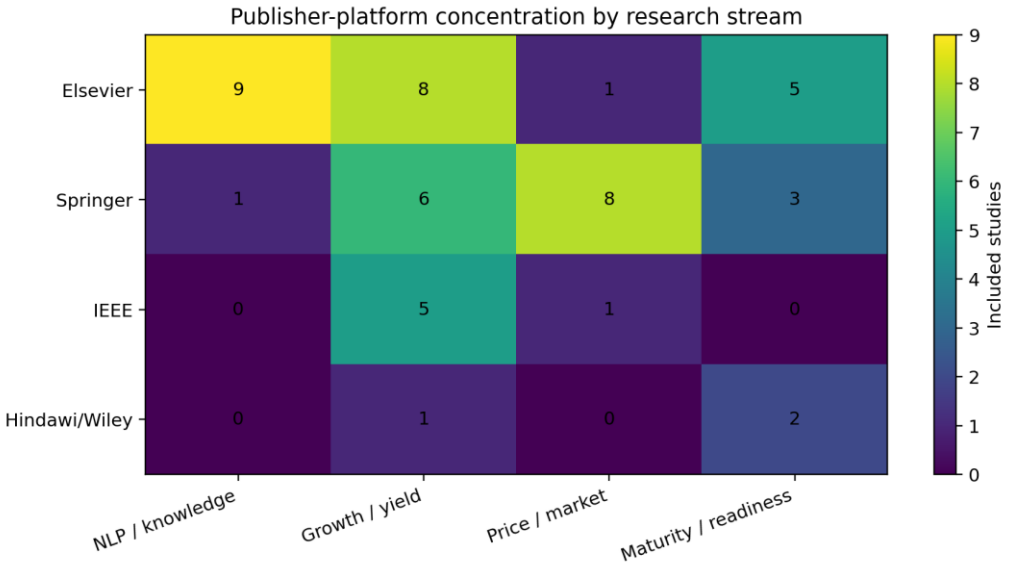


Figure 6. Publisher-platform distribution across the four research streams.

It is demonstrated in Figure 6 that the source concentration shifts depending on the task at hand. On the other hand, Springer is responsible for eight of the 10 market-price studies, while Elsevier is responsible for nine of the ten

NLP/knowledge research represented in the group. It is clear that the IEEE places a significant amount of importance on this field because five of the six publications that it includes are in the growth/yield prediction stream. The tendency does not indicate that the strategy utilized is either beneficial or detrimental for a particular publisher; however, it does indicate that findings derived at the stream level may incorporate platform-specific coverage and disciplinary emphases.

It is thru these patterns of distribution that we are able to obtain the necessary background information to comprehend the situation at the model level. As a result of the fact that natural language processing (NLP) and knowledge studies offer the semantic and event layer, which is not adequately represented in conventional crop forecasting pipelines, the first stream considered is this one. After that, the evidence associated with the crop's growth or yield, pricing, and maturity are examined in the same order that they have an impact on the ultimate judgment on whether or not the crop is ready to be sold.

4.3 Agricultural NLP and knowledge Intelligence models

Starting with straightforward text mining, agricultural natural language processing has progressed to include domain-aware transformer encoders, retrieval systems, ontologies, and generative models today. In their investigation of cross-lingual classification, ChatAgri demonstrated that prompt-based generative classification is not only flexible but also unstable [1]. Providing farmers with answers that were founded on evidence was accomplished by AgAsk thru the use of document retrieval and question asking [2]. The information on policies and plans was acquired thru the use of regional text analytics and social discourse analysis [3, 4], while ontology and graph-based work were utilized to arrange thoughts and connections in the agricultural sector [5].

According to one of the most significant discoveries, general-purpose generative models are not necessarily superior when it comes to organized agricultural extraction. When it came to agricultural named object recognition and relation extraction, ChatGPT did not perform as well as fine-tuned encoder models that included data addition [7]. When it comes to getting ready for the market, this discovery is significant since it is essential to consistently extract crop, location, disease, quantity, date, price, policy, and logistics entities from text in order to transform it into characteristics that are ready to be used in models. Regarding the importance of retrieval grounding, topic adaptability, multilingual coverage, and origin, reviews of answering agricultural questions and LLMs additionally emphasize the significance of these aspects [8–10].

In spite of all of this expansion, natural language processing was rarely evaluated based on how well it worked on farms or in marketplaces. Text-task metrics, usability, and possibilities at the architectural level were discussed in the majority of the research. There was not a single one of the ten natural language processing research that accurately forecasted the phenological stage, yield, maturity, future price, or the optimal time to sell together. A knowledge and feature interface is the final stage that is required. This interface should enable numerical forecasting services to view extracted events together with timestamps, locations, confidence scores, and source provenance.

Table 6. NLP capabilities and their translation into market-readiness variables.

NLP capability	Representative studies	Typical model form	Market-readiness variable that can be structured	Critical evaluation requirement
Cross-lingual agricultural text classification	[1]	Prompted generative model and supervised classifiers	Crop/topic labels, event categories and language-normalized alerts	Prompt stability, cross-language consistency and domain grounding
Evidence retrieval and question answering	[2], [8]	Retrieval, reranking, extractive or neural QA	Evidence-linked agronomic constraints and advisory facts	Source precision, answer attribution and corpus coverage
Regional discourse and policy analytics	[3], [4]	Topic, sentiment and text-mining pipelines	Policy, demand, risk and stakeholder-sentiment signals	Temporal/geographic validity and social-media bias control

Ontology, NER and relation extraction	[5], [7]	Domain encoders, relation extraction, ontology/GNN methods	Crop, location, date, disease, price, quantity and logistics relations	Entity/relation F1, schema coverage, provenance and annotation cost
Generative and planning intelligence	[6], [9], [10]	LLMs, retrieval augmentation and decision-support synthesis	Contextual summaries, planning constraints and human-facing explanations	Hallucination control, governance, multilingual equity and downstream utility

An example of how the NLP literature is transformed into a methodology for prediction is presented in Table 6. Once upon a time, it was a list of text jobs. It is recommended that you make extensive use of named entities and linkages since these features enable you to geocode and timestamped the entities before they are combined with market and sensor series. In order to uncover proof and provide an explanation for it, it is preferable to employ the retrieval and generating parts rather than to make a random number forecast. It is for this reason that the subsequent stage of the evaluation process must determine whether the extracted text characteristics improve the system in terms of crop-state, price-risk, or decision-utility, in addition to improving its accuracy in terms of text-task accuracy.

When attempting to determine the state of the crop, the interface viewpoint that is presented in Table 6 is the most essential aspect to comprehend initially. After the textual events have been normalized into components that are based on crop, location, time, trust, and provenance, it is possible to link them with structured and remotely sensed data that is utilized by growth models. That being said, the following section will discuss how to make an educated guess regarding the biophysical production state, which is the basis for any decision that is taken later on in the market scenarios.

4.4 Crop growth and yield estimation

There is no doubt that the tree groups that are most effective for structured regional datasets are the ones that are the best. Random Forest performed the best in a number of tests, including [11], [23], and [29]. In the cases where mixed feature types and nonlinear interactions were the most critical factors, we found that Optimized XGBoost and other boosting algorithms performed exceptionally well [19], [20], and [29]. In addition to being able to perform feature Importance or SHAP analysis, these models require less data to execute, and it is simple to observe how different items interact with one another. One of their shortcomings is that they are unable to accurately display time; they do not learn crop development in a linear fashion unless they are provided with lagged features and phenological summaries to fulfilll this requirement.

It has been resolved thru the application of deep learning in both time and space. Once upon a time, the TCN was able to make educated guesses about rice harvests by learning patterns that appear annually from a variety of bands [13]. ConvLSTM–ViT coupled minor changes in space and time with a focus on the entire globe [24], and multimodal ensemble models incorporated input from weather, crops, and remote sensing [27]. Both of these models were developed by ConvLSTM in 2015. At first glance, it appears that hybrid process-data models will be the most successful. The Nash–Sutcliffe efficiency was determined to be between 0.954 to 0.996 [17] by Jeong et al., who utilized crop modeling, remote sensing, and deep learning in conjunction with one another. In order to get estimates that are based on a smaller amount of data, we can make use of what we know about the functioning of crops and hybrids.

There is still a significant gap between the amount of transferability and interpretability. As Paudel et al. shown [12], it is essential to provide explanations, particularly in situations where the weather is unusual. Thru the application of the SHAP approach, Khodjaev and colleagues were able to determine that the crop height and red-edge indices were important. In addition, they said that the improved RF/XGBoost algorithm succeeded in reaching an R2 value of 0.89 [29]. Nevertheless, a great number of studies utilized random train–test splits on time-ordered data, which would have resulted in more precise estimations in the event that close seasons, places, or image tiles were blended together. There were a greater number of tests that were conducted throughout the same dataset as opposed to tests that were conducted across different seasons, areas, and extreme occurrences.

Table 7. Comparative growth/yield model families and their deployment implications.

Growth/yield model family	Representative studies	Best data regime	Analytical strength	Market-readiness contribution	Dominant deployment risk
Tree ensembles and boosting	[11], [19]–[23], [29]	Structured regional weather, soil and production variables	Nonlinear interactions, strong small/medium-data baseline, SHAP compatibility	Yield/supply estimate and feature attribution	Regional overfitting and weak native sequence representation
Temporal convolution and recurrent models	[13], [25]	Seasonal multispectral/weather sequences	Learns stage-dependent temporal patterns	Dynamic yield trajectory and early forecast update	Long-horizon drift and sensitivity to sequence construction
ConvLSTM–ViT / attention models	[24], [25]	Dense spatiotemporal remote-sensing data	Combines local dynamics with global attention	Spatially resolved production forecast	Compute demand, dense imagery dependence and calibration
Process-data hybrid modeling	[17]	Crop-model states plus remote sensing and environmental data	Physical plausibility with learned residual structure	Biophysically constrained yield confidence	Site-specific process parameters and calibration burden
Multimodal deep ensembles	[27]	SAR, optical imagery and meteorology	Complementary sensor fusion and ensemble robustness	Marketable supply estimate under heterogeneous sensing	Synchronization, missingness and independent replication

In Table 7, it is demonstrated that the data regime, and not the veracity of the headline, should be the one to determine which model to apply. As long as the inputs are limited and structured, tree groups will continue to be robust. In situations where sequence information is accurate, temporal networks become more defendable. In addition, when dense distant sensing streams are accessible, attention and multimodal structures are able to make sense of how difficult they are. A further layer of protection against expansion that is physically impossible is added by hybrids of data and processes. In order to make something ready for the market, you need more than just return on investment. In addition to this, you need to have a solid understanding of the degree of uncertainty surrounding the marketable quantity as well as the quantity that is currently available.

However, having faith in the progress of the crop is not sufficient to determine when it is appropriate to sell it in process. It is necessary to make a comparison between the anticipated supply path and the manner in which prices are projected to move, the dynamics of the market, and the degree of volatility. During the course of Section 4.5, the topic shifts from biological production to economic timestamp sets.

4.5 Agricultural price and market forecasting

When it came to predicting prices in the past, a straightforward autoregressive baseline was utilized. The employment of transformers, recurrent networks, and metaheuristic optimization are all examples of techniques that are frequently employed in modern times. Over the course of sixty days, Sari and his colleagues [31] conducted a study in which they examined eleven distinct goods under consideration. The researchers discovered that an extreme learning machine that had been updated with

genetic algorithms performed significantly better than GA-LSTM and Auto-ARIMA in terms of overall performance. Thru the utilization of stacked LSTM and vector autoregression, Meena and Chaitra were able to determine the amount of money that each market would be willing to spend for ragi [32]. Finding anomalies [33], forecasting volatility [34], metaheuristic transformer tuning [35], and bidirectional recurrent architectures [36, 38] were some of the topics that were investigated in the study that followed this one. They were also among the topics that were investigated. This research found that patch-based transformers and decomposition were effective methods for determining the appropriate price for multi-horizon potatoes [40].

