AGRICULTURAL POWER JOURNAL
Vol. 3 No. 4, November 2026
Page 1 – 14 | e-ISSN: 3062-8563

Machine Learning-Based Yield and Harvest Timing Prediction for High-Value Horticultural Crops Under Variable Environmental Conditions

Naima Haruna [1]*, Johannis P. Haumahu [2], Rudy Soplanit [3], Simson Liubana [4], and Zuyasna [5]
[1] Universitas Andi Djemma, [2] Universitas Pattimura, [3] Universitas Pattimura, [4] Universitas Pattimura, [5] Universitas Syiah Kuala
* Correspondence: naimaharuna@unanda.ac.id

ABSTRACT

Accurate prediction of crop yield and harvest timing is essential for improving production planning, resource allocation, and postharvest management in high-value horticultural crops. However, yield and maturity are strongly influenced by variable environmental conditions, including temperature, rainfall, humidity, soil characteristics, and crop growth dynamics, making conventional prediction approaches less responsive to changing field conditions. This study aims to develop a machine learning-based predictive framework for estimating crop yield and harvest timing by integrating crop growth and environmental variables. The proposed approach utilizes historical horticultural production data combined with meteorological and environmental observations as input features. Several machine learning algorithms are evaluated and compared based on their predictive performance using appropriate regression and classification or temporal prediction metrics. The framework is designed to identify the environmental and crop-related variables that contribute most strongly to yield and harvest timing predictions while capturing nonlinear relationships within the data. The expected results provide an integrated prediction approach that can support more accurate yield estimation and timely harvest decisions under variable environmental conditions. The proposed framework is expected to contribute to precision horticulture by providing data-driven information for crop management, harvest planning, and efficient utilization of agricultural resources.

Keywords: machine learning; crop yield prediction; harvest timing; horticultural crops; environmental variability; precision agriculture; systematic literature review

Article Information
Received: October 05, 2026
Revised: November 05, 2026
Online: November 19, 2026

This work is licensed under a Creative Commons Attribution 4.0 International license
Agricultural Power Journal, November **2026**, Vol 3, No 4

1. Introduction

Horticultural crops such as tomato, grape, mango, apple, and strawberry have substantial economic value because their quality, productivity, and market price are strongly associated with appropriate cultivation and harvesting decisions. However, horticultural production is highly sensitive to variations in environmental conditions, including temperature, rainfall, humidity, soil characteristics, and other weather-related factors. These variations can affect crop growth, fruit development, yield formation, and maturity, creating challenges for producers in estimating expected production and determining the appropriate harvest period. Consequently, accurate prediction of crop yield and harvest timing is increasingly important for supporting production planning, resource allocation, harvesting operations, and postharvest management [1,2].

The development of artificial intelligence has provided new opportunities to improve agricultural prediction systems. In particular, machine learning (ML) can identify nonlinear relationships and complex patterns among crop, environmental, and historical production variables that may be difficult to capture using conventional statistical approaches. A systematic review by Shawon *et al.* reported that ML has increasingly been applied to agricultural yield prediction, with environmental and vegetation-related variables among the important predictors used in these models [1]. Similarly, a comprehensive review of fruit yield estimation highlighted the increasing application of machine learning, computer vision, and deep learning for estimating fruit production in orchard systems [3]. These developments indicate that data-driven approaches have considerable potential for improving the accuracy and timeliness of horticultural production prediction.

Recent empirical studies have demonstrated the application of ML to different high-value horticultural crops. Kasimatis *et al.* developed ML models to predict industrial tomato yield using historical production data, cultivated hybrids, and environmental conditions. Their model was trained using data from open-field cultivation and subsequently incorporated in-season weather information, demonstrating the potential of ML for yield prediction during the crop production period [2]. In mango production, Torgbor *et al.* integrated remote sensing and weather variables into an ML framework for yield prediction, demonstrating the value of combining spatial crop information with environmental data [4]. Similarly, Andrade *et al.* evaluated different ML algorithms for grape yield prediction using soil and weather-related variables, illustrating that environmental information can contribute to the prediction of yield in perennial horticultural systems [5]. These studies demonstrate that yield prediction can be improved by integrating multiple sources of crop and environmental information rather than relying exclusively on historical yield data.