From a technical standpoint, the general direction of the model is pretty clear: tiny changes are handled by local convolution or decomposition, time dependency is managed by recurrent or attention mechanisms, and hyperparameters are controlled using optimization approaches. All of these mechanisms are used to handle the model. The rate at which one can acquire new knowledge is not as rapid as it has the potential to be. These, coupled with historical prices, lags, and a few figures from sources outside of the study, are utilized in the majority of studies that investigate prices. The researchers Vairo et al. examined the weather in great detail and demonstrated that the value of the weather varies depending on the model and the sky [39]. Many times, textual signals such as rules on buying and selling and price changes that occur before they can be noticed in a lagged series were not taken into account in the process. This was the case in a significant number of instances. The demand for events, the development of diseases, difficulties with transportation, and how people feel about potential market scenarios are some of the factors that fall under this category.

Table 8. Agricultural price-forecasting architectures and their relevance to market-readiness decisions.

Forecasting strategy	Representative studies	Core mechanism	Useful market-readiness signal	Primary unresolved issue
Optimized classical / shallow learning	[31]	GA-optimized ELM with ARIMA and LSTM comparison	Short-to-medium horizon price level	Limited textual and structural-shock representation
VAR + stacked recurrent forecasting	[32]	Cross-series autoregression followed by stacked LSTM	Market-wise price trajectory	Commodity/state specificity and shock sensitivity
Transformer + anomaly-aware recurrent pipeline	[33]	Attention-based forecasting with LSTM anomaly detection	Price forecast plus abnormal-market warning	Weak coupling to maturity, yield uncertainty and logistics
Volatility and bidirectional recurrent models	[34], [36], [38]	Hybrid volatility modeling, LSTM/GRU/BiGRU and attention	Volatility exposure, trend direction and nonlinear dependence	Exogenous context and causal driver transparency
Metaheuristic / CNN / patch transformers	[35], [37], [40]	Transformer tuning, local convolution and patch-based long-horizon modeling	Multi-horizon opportunity window	Optimization overhead, regime transfer and crop-specific validation
Weather-conditioned price forecasting	[39]	XGBoost and LSTM with/without weather	Conditional price sensitivity to weather	Weather does not encode policy, trade, storage or demand shocks

When looking at Table 8, it is simple to observe how the architecture shifts from better classical learners to recurrent, attention, and patch-transformer models. It is not a lack of time modeling that is the source of the operational deficiency; rather, it is incomplete information coming from outside the system. On the other hand, historical pricing can be used to

demonstrate how persistent something is; but, they are not able to accurately foresee policy announcements, transportation challenges, disease-related supply shocks, or changes in procurement before they demonstrate themselves in numbers. Rather than serving as alternatives for time-series models, natural language processing (NLP)-derived events are most beneficial when they are used as provenance-related features that are external to the system.

It is only possible to employ price chance if it is possible to strike a balance between the fact that the crop can be harvested and the fact that its quality can be maintained. This reliance establishes a connection between market forecasts and the subsequent section, which discusses maturity and anticipations. Visual and remote sensing models each provide a description of the biological window that allows for the possibility of economic optimization in this section's discussions.

4.6 Maturity, quality and harvest readiness

In the field of maturity research, computer vision is the primary instrument utilized. Using rules for locating, separating, and measuring the size of brinjal, we were able to determine the degree to which it had reached its ripeness [41]. In order to map sugarcane that was ready to be harvested, we utilized Sentinel-2 pictures [42], and in order to determine the amount of citrus that should be grown, we utilized a workflow that involved a UAV and the web [43]. [44, 47, 48] Research conducted on blueberries, dragon fruits, and oil palms utilized object detection or fused deep features in order to locate the fruit and evaluate its quality or freshness when it was discovered. A wide variety of cognitive, spectral, and visual signals of maturity (45, 46) have been confirmed by review studies conducted on cashews and fresh fruit bunches by researchers.

It is possible for the stated accuracy to be quite high in controlled samples. Nikam et al. discovered that Random Forest had an accuracy rate of 99.99% when it came to determining the quality of dragon fruit, while AdaBoost had a rate of 99.69% when it came to determining when the fruit was ready to be consumed [47]. It is not a good idea to apply these numbers as a general guideline without taking into consideration other factors such as the field illumination, the diversity of the background, the cultivar difference, the camera distance, and the class balance. Not only does it appear as tho the fruit is ready to be picked, but there is more to it than that. The research on broccoli then moved on to the planning of the harvest [49], and light six-stage cherry models assisted with the classification of resources that were limited [50]. Nevertheless, the objective was not to investigate the possibility of price or the danger of logistics.

Table 9. Maturity and harvest-readiness methods with field-deployment constraints.

Readiness method	Representative studies	Observed signal	Operational output	Field-deployment constraint
Detection + segmentation + morphology	[41]	Fruit/object appearance, segmentation and size	Maturity class or harvestability rule	Shadow, occlusion, cultivar morphology and rule sensitivity
Remote-sensing semantic mapping	[42]	Sentinel-2 spectral/spatial patterns	Harvest-ready area map	Clouds, pixel labels and regional crop-calendar dependence
UAV counting and harvest planning	[43], [49]	High-resolution aerial imagery	Fruit count, yield support or harvest plan	Occlusion, counting error, crop specificity and market coupling
Deep object detection	[44], [48]	Field/UAV imagery of fruit bunches	Localized maturity or ripe-object detection	Small objects, lighting, canopy occlusion and domain shift
Feature-fusion grading / lightweight CNN	[47], [50]	Fruit appearance and learned deep features	Quality grade or multi-stage maturity class	Controlled Image bias and incomplete representation of internal quality
Review-based maturity evidence	[45], [46]	Visible, spectral and learned indicators	Method taxonomy and comparative evidence	Heterogeneous datasets and absence of common field benchmarks

A distinction can be made between biological maturity and economic readiness, as demonstrated in Table 9 of this text. If internal quality, postharvest deterioration, labor supply, price opportunities, and logistics are not taken into consideration, vision systems are very good at estimating how ripe something is or how much of it can be harvested by this. However, this information is incomplete if these factors are not taken into account. A calibrated harvestability probability or grade distribution should be provided to the decision layer by the readiness subsystem rather than a single harvest suggestion being made.

It has been determined that each of the four lines has been examined on its own at this stage. Although they are complimentary in nature, their outputs are derived from distinct methodologies and are evaluated according to distinct standards. It is necessary for us to be aware of the specific validation risks that are associated with each approach in order for fusion to make sense. Section 4.7 makes it abundantly evident that heterogeneity exists before the review has the opportunity to measure the remaining end-to-end integration gap.

4.7 Data modalities and validation practices

Table 10. Data modalities and validation concerns - Part I: text, structured farm/environment and remote-sensing inputs

Modality	Typical inputs	Common metrics	Recurring validation concern
Text and documents	Scientific articles, advisories, social media, farmer questions, policy discourse	Classification F1, retrieval rank, answer relevance, NER/RE F1	Weak linkage to agronomic and economic outcomes
Structured farm/environment	Soil, weather, management, historical production	R2, RMSE, MAE, accuracy	Random splits, local datasets, uncertain transfer
Remote sensing/UAV	Multispectral, RGB, vegetation indices, crop height	R2/RMSE, IoU, mAP, accuracy	Clouds, occlusion, spatial leakage, acquisition cost

These three methods of describing information and crop state are broken down into their respective categories in Part I. Text sources give early contextual clues but need to be checked for accuracy later on; structured farm/environment variables are still good for interpretable regression but can be affected by local data and random-split bias; remote sensing increases the amount of space covered but comes with risks of cloud, occlusion, resolution, and spatial leakages.

Table 11. Data modalities and validation concerns - Part II: market time series and multimodal fusion.

Modality	Typical inputs	Common metrics	Recurring validation concern
Market time series	Prices, arrivals, volatility, selected weather/exogenous variables	RMSE, MAE, MAPE, R2, directional accuracy	Regime change, nonstationarity, limited text and policy signals
Multimodal fusion	Weather + imagery + soil + crop models; occasionally text	Mixed predictive metrics	Synchronization, missingness, calibration and provenance

Part II focuses on modalities closest to commercial timing. Although market time series are able to display fluctuations in pricing and arrivals, they are not always stationary and may experience regime adjustments. On the other hand, multimodal fusion is capable of combining weather, photos, soil, crop models, and even text in some cases; nevertheless, it is more difficult to synchronize, fill in missing data, and calibrate information. When the two groups are combined, the second group can serve as a natural foundation for a market-readiness system; however, this is only the case if the mode keeps the origin

and uncertainty that are unique to the manner.

It is not possible to directly rank works due to the existence of differential measures. The R2 and RMSE metrics are the most important ones for continuous yield and price prediction. For classification, accuracy and F1 are the most important factors, whereas for detection and segmentation, mAP and IoU are the most important factors. The number of research that discussed calibration error, prediction intervals, or decision utility was relatively low. A system that is prepared for the market, however, must provide an answer to a question that takes risk into consideration. It is possible that one model will make a different harvest choice than the other model with the same root mean square error (RMSE) if it is overconfident in its own capabilities near a critical price or maturity criterion sets.

When you look at an error number, it only makes sense when you consider it in the context of the data-gathering process, split design, forecast horizon, and decision threshold. Therefore, the comparison of modes in Tables 10 and 11 modifies the way you should understand the headline performance. Now that these discrepancies in validation have been discovered, the review can proceed to its primary systems inquiry, which is the amount of the various data that has been included into a single ready pipeline.

4.8 Evidence concentration and missing integration

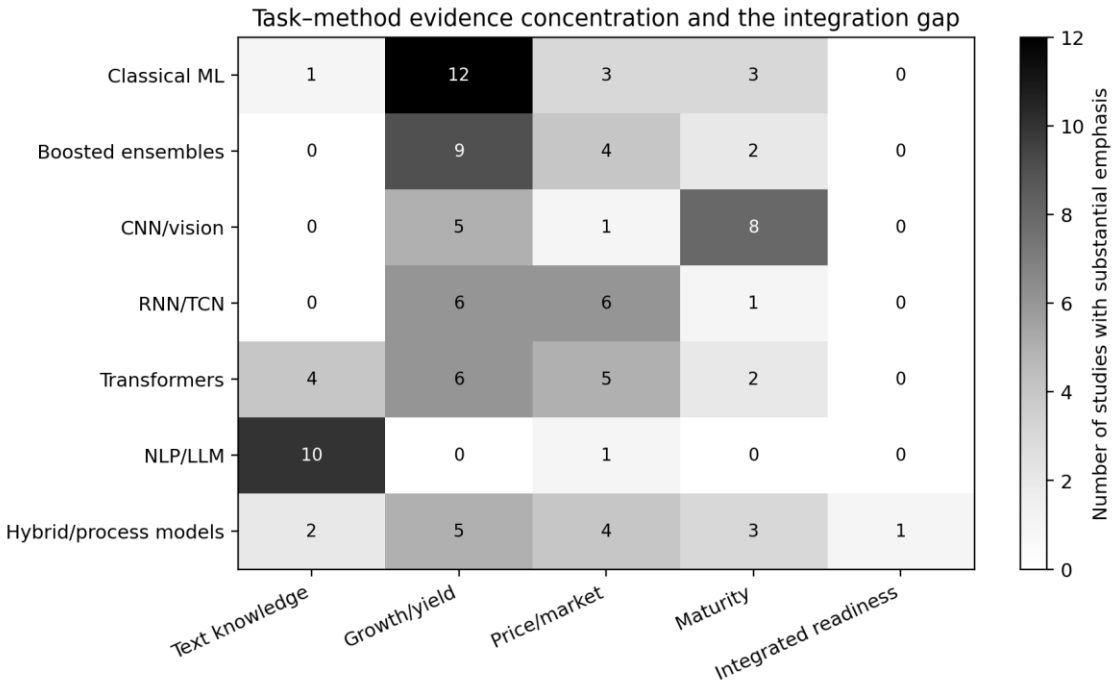


Figure 7. Evidence concentration across model families and tasks. The integrated-readiness column exposes the principal research gap.

Within each job, the heat map displays a great deal of methodological detail; yet, there is not a great deal of evidence from the beginning to the conclusion. In the integrated-readiness column, there is a single count that indicates that there has been some progress made toward harvest planning; nevertheless, the system is not yet fully enabled with natural language processing. None of the articles that were considered for inclusion made use of text-derived events, crop-state estimation, maturity assessment, multi-horizon market forecasting, calibrated uncertainty, and an action suggestion that was based on economics all at the same timestamp sets.

One of the most important aspects of the review is Figure 7. As seen in sections 4.3–4.7, there are already component models that are capable of doing the tasks that they are intended to perform. Open interfaces, calibration, and decision fusion are the most important aspects that are lacking. Comparing model families is done in Section 5, not with the intention of selecting a clear victor, but rather to identify the classes that are most capable of performing each duty in a modular market-readiness

system.

5. COMPARATIVE MODEL ANALYSIS

Table 12. Comparative model-family interpretation - Part I: classical, temporal and vision learners.

Model family	Best-suited setting	Principal strength	Principal weakness	Illustrative studies
RF / XGBoost / boosting	Structured and moderately sized datasets; mixed numeric features	Strong nonlinear baseline, efficient, explainable with SHAP	Weak native sequence modeling; regional overfitting	[11], [19]–[23], [29]
SVM / sparse regression	Small datasets and high-dimensional but stable features	Data-efficient and interpretable	Scaling and temporal limitations	[15], [16], [22], [30]
LSTM / GRU / TCN	Seasonal sequences and price histories	Learns temporal dependence	Long-horizon drift, tuning sensitivity, limited exogenous context	[13], [31], [32], [34], [36]
CNN / object detection	Fruit, canopy and field imagery	Spatial localization and maturity recognition	Occlusion, illumination and domain shift	[41]–[50]

Model classes that are still required as practical baselines are discussed in the first part of this article. When dealing with structured data, RF/XGBoost and sparse learners perform exceptionally well; LSTM/GRU/TCN designs are supported by real-time dependencies; CNN and detection models are required for identifying items in space and determining the level of maturity they possess. The comparison demonstrates that the shape of the data has a significant impact on the accuracy of the forecast. In spite of this, transfer, drift, and domain shift continue to be common ways in which the forecast can be incorrect for this process.

Table 13. Comparative model-family interpretation - Part II: transformers, domain NLP and hybrid process-data models.

Model family	Best-suited setting	Principal strength	Principal weakness	Illustrative studies
Transformer / ViT / PatchTST	Long sequences, multimodal imagery and market horizons	Global attention and scalable representation	Data/compute demand and calibration concerns	[24], [25], [35], [37], [40]
Domain NLP / LLM / RAG	Unstructured advisories, science, news and farmer text	Semantic extraction, multilingual access, explanation	Hallucination, provenance and downstream validation	[1]–[10]
Hybrid process-data models	Biophysical prediction under data constraints	Physical plausibility and stronger extrapolation potential	Calibration complexity and site-specific parameters	[17], [27]

It is in Part II that you will find the model families that are the most significant for integration. Domain Natural Language Processing and Language Learning (LLM) systems alter unstructured evidence, transformers add long-range representation to picture and market sequences, and process-data hybrids offer physical plausibility to crop forecasting. These are the things that make them the weakest.

There is not a single model family that dominates every situations. The utilization of classical ensembles as baselines ought to be mandated due to the fact that they are difficult to surpass when applied to little structured datasets. In situations when the raw data consists of true sequential information and the evaluation maintains the order of occurrences, temporal deep models

make perfect sense. The findings of vision models need to be calibrated and evaluated for quality before they can be used in business. Vision models are vital for maturity and field mapping, but before they can be utilized in business at all. The natural language processing and language learning components should not directly replace the physical evidence; rather, they should offer structure to the events, limits, explanations, and the process of collecting proof. Figure 8 is an addition to Tables 12 and 13, as it illustrates the frequency with which the primary model families appear in each of the fifty investigations for this process.

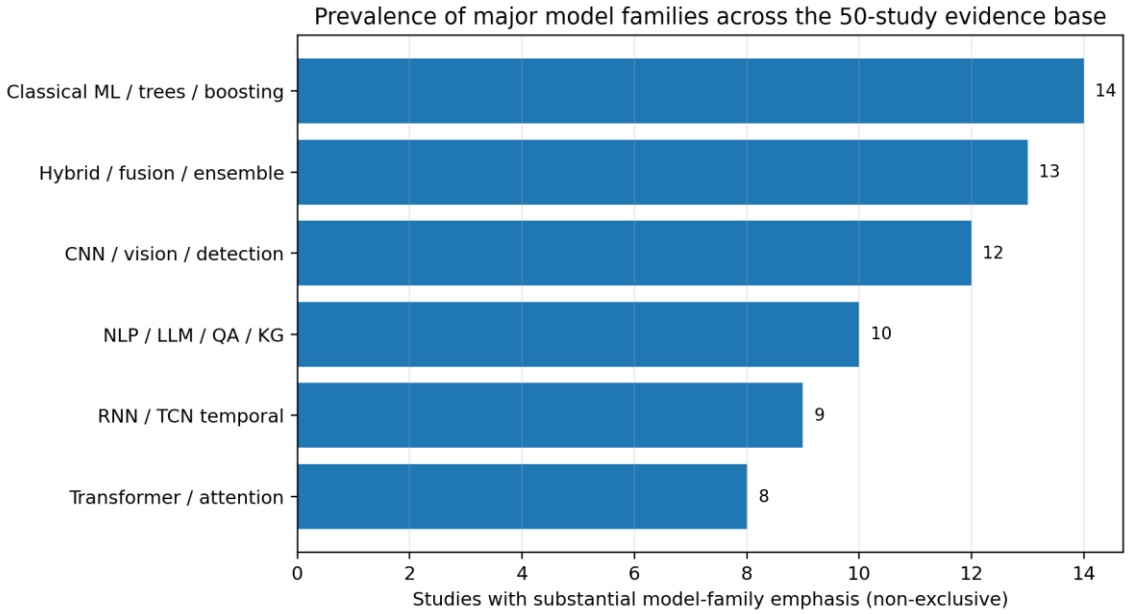


Figure 8. Cross-study prevalence of major model families; categories are non-exclusive.

Instead of naming a winner, Figure 8 presents a measurement of the breadth of the methodology. In total, there are a total of 14 research that employ traditional machine learning, tree, and boosting techniques, 13 studies that employ hybrid, fusion, and ensemble designs, and 12 studies that employ vision and detection techniques. In 10 of the works, either NLP, LLM, QA, or KG methods were utilized, nine of the works utilized temporal RNN or TCN approaches, and eight of the works utilized transformer or attention strategies. In light of the fact that research can be of use to more than one family, the counts are not exclusive on purpose. As a result of the distribution's support for a modular design, various students are able to remain in their own networks rather than being forced into a set of single network process.

A clear design concept emerges as a result of the comparative synthesis: outputs ought to be standardized prior to fusion, but specialization ought to be maintained. In the sixth section, the evidence is transformed into a clear readiness score, an operational description, a layered architecture, and a validation process that is based on this principle.

6. NLP-ENABLED CROP MARKET-READINESS FRAMEWORK

When it is time to harvest and sell a crop, crop market readiness refers to the period when it is appropriate to do so, taking into consideration the crop's biological state, the anticipated marketable volume and quality, the anticipated price and demand, the logistical and postharvest risk, and the uncertainty. This is not the same as the maturity of the crop. In the event that market prices are temporarily negatively impacted and the risk of storage is low, a crop that has reached its biological maturity may not be ready for sale. If, on the other hand, a crop is rapidly deteriorating, it may be necessary to sell it even if the price estimate is not favorable.

With this definition of ready, the decision goal is separated from any one proxy, such as price, yield, or ripeness, which is a significant advantage. Additionally, it establishes the framework of the system, which includes the following: a good prediction service must be able to guess each latent component; natural language processing (NLP) must offer context information in a manner that can be monitored; and the fusion layer must resolve ambiguity before any action is advised.

6.2 Five-layer architecture

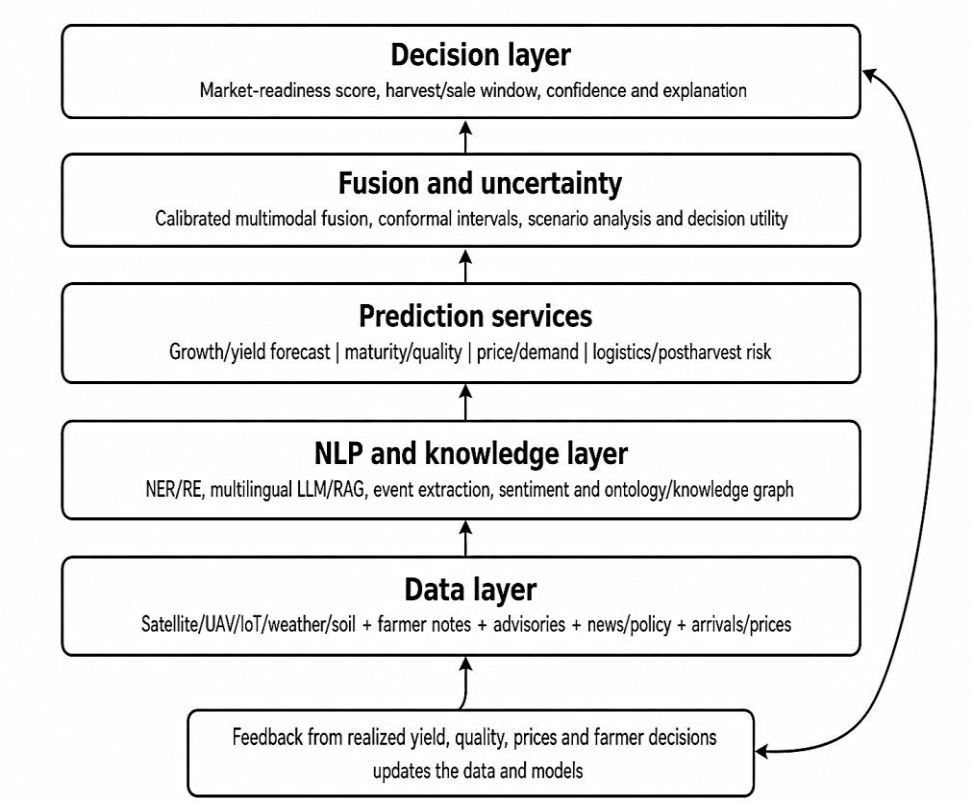


Figure 9. Proposed five-layer architecture for NLP-enabled crop growth estimation and market-readiness forecasting**.**

However, rather than being a single piece, the architecture is intended to be layered. Event types that are text-based are received thru the NLP/knowledge layer. On the other hand, prediction services for growth, maturity, pricing, and risk are still validated independently of one another. The process of uncertainty calibration must be completed before fusion can take place. Rather than being concealed in a covert representation, differences between modalities can be observed and evaluated in this manner. The loop for locating drift and updating the model is closed by the feedback path that is derived from real results. This is accomplished without the source's history being lost.

In synchrony, the data layer brings together photos captured by satellites or unmanned aerial vehicles (UAVs), data from field sensors, information about the soil and weather, historical yields, arrivals, pricing, and language that is not organized. The natural language processing layer is responsible for a variety of tasks, including language recognition, named entity and relation extraction, event detection, sentiment or stance analysis, retrieval-augmented evidence access, and ontology alignment tasks. It is necessary for every event that is extracted to have a source, a time, a place, a crop, a confidence interval, and a validity interval. The outcomes of the free-form language model are unable to be incorporated into the prediction because of the design, which prevents them from being tracked as facts.

In the past, prediction services operated independently before they merged. In the growth/yield service, biomass, phenological stage, and marketable yield are predicted; in the maturity/quality service, harvestable proportion and grade are predicted; in the price/demand service, predictions are made for more than one time period; and in the risk service, transportation, storage, weather shocks, and loss after harvest are modeled. Since there are different services for each component, it is possible for each component to make use of the appropriate model family and validation data samples. Following the utilization of calibrated probabilities or conformal prediction intervals, the outputs are distributed on a scale of uncertainty that is comparable.