This work is licensed under a Creative Commons Attribution 4.0 International license

Deep learning has further expanded the ability of agricultural prediction systems to process complex and high-dimensional data. Zheng *et al.* proposed a modified deep convolutional neural network for grape yield prediction, while von Bloh *et al.* introduced a knowledge-informed hybrid ML approach that combines ML with crop process information [6,7]. These approaches demonstrate two important directions in agricultural intelligence: the use of deep learning to extract complex patterns from crop-related data and the integration of data-driven models with agricultural knowledge. However, the effectiveness of these approaches may depend on crop type, environmental conditions, geographical location, data availability, and the timing at which observations are collected.

Although yield prediction is important, production management also requires information about when crops should be harvested. Yield and harvest timing represent related but distinct prediction problems. A crop may have a high expected yield but still require accurate maturity assessment to determine the appropriate harvesting period. Boechel *et al.* addressed this problem by developing a recurrent neural network (RNN)-based approach for predicting apple harvest time using temporal information associated with crop development and meteorological conditions [8]. More recently, Lin *et al.* developed a deep learning framework for strawberry harvest-date prediction using time-series images and ripeness information. Their study emphasized that many previous studies focused primarily on classifying generalized ripeness stages, whereas growers require more practical information concerning the specific timing of harvest [9]. This distinction is particularly important for high-value horticultural crops, for which harvesting too early or too late may affect fruit quality, market value, storage characteristics, and production efficiency.

Other studies have approached harvest timing through maturity and remote sensing information. Mimić *et al.* developed a ML approach for harvest-date detection and prediction using Sentinel-1 synthetic aperture radar (SAR) data, demonstrating the potential of remotely sensed temporal information for agricultural harvest monitoring [10]. Fruit maturity can also be assessed through characteristics associated with fruit size, color, spectral response, and internal quality. Takahashi *et al.* investigated tomato fruit-size prediction using several ML algorithms and environmental information, including temperature, to estimate fruit characteristics relevant to harvest [11]. Lee *et al.* applied hyperspectral imaging and analytical techniques to predict internal quality characteristics of strawberry fruit, illustrating how non-destructive sensing technologies can provide information relevant to maturity and harvest decisions [12]. In addition, deep learning has been applied to classify the maturity of mandarin oranges, showing the potential of image-based approaches for identifying fruit maturity stages [13].

Environmental variability remains a critical challenge for both yield and harvest-time prediction. Crop development does not occur under constant environmental conditions, and

This work is licensed under a Creative Commons Attribution 4.0 International license

changes in temperature, rainfall, humidity, and other meteorological factors can alter phenological development and the timing of maturity. This challenge becomes increasingly important when prediction models are applied across different locations or growing seasons. Chen *et al.* recently addressed this issue by coupling ML with a biophysical crop model to predict apple maturity dates across different apple-growing regions in China. Their study used phenological observations and meteorological data collected across multiple locations and years, highlighting the importance of accounting for environmental variability when predicting maturity dates [14]. The integration of ML with agricultural process knowledge provides an alternative to models that depend solely on historical observations from a specific production environment.

The emergence of sensor and Internet of Things (IoT) technologies also provides opportunities for obtaining environmental information in near real time. Mumtaz *et al.* proposed an IoT-enabled ML framework for strawberry yield prediction using environmental measurements and mobile decision support, demonstrating the potential for integrating continuous environmental monitoring with data-driven prediction [15]. Such systems can potentially improve prediction responsiveness because the model can incorporate updated environmental conditions during the crop growth period. Nevertheless, the integration of environmental monitoring with simultaneous prediction of production outcomes and harvest timing remains relatively limited, particularly for high-value horticultural crops.