Table 14. Interface specification for the proposed five-layer market-readiness architecture.

Architecture layer	Required input contract	Primary output contract	Mandatory integrity check
Data layer	Timestamped imagery, sensors, soil/weather, yield, arrivals/prices and raw text	Synchronized multimodal observations with source identifiers	Missingness, timestamp alignment, spatial consistency and access control
NLP / knowledge layer	Documents, advisories, policy/news, farmer notes and retrieved evidence	Entities, relations, events, confidence, source, location and validity interval	Factual attribution, contradiction checks, language normalization and provenance
Prediction services	Validated crop-state, maturity, price/demand and risk features	Calibrated forecasts per service and horizon	Chronological/spatial holdouts, drift checks and service-specific calibration
Fusion / uncertainty	Aligned service outputs with uncertainty estimates	Joint readiness evidence, conformal intervals and scenario sensitivity	Conflict resolution, uncertainty propagation and missing-modality stress testing
Decision layer	Readiness evidence, stakeholder constraints and feasible action windows	Ranked harvest/sale windows, utility, confidence and explanation	Constraint satisfaction, regret/utility evaluation and human override logging

The conceptual design is transformed into data contracts that may be checked with the use of Table 14. Because of the design, unstructured LLM output is prevented from evading provenance checks, and a single model family is prevented from taking over all of the jobs. In order for the fusion layer to provide a standard risk scale for the purpose of comparing the data, it is necessary for each service to display both a prediction and an uncertainty level. After that, the decision layer makes use of calibrated services rather than secret intermediate states. This not only makes it simpler to locate information, but it also enables examination of every source thru the use of ablations.

Figure 9 is no longer merely an idea; it has evolved into a flowchart that can be implemented in this. It illustrates the types of information that can be transferred from one layer to the next, as well as the conditions that must be met in terms of trustworthiness and provenance. Converting the outputs of the components that have been calibrated into a clear action requirement is the very last thing that has to be done. In Section 6.3, the ready score is intended to be used for this purpose & process.

6.3 Readiness score and decision rule

A transparent decision layer can begin with a constrained multi-criteria score rather than an opaque single network. Let Ht denote harvest readiness, Yt expected marketable-yield confidence, Pt forecast price advantage, Dt demand strength, Lt logistics feasibility, Qt quality retention, and Ut aggregate uncertainty and downside risk. A normalized market-readiness score may be written as:

$$MRSt = wH\,Ht + wY\,Yt + wP\,Pt + wD\,Dt + wL\,Lt + wQ\,Qt - wU\,Ut, \quad \Sigma wi = 1. \qquad (1)$$

For the purpose of achieving the most accurate outcomes in the past, it is necessary to first collect weights for the objectives of the stakeholders, the crops, and the areas, and then to stress-test those weights rather than choose them. During the window that maximizes expected utility over the possible time frame, taking into consideration restrictions on maturity, labor, storage, and risk, the optimal time to harvest or sell is during the window that maximizes the expected utility overall. It is necessary for the system to display the optimal window, a second-best choice, a confidence interval, the difference between the predicted income and the loss that was prevented, and the evidence that had the greatest influence on the

recommendations.

As a consequence of this, the score functions more as a technique for making decisions than as a new model for making forecasts. The usefulness of it is contingent upon the precision of the component forecasts as well as the degree to which weight sensitivity, alternative action windows, and missing modalities are tested. Consequently, as a result of this, the framework does not include a validation procedure that extends beyond the typical train–test accuracy.

6.4 Validation protocol

Table 15. Recommended validation protocol - Part I: temporal, spatial, modality and calibration validity.

Validation dimension	Minimum requirement
Temporal validity	Rolling-origin backtesting; no future leakage; multiple forecast horizons
Spatial validity	Farm, district or region holdout; cross-agro-climatic transfer
Modality robustness	Missing-sensor, cloud-cover, sparse-text and delayed-market scenarios
Calibration	Prediction Interval coverage, Brier score/ECE for probabilities, conformal coverage

The checks for predictive validity that need to be done before any practical claims can be made are discussed in the first part of this article. Rolling-origin backtesting and geographic holdouts are examples of techniques that are used to address leakage in time and space. Missing-sensor/cloud/sparse-text situations are used to evaluate the robustness of the modality, and calibration measures are used to determine whether or not confidence is statistically relevant. Compared to random train-test accuracy, these standards are more stringent because decisions regarding market readiness are directly influenced by factors such as time, location, and the absence of data samples.

Table 16. Recommended validation protocol - Part II: decision utility, NLP reliability, fairness and human factors.

Validation dimension	Minimum requirement
Decision utility	Revenue uplift, avoided postharvest loss, missed-market cost, timing regret
NLP reliability	Entity/event F1, factual attribution, source precision, contradiction rate
Fairness and access	Performance by farm size, language, connectivity and region
Human factors	Agronomist agreement, farmer comprehension, override rate and trust calibration

Table 16. Recommended validation protocol - Part II: decision utility, NLP reliability, fairness and human factors.

In Part II, the evaluation shifts from focusing on the quality of forecasts to focusing on the quality of deployment. The repercussions of a decision can be measured in a number of different ways, including revenue uplift, prevented loss, and timing regret. By limiting NLP error, entity/event F1 and factual attribution are utilized. The use of stratified analysis ensures that access and fairness are maintained across all regions. Agronomists and farmers come to an agreement to determine whether or not the system can be comprehended by human beings. The truth of the model is linked to its application in business and to responsible application thru the two components of validation that work together.

In Tables 15 and 16, the proposed system is completed by connecting each architectural layer to a matching validation duty. This brings the system to a successful conclusion in process. Because of this distinction, it is possible to differentiate between a model that functions effectively in a static measure and a decision system that continues to be beneficial even when time and location change and when real economic repercussions occur. Taking into consideration the more stringent deployment requirements, the Discussion now provides an explanation of what the evidence that was examined means.

7. DISCUSSION

7.1 What the literature establishes

The analysis demonstrates that crop-state prediction, price prediction, and maturity prediction are all technically mature

enough to function as independent systems, but they cannot be used together as a single system for decision-making. Deep temporal models are helpful when there is a lot of sequence structure; vision systems can find areas that are mature and ready to be harvested; and natural language processing (NLP) can turn papers and conversations into structured agricultural knowledge. Tree ensembles are good first-line models for structured datasets; deep temporal models are useful when there is a lot of sequence structure. When it comes to extrapolating beyond the seasons that are recognizable to us [17, 29], hybrid systems that combine learning with mechanical or domain limitations are extremely intriguing in process.

The fact that this amount of progress has occurred at the component level is not only encouraging, but it also clarifies the precise function that language intelligence plays. It is not because natural language processing can take the role of sensors, crop models, or market time series that it is valuable; rather, it is because it provides events, restrictions, and proof from outside sources that these numerical channels are unable to observe directly.

7.2 Why NLP matters for market readiness

In three different ways, NLP is useful. In the first place, it discovers events before they appear in a list. These events include policy releases, buyer requirements, disease alerts, transportation challenges, and changes in local demand. The second benefit is that it pulls together pieces of knowledge that are dispersed across a variety of languages and document types. Third, it provides explanations and evidence by establishing a connection between a change in the forecast and particular evidence. It is preferable to assign these duties to a general-purpose LLM rather than to expect them to explicitly anticipate yield or pricing based on free-form cues.

This aspect of the circumstance alters the manner in which performance ought to be evaluated. In the event that text-based signals have the ability to alter a decision on whether to harvest or sell, then F1, RMSE, R2, or accuracy values on their own are not sufficient. The downstream effect of these factors needs to be calibrated, transferred to a variety of realistic scenarios, and evaluated in terms of the decision outcome.

7.3 Why reported accuracy is not enough

A number of studies have shown extremely high results, such as a classification accuracy of approximately 99% [20, 47] and high R2 or efficiency measures [11, 17, 23, 36]. There is a possibility that these kinds of results will function in their original datasets; but, it is also possible that they will not work in a new region, crop, camera, market, or season. It is intended that agricultural deployment be flexible enough to accommodate changes in dispersion. It is therefore necessary for evaluation to differentiate between interpolation and transfer, incorporate uncertainty calibration, and determine the amount of money that the economy loses as a result of poor timing. When it comes to a crop that spoils rapidly, a single day's harvest error can be more significant than a slight shift in the overall root mean square error (RMSE).

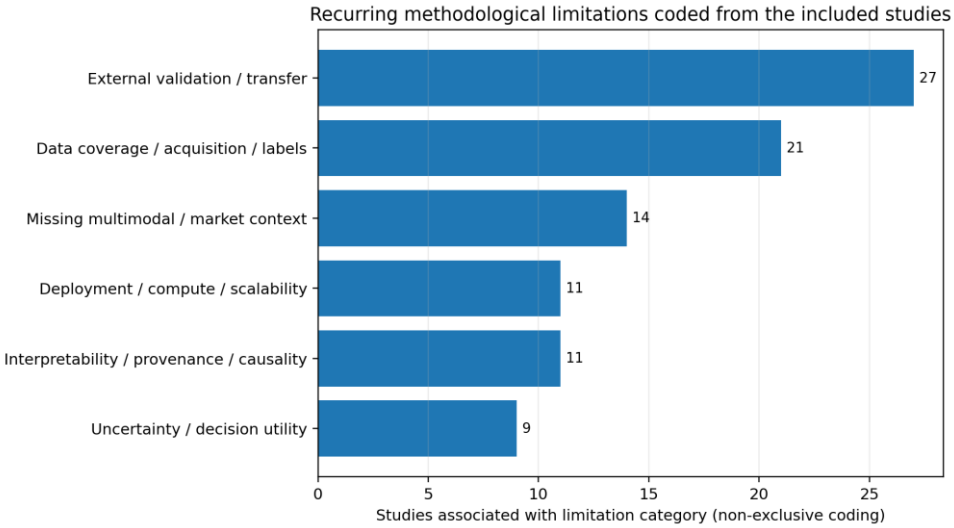


Figure 10. Frequency of recurring methodological limitations in the 50-study evidence base.

The coding of the primary limitation that was recorded for each of the studies that were included in the analysis is not exclusive, and it is the basis for Figure 10. Concerns concerning external validation or transfer are the types of issues that arise the most frequently (27 studies), followed by issues with data coverage, collecting, or labeling in 21 research. There are fourteen studies that do not have multimodal or market context, eleven studies that have issues with interpretability, provenance, causality, deployment, computing, or scaling, and nine research that have issues with uncertainty or decision-utility. Based on the coding, it has been established that the primary issues are not a deficiency in prediction algorithms but rather issues with deployment and the integration of evidence.

Figure 10 demonstrates that this view is confirmed by the data: the primary issues are not a shortage of high-performing students but rather transfer, data coverage, and missing context. Other concerns include missing context. Therefore, repeatability and governance become particularly crucial since an integrated system needs to be able to keep track of data lines and verification of where the evidence came from while still being able to be tested across different seasons, geographies, and market regimes.

7.4 Reproducibility and data governance

Inconsistent market data, private farm data, insufficient feature descriptions, and image labels that are not available to the public are all factors that make it difficult to reproduce. Due to the fact that information on buyers, farm operations, location, and production estimates is sensitive to business, there are also governance issues associated with an integrated system. Data reduction, access control, versioned provenance, and model cards are all essential requirements for the design, and they should all be considered fundamental. When it comes to LLM components, it is necessary to keep distinct records of the evidence that has been obtained and the summaries that have been prepared. This allows the user to differentiate between the source facts and the model analysis.

These rules of governance are not irrelevant to the modeling problem; they determine whether or not evidence can be believed, shared, updated, and reviewed when many data sources and stakeholders are linked together. In light of this, the future plan for Section 8 is centered on simultaneously addressing the deficiencies in science, validation, and deployment. This is in contrast to the previous plan, which merely proposed an additional improvement in accuracy.

8. RESEARCH GAPS AND FUTURE AGENDA

Table 17. Research gaps and future agenda - Part I: task definition, alignment, multilinguality, uncertainty and causality.

Research gap	Priority research direction
End-to-end task definition	Develop benchmark tasks that jointly predict crop state, harvestability, market conditions and action utility rather than isolated labels.
Text–sensor alignment	Create timestamped datasets linking advisories, news, policy and farmer narratives to satellite, sensor, phenology, yield and price observations.
Multilingual and low-resource NLP	Evaluate local-language extraction and retrieval under spelling variation, code-mixing and limited annotation.
Uncertainty-aware fusion	Use calibrated ensembles, Bayesian or conformal methods so conflicting modalities do not produce unjustified confidence.
Causal and counterfactual analysis	Distinguish predictive association from actionable drivers; estimate the consequences of harvesting now versus later.

The issue of scientific expression is discussed in the first part of this session. It is necessary to provide a comprehensive description of the task from beginning to end before that models can be optimized against the same goal. The alignment of text sensors and the extraction of many languages make it possible for machines to use external context. Without allowing contradicting evidence to give misleading confidence, uncertainty-aware fusion prevents this from happening in real time scenarios. In addition, causal and counterfactual analysis is required in order to differentiate between predicted association and actionable timing effects.

Table 18. Research gaps and future agenda - Part II: transfer, market shocks, decision metrics, human collaboration and open benchmarks.

Research gap	Priority research direction
Cross-region and cross-season transfer	Adopt leave-one-region/season/cultivar-out evaluation and report performance degradation.
Market regime shifts	Stress-test price models under policy shocks, export restrictions, extreme weather and abrupt supply-chain disruption.
Decision-centric metrics	Report timing regret, profit distribution, postharvest loss, labor utilization and risk-adjusted utility.
Human–AI collaboration	Design explanations and override mechanisms for farmers, traders, agronomists and cooperatives.
Open benchmark and reporting standard	Publish synchronized multimodal benchmarks, preprocessing code, split definitions, calibration results and model energy/cost.

During the second part, we will discuss deployment and how to make things more general. The cross-region/season review and market-regime stress tests are used to determine how robust something is when it comes to altering farming practices in the actual world. Decision-centric metrics are used to demonstrate how inaccurate predictions impact the environment. The monitoring and repeatability that are required to compare systems in an equitable manner are provided by open multimodal benchmarks and collaboration between humans and artificial intelligence sets. With the implementation of these objectives, the industry will transition from having a single correct benchmark to having decision support that can be utilized by anybody and can be verified in process.

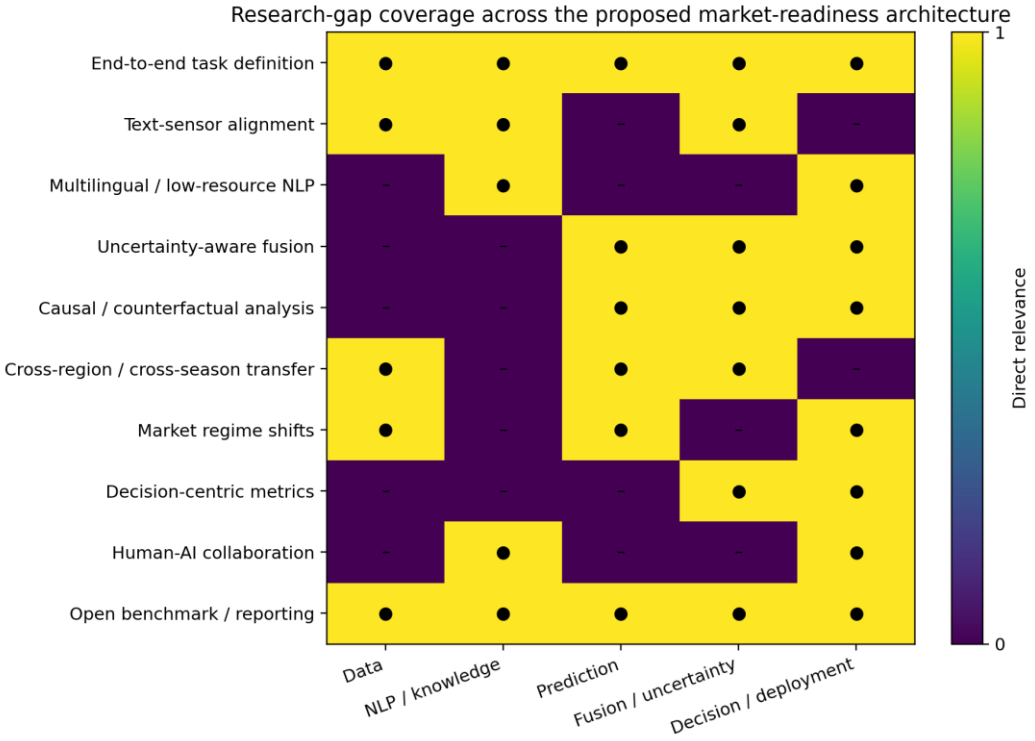


Figure 11. Mapping of the identified research gaps to the five-layer market-readiness architecture.

As can be seen in Figure 11, the gaps are not confined to a single section of the program but rather are dispersed throughout the entire system. End-to-end task definition and open benchmarking have an impact on all five layers; text-sensor alignment

is primarily found at the data, natural language processing, and fusion interfaces; uncertainty and causal analysis are primarily found at the prediction-to-decision boundaries; and it is essential for humans and artificial intelligence to collaborate when evidence is put into action. Because of this, the matrix is able to provide coordinated benchmark construction, which is preferable to the alternative of replacing single classifiers or forecasters each one individually in process.

At the moment, the most publishable opportunity is a multimodal benchmark in which text events are not merely additional information but also quantitative predictors. Changes in the prices of procurement, forecasts of approaching rain, illness outbreaks, and blockages are all examples of things that an event extractor might take into consideration. In accordance with the forecasts of crop stages and market arrivals, these occurrences would occur at the same time. Ablation experiments would then be conducted to determine whether or whether natural language processing (NLP) makes forecasting, early warning, and decision-making more effective than simply looking at weather, photographs, and historical prices.

Therefore, the approach makes sense as a whole, starting with the establishment of benchmarks and progressing to decision support that is quantified and human-auditable in this process. The way in which the review was organized, in particular the fact that the publisher had constraints, the 2026 coverage was not comprehensive, and the tasks and metrics that were reported were not all the same, must be taken into consideration before reaching a definitive decision in real time scenarios. However, before reaching this conclusion, these assertions must be limited.

9. REVIEW LIMITATIONS

The search was restricted to records written in English and originating from four different publishing families. It is possible that comparable work from other publishers, regional journals, and preprint repositories will not be found as a result of this. It is possible to find some proof in detailed publisher data rather than full-text access. This is because different publisher search interfaces have varied methods of indexing and exporting. Because of the disparity in objectives and measurements, it was unable to do a meta-analysis. Lastly, a lot of studies discovered that the performance within the dataset was quite good, but there was not enough validation from outside sources. As a result of this, rather than announcing a single model that is superior to all others, this evaluation takes the time to carefully examine comparison patterns.

Despite the fact that these issues restrict the amount of conclusions that can be formed, they do not alter the pattern that is most constant across all of the evidence, which is that powerful specialized models combine with weak end-to-end integration. Consequently, the conclusion is primarily concerned with the aspects that the study is able to reliably support. These aspects include the maturity of the component techniques, the recurrence of validation deficiencies, and the architectural requirements for a market-readiness system that makes use of natural language processing and can be defended.

10. CONCLUSION

The research field is shown to be algorithmically varied, but it is nonetheless structurally fragmented, according to the findings described. The evidence base is comprised of ten studies on natural language processing and knowledge intelligence, twenty studies on crop growth and yield, ten studies on price and market, and ten studies on maturity and readiness. The year 2024–2025 saw the publication of 36 (12%) of the fifty papers. Together, Elsevier and Springer publish 41 papers, which accounts for 82% of the total number of studies that are included. This not only demonstrates that there is a significant publication trend, but it also demonstrates that there is a focus that ought to be taken into consideration when applying the synthesis to other locations. The following research methods are utilized in the following investigations: 14 studies utilize classical machine learning/tree/boosting approaches, 13 studies utilize hybrid/fusion/ensemble designs, 12 studies utilize vision/detection methods, 10 studies utilize NLP/LLM/QA/knowledge-graph methods, and 9 studies utilize temporal RNN/TCN approaches. Eight studies, in conclusion, make use of transformer/attention designs. The pattern that has arisen is not one of algorithmic replacement; rather, it is one of increased specialization across a variety of market and agronomic tasks.

This specialized knowledge is supported by technical evidence. There is still a lot of success with tree ensembles for small structured farm datasets and regional datasets. On the other hand, temporal networks, ConvLSTM–ViT, transformers, and multimodal ensembles become more powerful as the amount of sequence or dense spatial information increases [11 thru 30]. Integration of process data can be particularly effective: the mixed crop modeling study that was examined discovered Nash–Sutcliffe efficiency values ranging from 0.954 to 0.996 [17]. Additionally, optimized recurrent, attention, and patch-based

transformer designs have been utilized in order to enhance market forecasting [31–40]. In the meantime, maturity systems are capable of achieving very high levels of discrimination within the same dataset. For instance, dragon-fruit quality grading and maturity detection were reported to be 99.99% and 99.69%, respectively [47]. In spite of this, there are a few issues with these extremely high levels. For instance, concerns regarding external validation or transfer are brought up in 27 studies, problems with data coverage, acquisition, or labeling are brought up in 21, missing multimodal or market context occurs in 14, interpretability, provenance, or causality issues are brought up in 11, and deployment, compute, or scalability issues are brought up in 9. Last but not least, problems with decision-utility or uncertainty are present in 9. Therefore, when it comes to being prepared for the market, having a high level of accuracy in generating forecasts does not yet equal having good understanding.

The most significant contribution that this review makes is the identification of an architectural gap rather than a deficiency in the number of useful component models. We looked at a number of other studies, but none of them were able to combine provenance-linked text events, crop-state estimation, maturity or quality evaluation, multi-horizon market forecasting, calibrated uncertainty, and an economically evaluated harvest/sale recommendation all at the same time. This issue is resolved by the proposed five-layer framework, which divides data collection, natural language processing and knowledge extraction, specialized prediction services, uncertainty-aware multimodal fusion, and decision support into distinct categories. In spite of the fact that it is necessary to have standard interfaces that communicate crop, location, time, confidence, and provenance, this modular structure enables each subsystem to maintain the model family that functions most effectively with its data. It is important to note that the market-readiness score should not be regarded as a substitute for crop physiology, market models, or expert restrictions; rather, it should be regarded as a clear method for making decisions regarding calibrated services.

What we are able to accomplish in the future will be determined by how effectively this kind of integration assists us in making judgments when things actually change. In addition to the standard RMSE, R2, F1, mAP, and accuracy tests, you should also perform the following: rolling-origin backtesting in chronological order, farm/region/cultivar holdouts, missing-modality stress tests, prediction Interval and conformal coverage, Brier score or expected calibration error, and market-regime shock evaluation. In addition, studies should demonstrate the amount of money that was made, whether or not there was a sense of regret over the time, whether or not postharvest losses were averted, how well the system performed in a variety of languages and areas, and whether or not humans were able to override or agree. Consequently, the most promising opportunity for research in the not-too-distant future is a synchronized multimodal benchmark. This benchmark use natural language processing (NLP) to time policy, disease, logistics, and demand events in such a way that they coincide with crop state, maturity, arrivals, and prices. Additionally, the additional value of these events is evaluated using ablation. A benchmark such as this would transform the field from a collection of precise predictors into a crop market-readiness system that is capable of being checked, takes into account uncertainty, and is economically relevant in real time scenarios.

Appendix A. Evidence Matrix of the 50 Included Studies

The evidence matrix records the task, input data, model family, principal reported contribution and the main limitation identified during synthesis. Reported numerical values are reproduced only where they were available from the publisher record or article summary; otherwise the finding is stated qualitatively.

Table 19. Evidence matrix - Part I: studies [1]–[5].