Based on previous studies, three important research gaps can be identified. First, existing studies have extensively examined yield prediction using ML, but many of them focus on estimating production as a single outcome without explicitly incorporating harvest timing into the same predictive framework [1–7]. Second, studies addressing harvest timing or maturity prediction have generally concentrated on individual crops and specific data modalities, such as meteorological data, time-series images, or remote sensing data [8–14]. Third, although environmental variables are increasingly incorporated into agricultural ML models, there remains a need for an integrated framework capable of utilizing variable environmental conditions to support both yield estimation and harvest-timing prediction. This gap is particularly relevant to high-value horticultural production, where yield quantity and harvesting time jointly influence production efficiency and economic value.

Therefore, this study aims to synthesize, through a systematic literature review, how ML has been applied to predict yield and harvest timing of high-value horticultural crops under variable environmental conditions. Specifically, the review (i) maps the ML approaches and data sources used for yield and harvest timing prediction, (ii) examines how environmental and crop-related variables are incorporated into these models and why they matter, and (iii)

This work is licensed under a Creative Commons Attribution 4.0 International license

identifies research gaps and future directions toward an integrated framework that supports production planning, harvest scheduling, and decision-making in precision horticulture.

2. Materials and Method

This study employed a qualitative research approach using a systematic literature review (SLR) to examine the application of ML for crop yield and harvest timing prediction in high-value horticultural crops under variable environmental conditions. The systematic literature review approach was selected because it enables the synthesis of empirical findings from previous studies concerning agricultural yield prediction, harvest timing, crop maturity, ML applications, and environmental variability. This approach also provides a basis for identifying methodological developments, recurring findings, and research gaps related to data-driven prediction in horticultural production [1].

The literature search was conducted using internationally recognized scientific databases, including ScienceDirect, SpringerLink, MDPI, Frontiers, Wiley Online Library, and Google Scholar. The search process employed combinations of keywords such as *machine learning*, *crop yield prediction*, *horticultural crops*, *harvest timing prediction*, *harvest date prediction*, *crop maturity*, *environmental variability*, *deep learning*, *precision agriculture*, and *agricultural prediction*. Boolean operators such as "AND" and "OR" were used to improve the relevance of the retrieved publications.

The inclusion criteria consisted of peer-reviewed journal articles published between 2022 and 2026 that examined ML applications for horticultural crop yield prediction, harvest timing or maturity prediction, or the integration of environmental and crop-related variables. Studies focusing on unrelated agricultural applications, non-horticultural commodities without relevant predictive objectives, conference materials, and publications that did not provide sufficient methodological or empirical information were excluded from the review. Based on these criteria, fifteen relevant journal articles were selected for detailed analysis.

The selected studies covered several high-value horticultural crops, including tomato, grape, mango, apple, strawberry, and mandarin. The reviewed literature employed various ML and deep learning approaches, including Random Forest, Gradient Boosting, XGBoost, convolutional neural networks, recurrent neural networks, and hybrid ML models [2,4–11]. The diversity of crops and analytical approaches allowed the review to identify how ML has been applied across different production environments and prediction objectives.

The selected literature was analyzed using thematic content analysis. Each article was examined according to several analytical dimensions, including crop type, research objective, data sources, environmental variables, ML methods, prediction target, and principal findings. The extracted information was subsequently synthesized into three major themes: ML-based

This work is licensed under a Creative Commons Attribution 4.0 International license

crop yield prediction, ML-based harvest timing and maturity prediction, and the role of environmental variability in agricultural prediction. This procedure enabled comparison of findings across studies and identification of common patterns and limitations in the existing literature [4,7,14].

The synthesis further examined the extent to which environmental and crop-related variables were incorporated into prediction models. Variables such as temperature, rainfall, humidity, soil conditions, remote sensing information, crop growth characteristics, phenological stages, and fruit quality indicators were considered where reported by the reviewed studies [4,9,11,12]. Particular attention was given to studies that demonstrated the integration of multiple data sources because such integration is relevant to developing more comprehensive prediction frameworks for horticultural production under changing environmental conditions.

Finally, the findings from the fifteen selected studies were compared to identify research gaps and potential directions for future research. The comparison focused on the separation between yield prediction and harvest timing prediction, the diversity of environmental data used, the development of deep learning and hybrid approaches, and the potential integration of yield and harvest timing into a single predictive framework. The results of this synthesis were subsequently used to formulate the discussion and identify the potential contribution of an integrated ML-based approach for high-value horticultural crops.