Ref.	Year / publisher	Study and focus	Data / modality	Model(s)	Main reported contribution	Primary limitation
[1]	2023 Elsevier	Zhao et al. ChatAgri: Exploring potentials of ChatGPT on cross-linguistic agricultural text classification	Agricultural text classification; multilingual corpora	Prompted ChatGPT and supervised text classifiers	Demonstrated that generative models can support cross-lingual agricultural classification, while performance varies by language, prompt design and label	Prompt sensitivity, limited domain grounding, and no linkage to crop-state or market outcomes.

					granularity.	
[2]	2024 Springer	Koopman et al. AgAsk: an agent to help answer farmers' questions from scientific documents	Scientific agricultural documents and farmer questions	Retrieval, neural reranking and extractive question answering	Showed that an evidence-retrieval agent can connect practical questions to scientific passages and improve access to agronomic knowledge.	Answer quality depends on corpus coverage; field and market signals are not modeled.
[3]	2023 Elsevier	Bagheri et al. A text analytics model for agricultural knowledge discovery and sustainable food production: A case study from Oklahoma Panhandle	Regional agricultural documents and stakeholder text	Text mining, topic extraction and decision analytics	Converted unstructured regional information into interpretable agricultural knowledge themes useful for planning.	Case-study scope and limited temporal validation.
[4]	2024 Elsevier	Gutiérrez Domínguez et al. Natural language processing of social network data for the evaluation of agricultural and rural policies	Social-media discourse related to agricultural and rural policy	NLP, sentiment/topic analysis	Illustrated how public discourse can reveal policy perceptions and emerging rural concerns.	Social-media bias, noisy geolocation and indirect relationship to farm production.
[5]	2024 Elsevier	Saravanan and Bhagavathiappan Innovative agricultural ontology construction using NLP methodologies and graph neural network	Agricultural text and domain concepts	NLP-based concept extraction, ontology construction and GNN	Produced a structured semantic representation for agricultural entities and relations.	Ontology coverage and portability require broader expert validation.

Studies [1]–[5] establish the semantic foundation: cross-lingual classification, evidence retrieval, regional text analytics, policy discourse analysis and ontology construction. Their shared limitation is that semantic outputs are rarely evaluated against downstream crop-state or market outcomes, so provenance-rich feature extraction remains the necessary bridge to forecasting.

Table 20. Evidence matrix - Part II: studies [6]–[10].

Ref.	Year / publisher	Study and focus	Data / modality	Model(s)	Main reported contribution	Primary limitation
[6]	2025 Elsevier	Das et al. Machine Learning-Powered NLP for Optimal Crop	Agricultural advisory text and production-	NLP feature extraction and machine-learning	Positioned textual knowledge as an input to crop planning and recommendation.	Limited evidence of cross-region deployment and real-time market

		Production Planning	planning variables	planning models		integration.
[7]	2025 Elsevier	De et al. Fine-tuned encoder models with data augmentation beat ChatGPT in agricultural named entity recognition and relation extraction	Agricultural NER and relation-extraction corpora	Fine-tuned transformer encoders, data augmentation and ChatGPT comparison	Reported that domain-adapted encoders can outperform general-purpose ChatGPT prompting on structured information extraction.	Task-specific labels and annotation costs constrain scaling.
[8]	2024 Elsevier	Yang et al. Application of question answering systems for intelligent agriculture production and sustainable management: A review	Agricultural question-answering literature	Review of retrieval, knowledge-base and neural QA systems	Mapped QA architectures and their roles in production support and sustainable management.	Few studies evaluate downstream agronomic or financial utility.
[9]	2025 Elsevier	Li et al. A review on enhancing agricultural intelligence with large language models	LLM applications across agriculture	Review of prompting, fine-tuning, retrieval augmentation and agents	Identified LLM opportunities in advisory, knowledge extraction and multimodal decision support.	Hallucination, provenance, multilingual equity and evaluation remain unresolved.
[10]	2025 Elsevier	Shahriar et al. The role of generative artificial intelligence in digital agri-food	Digital agri-food use cases and governance evidence	Generative AI and digital agri-food synthesis	Connected generative AI with food-chain communication, analytics and operational support.	Limited field evidence and unresolved reliability, data-governance and accountability issues.

Studies [6]–[10] extend agricultural NLP toward production planning, domain-adapted information extraction, question-answering synthesis and generative intelligence. The evidence increasingly supports retrieval grounding and task-specific encoders, but hallucination control, multilingual equity, governance and cross-region deployment remain unresolved before these tools can influence commercial timing decisions.

Ref.	Year / publisher	Study and focus	Data / modality	Model(s)	Main reported contribution	Primary limitation
[11]	2023 Elsevier	Jhajharia et al. Crop Yield Prediction using Machine Learning and Deep Learning Techniques	Structured crop, soil and weather data	RF, regression, SVM and deep-learning models	Random Forest was reported as the strongest tested model (R2 about 0.963) on the study dataset.	Single-dataset performance may not transfer across seasons and regions.
[12]	2023 Elsevier	Paudel et al. Interpretability of deep learning models for crop yield forecasting	Weather, remote sensing and crop-yield records	Deep neural forecasting with interpretability analysis	Showed that interpretability can expose how weather and remote-sensing factors drive yield predictions, including problematic behavior under extremes.	Explanations remain model- and region-dependent.
[13]	2023 Elsevier	Mohan et al. Temporal convolutional network based rice crop yield prediction using multispectral satellite data	Multispectral satellite time series for rice	Temporal convolutional network	Demonstrated the value of temporal convolutions for learning phenological patterns from satellite sequences.	Cloud effects, sensor resolution and local calibration limit generalization.
[14]	2023 Springer	Talaat Crop yield prediction algorithm (CYPA) in precision agriculture based on IoT techniques and climate changes	IoT, climate and crop-yield observations	IoT-enabled predictive algorithm and neural learning	Integrated climate and IoT variables for yield estimation within a precision-agriculture workflow.	Deployment assumptions and external validation were limited.
[15]	2023 Hindawi/Wiley	Ayalew et al. Prediction of Crop Yield by Support Vector Machine Coupled with Deep Learning Algorithm Procedures in Lower Kulfo Watershed of Ethiopia	Farm, climate and watershed variables	SVM coupled with deep-learning procedures	Found that combined nonlinear learners can improve farm-level yield estimation in the study watershed.	Local sample size and watershed specificity.

Table 21. Evidence matrix - Part I: studies [11]–[15].

Studies [11]–[15] emphasize structured ML, temporal learning and increasingly multimodal crop-state estimation. The strongest contributions improve nonlinear or spatiotemporal representation, while the recurring weakness is transfer beyond the original season, region or sensor configuration. This group therefore strengthens the supply-estimation component but does not yet close the market-context gap.

Ref.	Year / publisher	Study and focus	Data / modality	Model(s)	Main reported contribution	Primary limitation
[16]	2023 IEEE	Sharma et al. Predicting Agriculture Yields Based on Machine Learning Using Regression and Deep Learning	Agricultural production and environmental variables	Regression, machine learning and deep learning	Compared conventional and deep models and reinforced the competitiveness of ensemble/nonlinear approaches.	Limited discussion of leakage-resistant temporal validation.
[17]	2024 Elsevier	Jeong et al. Deep learning-enhanced remote sensing Integrated crop modeling for rice yield prediction	Process-based crop model, remote sensing and field yield	Hybrid crop model and deep learning	Reported very high simulation agreement (NSE 0.954–0.996) by coupling remote sensing with crop modeling.	High data and calibration requirements; region-specific process parameters.
[18]	2024 Elsevier	Subramaniam and Marimuthu Crop yield prediction using effective deep learning and dimensionality reduction approaches for Indian regional crops	Indian regional crop, soil and weather data	Dimensionality reduction and deep learning	Reduced input redundancy while maintaining strong predictive performance across selected crops.	Transfer to unseen agro-climatic zones was not fully tested.
[19]	2024 Springer	Abdel-salam et al. A proposed framework for crop yield prediction using hybrid feature selection approach and optimized machine learning	Multifactor agricultural datasets	Hybrid feature selection and optimized ML	Improved prediction by combining feature screening with hyperparameter optimization.	Optimization cost and sensitivity to dataset composition.
[20]	2024 Springer	Nagesh et al. Boosting enabled efficient machine learning technique for accurate prediction of crop yield towards precision agriculture	Structured crop and environmental data	AdaBoost and comparative ML	Reported high classification-style performance, including about 98% accuracy and 99% recall in the evaluated setup.	Accuracy-based reporting may not represent continuous yield error or cross-season robustness.

Table 22. Evidence matrix - Part II: studies [16]–[20].

Studies [16]–[20] broaden growth/yield forecasting through hybrid process-data models, optimization, remote sensing and explainability. Their collective value lies in more physically or spatially informed yield estimates, yet calibration burden, dataset locality and independent replication continue to limit deployment-scale confidence.

Ref.	Year / publisher	Study and focus	Data / modality	Model(s)	Main reported contribution	Primary limitation
[21]	2024 IEEE	Badshah et al. Crop Classification and Yield Prediction Using Robust Machine Learning Models for Agricultural Sustainability	Crop, weather and soil attributes	Robust ML for classification and yield prediction	Unified crop identification and yield estimation in a sustainability-oriented workflow.	Limited uncertainty quantification and real-world decision evaluation.
[22]	2024 IEEE	Ashfaq et al. Accurate Wheat Yield Prediction Using Machine Learning and Climate-NDVI Data Fusion	Climate records and NDVI for wheat	RF, SVM and data fusion	Climate-NDVI fusion improved wheat-yield prediction; RF was reported around 97% and SVM around 93% in the study metric.	Performance is tied to regional data quality and spatial scale.
[23]	2024 IEEE	Erlin et al. Enhancing Rice Production Prediction in Indonesia Using Advanced Machine Learning Models	2018–2023 provincial production, rainfall, humidity and temperature	RF, gradient boosting, DT, SVM and ANN	Random Forest remained strongest after tuning, with reported R2 improving to approximately 0.91.	Provincial aggregation masks farm-level heterogeneity.
[24]	2024 IEEE	Nejad et al. ConvLSTM–ViT: A Deep Neural Network for Crop Yield Prediction Using Earth Observations and Remotely Sensed Data	Earth observation and remote-sensing sequences	ConvLSTM and Vision Transformer	Combined local spatiotemporal encoding with global attention for crop-yield prediction.	Computational cost, limited interpretability and dependence on dense imagery.
[25]	2025 Elsevier	Gaso et al. Predicting spatial and temporal variability in soybean yield using deep learning and open source data	Open-source weather, soil and remote-sensing time series	1D-CNN, LSTM and Transformer	Compared temporal deep learners for soybean-yield variability and showed that architecture choice depends on spatial-temporal structure.	Open data resolution and transfer across production systems remain constraints.

Table 23. Evidence matrix - Part I: studies [21]–[25].

Studies [21]–[25] emphasize structured ML, temporal learning and increasingly multimodal crop-state estimation. The strongest contributions improve nonlinear or spatiotemporal representation, while the recurring weakness is transfer beyond the original season, region or sensor configuration. This group therefore strengthens the supply-estimation component but does not yet close the market-context gap.