3. Result

The synthesis of the fifteen selected studies indicates that ML has become an increasingly important approach for predicting crop yield, harvest timing, and maturity in high-value horticultural crops. The reviewed studies covered several major horticultural commodities, including tomato, grape, mango, apple, strawberry, and mandarin, with different environmental conditions and data characteristics. Overall, the literature demonstrates that ML can identify complex and nonlinear relationships between crop performance, environmental conditions, and crop development that are difficult to capture using conventional prediction approaches [1,7].

The reviewed studies can generally be categorized into three major research areas: crop yield prediction, harvest timing and maturity prediction, and the integration of environmental variables into ML models. Yield prediction studies predominantly focused on estimating production based on weather, remote sensing, crop characteristics, and historical production data. Meanwhile, harvest-related studies focused on predicting harvest dates, maturity stages, fruit size, and internal fruit quality using temporal, imaging, hyperspectral, and satellite-based information. The distribution of these research areas indicates that ML applications in

This work is licensed under a Creative Commons Attribution 4.0 International license

horticulture have expanded beyond yield estimation toward more precise crop development and harvest management.

3.1. Machine Learning-Based Crop Yield Prediction

The literature demonstrates that ML has been widely applied to estimate horticultural crop yield. Kasimatis *et al.* investigated ML models for industrial tomato yield prediction and demonstrated the potential of data-driven approaches for estimating tomato production [2]. Torgbor *et al.* combined remote sensing and weather variables in a ML approach for mango yield prediction, showing the importance of integrating environmental information into agricultural prediction [4]. Andrade *et al.* compared different ML algorithms for grape yield prediction and demonstrated that algorithm selection is an important consideration in developing yield prediction models [5].

The development of deep learning has further expanded the application of ML for crop yield prediction. Zheng *et al.* proposed a modified deep convolutional neural network for grape yield prediction, demonstrating the application of deep learning to complex agricultural data [6]. Von Bloh *et al.* developed a knowledge-informed hybrid ML approach that incorporated agricultural knowledge into yield prediction, indicating the potential of combining ML with domain-specific information [7].

The reviewed studies also demonstrate that yield prediction increasingly relies on multiple data sources. Weather observations, remote sensing information, crop characteristics, and historical production data have been incorporated into ML models to represent variations in agricultural production [4]. This indicates that environmental conditions are an important component in developing predictive models for horticultural yield.

3.2. Machine Learning-Based Harvest Timing and Maturity Prediction

The second major finding concerns the application of ML to harvest timing and crop maturity prediction. Boechel *et al.* developed an RNN-based approach for predicting apple harvest time, demonstrating the usefulness of recurrent neural networks for capturing temporal patterns in crop development [8]. Lin *et al.* applied multi-feature fusion deep learning to predict strawberry harvest dates in a plant factory, showing that multiple crop features can be combined to estimate harvest timing [9].

Mimić *et al.* investigated ML-based harvest date detection and prediction using SAR data, demonstrating the potential of satellite-based information for identifying agricultural harvest periods [10]. Takahashi *et al.* applied ML algorithms to predict tomato fruit size, providing information related to crop development and harvest readiness [11]. Lee *et al.* used organic

This work is licensed under a Creative Commons Attribution 4.0 International license

analysis and hyperspectral imaging to predict internal strawberry fruit quality, demonstrating the potential of spectral information for evaluating fruit conditions before harvest [12].

Singh *et al.* applied deep learning to classify mandarin orange maturity, indicating that ML can support the identification of maturity stages relevant to harvest decisions [13]. These studies demonstrate that harvest-related prediction has expanded beyond conventional calendar-based approaches toward the use of temporal, spectral, imaging, and crop-development information.

3.3. Environmental Variables in Horticultural Prediction

Environmental variability is an important component identified across the reviewed literature. Torgbor *et al.* demonstrated the integration of weather variables and remote sensing data for mango yield prediction, highlighting the contribution of environmental observations to agricultural forecasting [4]. Von Bloh *et al.* further demonstrated that incorporating agricultural knowledge into ML can improve the representation of relationships between environmental conditions and crop yield [7].