Ref.	Year / publisher	Study and focus	Data / modality	Model(s)	Main reported contribution	Primary limitation
[26]	2025 Elsevier	Menon et al. Deep learning based farm-level crop yield prediction using multi-temporal satellite data for complex engineering application	Farm-level multi-temporal satellite imagery	Deep-learning regression	Demonstrated farm-scale estimation from temporal satellite observations.	Small-field boundary noise and cloud gaps affect reliability.
[27]	2025 Elsevier	Yewle et al. Multi-modal data fusion and deep ensemble learning for rice yield prediction	Mekong Delta multimodal remote sensing, weather and crop data	RicEns-Net deep ensemble and multimodal fusion	Reported improved rice-yield prediction by combining complementary modalities and ensemble learners.	Complex pipeline, high data harmonization burden and limited independent replication.
[28]	2025 Springer	El-Kenawy et al. Predicting Potato Crop Yield with Machine Learning and Deep Learning for Sustainable Agriculture	Potato production, soil and environmental variables	ML and DL comparison with optimization	Demonstrated that tuned nonlinear models can support potato-yield forecasting and sustainability planning.	Dataset locality and limited causal interpretation.
[29]	2025 Springer	Khodjaev et al. Optimizing machine learning models for wheat yield estimation using a comprehensive UAV dataset	UAV vegetation indices, crop height, solar radiation and soil properties	RF, XGBoost, boosting, trees and SHAP	Optimized RF and XGBoost reached R2 values up to about 0.89; SHAP highlighted red-edge indices and crop height.	UAV acquisition and site-specific calibration limit large-area scalability.
[30]	2025 Springer	Seck et al. Crop yield forecasting in Senegal: application of machine learning methods	Groundnut, millet and cotton yield with climate/agronomic predictors	Stepwise regression, LASSO and RF	Compared sparse statistical and ensemble models for multiple Senegalese crops.	National data scarcity and limited deep temporal modeling.

Table 24. Evidence matrix - Part II: studies [26]–[30].

Studies [26]–[30] broaden growth/yield forecasting through hybrid process-data models, optimization, remote sensing and explainability. Their collective value lies in more physically or spatially informed yield estimates, yet calibration burden, dataset

locality and independent replication continue to limit deployment-scale confidence.

Read together, Tables 21–24 show how growth/yield research progresses from conventional structured learners toward temporal, multimodal and process Informed forecasting. The next evidence block, studies [31]–[40], moves downstream from estimating supply to modeling the price and market trajectory against which that supply must be timed.

Ref.	Year / publisher	Study and focus	Data / modality	Model(s)	Main reported contribution	Primary limitation
[31]	2024 Springer	Sari et al. Various optimized machine learning techniques to predict agricultural commodity prices	Eleven agricultural commodity price series; 60-day horizon	GA-ELM, GA-LSTM and Auto-ARIMA	GA-ELM provided the strongest overall forecasts among the compared optimized and statistical models.	Univariate emphasis and limited use of textual demand or policy signals.
[32]	2024 IEEE	Meena and Chaitra A Novel Framework Using Deep Learning Techniques for Ragi Price Prediction in Karnataka	Monthly ragi prices from multiple Karnataka markets	VAR-Stacked LSTM	The hybrid framework reported lower forecasting error than comparison models and produced market-wise price ranges.	Commodity- and state-specific design; external shocks and text signals were not modeled.
[33]	2025 Elsevier	Sediyono et al. An integrated framework for multi-commodity agricultural price forecasting and anomaly detection using attention-boosted models	Multiple agricultural commodity price series	Transformer price prediction and attention-boosted LSTM anomaly detection	Integrated forecasting with anomaly detection across commodities, strengthening monitoring of sudden market deviations.	Limited linkage to crop maturity, yield uncertainty and logistics constraints.

[34]	2025 Springer	Manogna et al. A novel hybrid neural network-based volatility forecasting of agricultural commodity prices: empirical evidence from India	Indian agricultural commodity prices and volatility series	Hybrid neural network volatility model	Improved volatility forecasting relative to conventional benchmarks on the examined Indian commodities.	Volatility is not equivalent to price level or selling readiness.
[35]	2025 Springer	Nayak et al. Meta-transformer: leveraging metaheuristic algorithms for agricultural commodity price forecasting	Agricultural commodity price series	Transformer optimized with GWO, WOA and PSO	Metaheuristic tuning improved transformer forecasting and reduced manual hyperparameter selection.	Optimization overhead and risk of market-specific overfitting.

Table 25. Evidence matrix - Part I: studies [31]–[35].

Studies [31]–[35] show the transition from optimized shallow learners and stacked recurrent models toward transformer and anomaly-aware market forecasting. Accuracy improvements are meaningful, but the models still rely mainly on numerical histories and seldom incorporate crop maturity, logistics or provenance-linked text events.

Ref.	Year / publisher	Study and focus	Data / modality	Model(s)	Main reported contribution	Primary limitation
[36]	2025 Springer	Tewachew et al. Forecasting Sesame Price in Ethiopia Using Recurrent Neural Network Based Deep Learning Algorithms	Ethiopian sesame price time series	LSTM, GRU, BiLSTM and BiGRU	BiGRU was the best recurrent architecture, with reported R2 around 97.34% before tuning in one evaluation.	Single-commodity setting and limited exogenous features.
[37]	2026 Springer	Varshini et al. CNN-transformer hybrid deep learning model for agricultural commodities price forecasting	Multi-commodity price sequences	CNN-Transformer hybrid	Combined local pattern extraction with long-range attention for price forecasting.	Early evidence; stability across structural market changes requires longer evaluation.

[38]	2026 Springer	Farzana et al. Unlocking agricultural price trends with a BiGRU attention framework	Agricultural price time series	BiGRU with attention	Reported strong trend prediction, including an R2 near 0.88 in the evaluated setting.	Limited transparency regarding causal market drivers and deployment latency.
[39]	2026 Springer	Vairo et al. Assessing the role of weather in commodity price forecasting	Durum-wheat commodity prices with meteorological covariates	XGBoost and LSTM with/without weather	Directly tested whether weather adds predictive value and showed that gains depend on model and forecast horizon.	Weather alone cannot represent policy, trade, storage and demand shocks.
[40]	2026 Springer	Kalaiarasan et al. Enhancing the Accuracy of Multi-Horizon Potato Price Forecasts through Decomposition-based PatchTST Modelling	Potato prices across forecasting horizons	Signal decomposition and PatchTST	Improved multi-horizon forecasting by combining decomposition with patch-based transformer learning.	Commodity-specific model and no harvest-readiness fusion.

Table 26. Evidence matrix - Part II: studies [36]–[40].

Studies [36]–[40] extend price forecasting with bidirectional recurrence, attention, weather conditioning and decomposition-based PatchTST modeling. These methods improve temporal representation and multi-horizon forecasting, yet regime shifts, exogenous policy/trade shocks and direct harvest-readiness fusion remain the main barriers to actionable selling recommendations.

The market evidence therefore contributes an economic timing signal but does not establish biological feasibility. Tables 27–28 complete the matrix with maturity and harvest-readiness studies, providing the visual and operational evidence needed to delimit when a crop can actually be harvested.

Ref.	Year / publisher	Study and focus	Data / modality	Model(s)	Main reported contribution	Primary limitation
[41]	2024 Elsevier	Tamilarasi and Muthulakshmi Machine Vision Algorithm for Detection and Maturity Prediction of Brinjal	Field mobile images of KKM-1 brinjal	Detection, k-means segmentation and size-based maturity rules	Detected brinjal and classified moderate maturity, with reported performance near 95.9%.	Rule sensitivity to shadows, occlusion and cultivar morphology.
[42]	2024	Muqaddas et al.	Sentinel-2	Deep semantic	Mapped	Pixel labels,

	Elsevier	A Comprehensive Deep Learning Approach for Harvest Ready Sugarcane Pixel Classification in Punjab, Pakistan Using Sentinel-2 Multispectral Imagery	multispectral imagery	segmentation/classification	harvest-ready sugarcane fields at regional scale using multispectral imagery.	cloud cover and regional crop-calendar dependence.
[43]	2024 Elsevier	Subeesh et al. UAV imagery coupled deep learning approach for the development of an adaptive in-house web-based application for yield estimation in citrus orchard	UAV citrus-orchard imagery	Deep learning, fruit segmentation/counting and web application	Operationalized orchard yield estimation in the DeepYield web application.	Counting errors under occlusion and limited connection to prices or demand.
[44]	2025 Springer	Zhao et al. Smart UAV-assisted blueberry maturity monitoring with deep learning	UAV imagery of blueberry orchards	Object detection and maturity classification	Enabled spatial monitoring of blueberry maturity using UAV-based computer vision.	Small fruit size, illumination and canopy occlusion affect field transfer.
[45]	2025	Colaco et al.	Cashew maturity/quality	Review and comparative synthesis of AI models	Synthesized image, spectral	Heterogeneous datasets and

	Springer	Artificial intelligence advances for cashew fruit maturity and quality assessment	datasets and prior AI studies		and learning approaches for cashew maturity and quality assessment.	absence of common benchmarks.

Table 27. Evidence matrix - Part I: studies [41]–[45].

Studies [41]–[45] move harvest-readiness analysis from rule-based machine vision toward remote-sensing segmentation, UAV workflows and deep object detection. The models can localize or classify maturity effectively, but field transfer is sensitive to occlusion, illumination, crop calendars and benchmark heterogeneity.

Table 28. Evidence matrix - Part II: studies [46]–[50].

Ref.	Year / publisher	Study and focus	Data / modality	Model(s)	Main reported contribution	Primary limitation
[46]	2025 Springer	Goh et al. Fresh Fruit Bunch Ripeness Classification Methods: A Review	Oil-palm fresh-fruit-bunch ripeness literature	Review of vision, spectral and learning methods	Mapped ripeness indicators and classification approaches relevant to mechanized harvesting.	Limited standardized field datasets and cross-estate validation.
[47]	2025 Hindawi/Wiley	Nikam et al. Predictive Classification Model for Quality Grading and Maturity Detection of Dragon Fruit Using Fused Deep CNN Feature and Ensemble Learning	Dragon-fruit images	Fused deep CNN features, RF and AdaBoost	Reported 99.99% quality-grading accuracy with RF and 99.69% maturity accuracy with AdaBoost.	Likely controlled image conditions and limited cultivar/domain-shift analysis.
[48]	2025 Hindawi/Wiley	Chotikawanid et al. Optimizing an Object Detection Algorithm for Detecting Oil Palm Fruit Ripeness	Oil-palm fruit-bunch images	Optimized YOLOv7	Optimized object detection for ripeness localization and classification in oil-palm fruit bunches.	Performance can degrade with occlusion, lighting variation and estate-specific appearance.
[49]	2026 Elsevier	Zhang et al. A drone- and deep-learning-based harvest planning model for broccoli	Drone imagery and broccoli field observations	Drone vision and deep-learning harvest planning	Moved beyond maturity detection toward field-level harvest scheduling.	Early-stage evidence, crop specificity and limited market-price coupling.
[50]	2026	Hemrattrakun et al. Lightweight Deep	Cherry images	Lightweight CNN/deep-	Demonstrated fine-grained maturity	Visual maturity may not fully

	Elsevier	Learning Models for Six-Stage Maturity Classification of Cherries	across six maturity stages	learning classifiers	recognition suitable for resource-constrained deployment.	represent internal quality or market demand.

Studies [46]–[50] emphasize ripeness review evidence, fused deep features, optimized detection, drone-based harvest planning and lightweight maturity classification. They move closest to operational harvest timing, yet internal quality, domain shift, market-price coupling and commercial decision utility are still incompletely represented.

Across Tables 19–28, the appendix reproduces the same systems-level pattern observed in the main synthesis: semantic context, crop state, market dynamics and harvestability are each well represented, but their outputs are seldom evaluated as one coordinated decision chain. The numbered reference list that follows provides the bibliographic traceability for all 50 included studies and the two PRISMA reporting standards.

REFERENCES

[1] Zhao, B., Jin, W., Del Ser, J., & Yang, G. (2023). ChatAgri: Exploring potentials of ChatGPT on cross-linguistic agricultural text classification. Neurocomputing, 557, 126708. https://doi.org/10.1016/j.neucom.2023.126708

[2] Koopman, B., Mourad, A., Li, H., van der Vegt, A., Zhuang, S., Gibson, S., Dang, Y., Lawrence, D., & Zuccon, G. (2024). AgAsk: An agent to help answer farmers' questions from scientific documents. International Journal on Digital Libraries, 25, 569–584. https://doi.org/10.1007/s00799-023-00369-y

[3] Bagheri, A., Taghvaeian, S., & Delen, D. (2023). A text analytics model for agricultural knowledge discovery and sustainable food production: A case study from Oklahoma Panhandle. Decision Analytics Journal, 9, 100350. https://doi.org/10.1016/j.dajour.2023.100350

[4] Gutiérrez Domínguez, A., Roig-Tierno, N., et al. (2024). Natural language processing of social network data for the evaluation of agricultural and rural policies. Journal of Rural Studies, 109, 103341.

[5] Saravanan, K. S., & Bhagavathiappan, V. (2024). Innovative agricultural ontology construction using NLP methodologies and graph neural network. Engineering Science and Technology, an International Journal, 52, 101675. https://doi.org/10.1016/j.jestch.2024.101675

[6] Das, S., et al. (2025). Machine Learning-Powered NLP for Optimal Crop Production Planning. Procedia Computer Science. https://doi.org/10.1016/j.procs.2025.04.454

[7] De, S., Sanyal, D. K., & Mukherjee, I. (2025). Fine-tuned encoder models with data augmentation beat ChatGPT in agricultural named entity recognition and relation extraction. Expert Systems with Applications, 277, 127126. https://doi.org/10.1016/j.eswa.2025.127126

[8] Yang, T., Mei, Y., Xu, L., Yu, H., & Chen, Y. (2024). Application of question answering systems for intelligent agriculture production and sustainable management: A review. Resources, Conservation & Recycling, 204, 107497. https://doi.org/10.1016/j.resconrec.2024.107497

[9] Li, H., Wu, H., Li, Q., & Zhao, C. (2025). A review on enhancing agricultural intelligence with large language models. Artificial Intelligence in Agriculture, 15, 671–685. https://doi.org/10.1016/j.aiia.2025.05.006

[10] Shahriar, S., Corradini, M. G., Sharif, S., Moussa, M., & Dara, R. (2025). The role of generative artificial intelligence in digital agri-food. Journal of Agriculture and Food Research, 20, 101787. https://doi.org/10.1016/j.jafr.2025.101787

[11] Jhajharia, K., Mathur, P., Jain, S., & Nijhawan, S. (2023). Crop yield prediction using machine learning and deep learning techniques. Procedia Computer Science, 218, 406–417. https://doi.org/10.1016/j.procs.2023.01.023

[12] Paudel, D., de Wit, A., Boogaard, H., Marcos, D., Osinga, S., & Athanasiadis, I. N. (2023). Interpretability of deep learning models for crop yield forecasting. Computers and Electronics in Agriculture, 206, 107663. https://doi.org/10.1016/j.compag.2023.107663

[13] Mohan, A., Venkatesan, M., Prabhavathy, P., & Jayakrishnan, A. (2023). Temporal convolutional network based rice crop yield prediction using multispectral satellite data. Infrared Physics & Technology, 135, 104960. https://doi.org/10.1016/j.infrared.2023.104960

[14] Talaat, F. M. (2023). Crop yield prediction algorithm (CYPA) in precision agriculture based on IoT techniques and climate changes. Neural Computing and Applications, 35, 17281–17292. https://doi.org/10.1007/s00521-023-08619-5

[15] Ayalew, A. T., et al. (2023). Prediction of crop yield by support vector machine coupled with deep learning algorithm procedures in Lower Kulfo Watershed of Ethiopia. Journal of Engineering, 2023, 6675523. https://doi.org/10.1155/2023/6675523

[16] Sharma, P., Dadheech, P., Aneja, N., & Aneja, S. (2023). Predicting agriculture yields based on machine learning using regression and deep learning. IEEE Access, 11, 111255–111264. https://doi.org/10.1109/ACCESS.2023.3321861

[17] Jeong, S., Ko, J., Ban, J.-O., Shin, T., & Yeom, J.-M. (2024). Deep learning-enhanced remote sensing Integrated crop modeling for rice yield prediction. Ecological Informatics, 84, 102886. https://doi.org/10.1016/j.ecoinf.2024.102886

[18] Subramaniam, L. K., & Marimuthu, R. (2024). Crop yield prediction using effective deep learning and dimensionality reduction approaches for Indian regional crops. e-Prime, 8, 100611. https://doi.org/10.1016/j.prime.2024.100611

[19] Abdel-salam, M., Kumar, N., & Mahajan, S. (2024). A proposed framework for crop yield prediction using hybrid feature selection approach and optimized machine learning. Neural Computing and Applications, 36, 20723–20750. https://doi.org/10.1007/s00521-024-10226-x

[20] Nagesh, O. S., et al. (2024). Boosting enabled efficient machine learning technique for accurate prediction of crop yield towards precision agriculture. Discover Sustainability, 5. https://doi.org/10.1007/s43621-024-00254-x

[21] Badshah, A., Alkazemi, B. Y., Din, F., Zamli, K. Z., & Haris, M. (2024). Crop classification and yield prediction using robust machine learning models for agricultural sustainability. IEEE Access, 12, 162799–162813. https://doi.org/10.1109/ACCESS.2024.3486653

[22] Ashfaq, M., et al. (2024). Accurate wheat yield prediction using machine learning and climate-NDVI data fusion. IEEE Access, 12, 40947–40961. https://doi.org/10.1109/ACCESS.2024.3376735

[23] Erlin, Yunianta, A., Wulandhari, L. A., Desnelita, Y., Nasution, N., & Junadhi. (2024). Enhancing rice production prediction in Indonesia using advanced machine learning models. IEEE Access, 12, 151161–151177. https://doi.org/10.1109/ACCESS.2024.3478738

[24] Nejad, S. M. M., Abbasi-Moghadam, D., & Sharifi, A. (2024). ConvLSTM–ViT: A deep neural network for crop yield prediction using Earth observations and remotely sensed data. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 17, 17489–17502. https://doi.org/10.1109/JSTARS.2024.3464411

[25] Gaso, D. V., et al. (2025). Predicting spatial and temporal variability in soybean yield using deep learning and open source data. European Journal of Agronomy, 164, 127498.

[26] Menon, A. S., et al. (2025). Deep learning based farm-level crop yield prediction using multi-temporal satellite data for complex engineering application. Smart Agricultural Technology, 12.

[27] Yewle, A. D., et al. (2025). Multi-modal data fusion and deep ensemble learning for rice yield prediction. Smart Agricultural Technology.

[28] El-Kenawy, E. S. M., Alhussan, A. A., Khodadadi, N., Mirjalili, S., & Eid, M. M. (2025). Predicting potato crop yield with machine learning and deep learning for sustainable agriculture. Potato Research, 68, 759–792. https://doi.org/10.1007/s11540-024-09753-w

[29] Khodjaev, S., Bobojonov, I., Kuhn, L., & Glauben, T. (2025). Optimizing machine learning models for wheat yield estimation using a comprehensive UAV dataset. Modeling Earth Systems and Environment, 11, 15. https://doi.org/10.1007/s40808-024-02188-9

[30] Seck, N. K. G., Ngom, A., Ngom, P., & Noba, K. (2025). Crop yield forecasting in Senegal: Application of machine learning methods. Discover Agriculture, 3, 192. https://doi.org/10.1007/s44279-025-00381-7

[31] Sari, M., Duran, S., Kutlu, H., Guloglu, B., et al. (2024). Various optimized machine learning techniques to predict agricultural commodity prices. Neural Computing and Applications, 36, 11439–11459. https://doi.org/10.1007/s00521-024-09679-x

[32] Meena, K., & Chaitra, B. (2024). A novel framework using deep learning techniques for Ragi price prediction in Karnataka. IEEE Access, 12, 136103–136119. https://doi.org/10.1109/ACCESS.2024.3455892

[33] Sediyono, E., et al. (2025). An integrated framework for multi-commodity agricultural price forecasting and anomaly detection using attention-boosted models. Journal of Agriculture and Food Research, 22, 102048.

[34] Manogna, R. L., Dharmaji, V., & Sarang, S. (2025). A novel hybrid neural network-based volatility forecasting of agricultural commodity prices: Empirical evidence from India. Journal of Big Data, 12, 85. https://doi.org/10.1186/s40537-025-01131-8

[35] Nayak, G. H. H., Alam, M. W., Naik, B. S., Varshini, B. S., Avinash, G., Kumar, R. R., Ray, M., et al. (2025). Meta-transformer: Leveraging metaheuristic algorithms for agricultural commodity price forecasting. Journal of Big Data, 12, 138. https://doi.org/10.1186/s40537-025-01196-5

[36] Tewachew, Y., Alemu, A., Wodaynew, A., & Mekuriaw, Y. (2025). Forecasting sesame price in Ethiopia using recurrent neural network based deep learning algorithms. International Journal of Computational Intelligence Systems, 18, 316. https://doi.org/10.1007/s44196-025-00988-0

[37] Varshini, B. S., Alam, M. W., Nayak, G. H. H., Avinash, G., Vinayaka, Kumar, R. R., & Ray, M. (2026). CNN-transformer hybrid deep learning model for agricultural commodities price forecasting. Discover Computing, 29, 139. https://doi.org/10.1007/s10791-026-09966-2

[38] Farzana, G. S., et al. (2026). Unlocking agricultural price trends with a BiGRU attention framework. Discover Computing. https://doi.org/10.1007/s10791-026-10186-x

[39] Vairo, A., et al. (2026). Assessing the role of weather in commodity price forecasting. Computational Management Science. https://doi.org/10.1007/s10287-026-00576-3

[40] Kalaiarasan, G., Anbuchchelvi, I., & Premjith, B. (2026). Enhancing the accuracy of multi-horizon potato price forecasts through decomposition-based PatchTST modelling. Potato Research, 69, 121. https://doi.org/10.1007/s11540-026-10074-3

[41] Tamilarasi, T., & Muthulakshmi, P. (2024). Machine vision algorithm for detection and maturity prediction of brinjal. Smart Agricultural Technology, 7, 100402.

[42] Muqaddas, S., Qureshi, W. S., Jabbar, H., Munir, A., & Haider, A. (2024). A comprehensive deep learning approach for harvest ready sugarcane pixel classification in Punjab, Pakistan using Sentinel-2 multispectral imagery. Remote Sensing Applications: Society and Environment, 35, 101225.

[43] Subeesh, A., Kumar, S. P., Chakraborty, S. K., Upendar, K., Chandel, N. S., Jat, D., et al. (2024). UAV imagery coupled deep learning approach for the development of an adaptive in-house web-based application for yield estimation in citrus orchard. Measurement, 234, 114786. https://doi.org/10.1016/j.measurement.2024.114786

[44] Zhao, F., et al. (2025). Smart UAV-assisted blueberry maturity monitoring with deep learning. Precision Agriculture. https://doi.org/10.1007/s11119-025-10252-2

[45] Colaco, L., et al. (2025). Artificial intelligence advances for cashew fruit maturity and quality assessment. Journal of Big Data. https://doi.org/10.1186/s40537-025-01296-2

[46] Goh, J. Y., et al. (2025). Fresh fruit bunch ripeness classification methods: A review. Food and Bioprocess Technology. https://doi.org/10.1007/s11947-024-03483-0

[47] Nikam, S. B., et al. (2025). Predictive classification model for quality grading and maturity detection of dragon fruit using fused deep CNN feature and ensemble learning. Journal of Food Processing and Preservation, 2025, 6938071. https://doi.org/10.1155/jfpp/6938071

[48] Chotikawanid, P., et al. (2025). Optimizing an object detection algorithm for detecting oil palm fruit ripeness. Applied Computational Intelligence and Soft Computing, 2025, 6263757. https://doi.org/10.1155/acis/6263757

[49] Zhang, X., et al. (2026). A drone- and deep-learning-based harvest planning model for broccoli. Computers and Electronics in Agriculture.

[50] Hemrattrakun, P., et al. (2026). Lightweight deep learning models for six-stage maturity classification of cherries. Smart Agricultural Technology.

[51] Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., et al. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. BMJ, 372, n71. https://doi.org/10.1136/bmj.n71

[52] Rethlefsen, M. L., Kirtley, S., Waffenschmidt, S., Ayala, A. P., Moher, D., Page, M. J., Koffel, J. B., & PRISMA-S Group. (2021). PRISMA-S: An extension to the PRISMA Statement for Reporting Literature Searches in Systematic Reviews. Systematic Reviews, 10, 39. https://doi.org/10.1186/s13643-020-01542-z