Chen *et al.* combined ML with a biophysical model to predict apple maturity dates across different apple-growing regions in China, demonstrating the potential of integrating environmental and biological processes into maturity prediction [14]. Mumtaz *et al.* developed an IoT-enabled ML framework for strawberry yield prediction with mobile decision support, indicating the increasing integration of real-time agricultural data and ML technologies [15].

He *et al.* emphasized that fruit yield estimation can utilize various direct and indirect data sources, including remote sensing and other technological approaches, reflecting the increasing diversity of information available for horticultural prediction [3]. Overall, the reviewed literature shows that environmental and crop-related variables provide important information for understanding variations in yield and crop development.

3.4. Synthesis of the Reviewed Literature

The synthesis identifies that existing studies have primarily concentrated on individual prediction objectives. Yield prediction has been examined extensively using weather, remote sensing, crop, and production variables [2]. In contrast, harvest timing and maturity prediction has generally emphasized temporal, imaging, spectral, and phenological information [8]. This separation indicates that yield estimation and harvest timing are commonly treated as distinct prediction problems. The reviewed studies also demonstrate differences in data sources and modeling approaches. Deep learning models have been increasingly applied to complex agricultural prediction problems [6], while remote sensing and SAR have been utilized to obtain spatial and environmental information [4]. IoT-based approaches have additionally introduced opportunities for incorporating real-time agricultural observations into ML systems [15].

This work is licensed under a Creative Commons Attribution 4.0 International license

Despite these developments, the literature indicates a limited integration of crop yield prediction and harvest timing prediction within a single framework. Chen *et al.* demonstrated the potential of combining ML with biophysical information for maturity prediction [14], while Mumtaz *et al.* demonstrated the integration of ML with IoT-based decision support for yield prediction [15]. However, these approaches address different prediction objectives. Therefore, an integrated approach that simultaneously considers environmental variability, crop development, yield, and harvest timing represents an area for further investigation.

4. Discussion

The findings of this systematic literature review indicate that ML offers substantial opportunities for improving yield and harvest timing prediction in high-value horticultural crops. Rather than restating the individual studies summarized in the Result section, this section interprets why the reviewed studies converge on the value of environmental information, how different modeling approaches address the resulting prediction challenges, and what the evidence implies for future prediction frameworks.

The first question is why integrating environmental variables benefits prediction. Across the reviewed literature, models that incorporate weather, soil, or remote sensing information are presented as an advance over approaches that rely on historical production data alone [2,4,5]. A plausible explanation lies in the biological role of these variables. Temperature, rainfall, humidity, and soil conditions are not merely statistical correlates of production; they are the drivers of the physiological processes that determine yield and maturity, including phenological development, flowering and fruit set, fruit growth, and ripening. Historical yield records capture only the net outcome of these processes in previous seasons. A model trained on such records alone cannot distinguish between seasons that produced similar yields through different environmental pathways, and it has limited ability to anticipate a season whose conditions fall outside the historical range. Environmental inputs supply the explanatory information that is otherwise missing, allowing a model to account for interannual variation rather than merely extrapolating from past outcomes. This interpretation is consistent with the tomato models of Kasimatis *et al.* [2], which incorporated in-season weather information, and with the mango framework of Torgbor *et al.* [4], which combined weather variables with remote sensing.

The second reason concerns how ML represents these relationships. The influence of an environmental variable on crop outcomes is rarely linear or independent of other factors: the same temperature may be favorable or damaging depending on the developmental stage, water availability, and cultivar, and its effects accumulate over time rather than acting instantaneously. Ensemble methods such as Random Forest, Gradient Boosting, and XGBoost

This work is licensed under a Creative Commons Attribution 4.0 International license

can capture threshold effects and interactions among variables without a predefined functional form, whereas recurrent architectures can represent the sequential accumulation of conditions during the growing season, as illustrated by the RNN-based apple harvest-time model [8]. Convolutional architectures, in turn, extract spatial and spectral patterns from imagery that summarize crop responses to environmental conditions [6,13]. These characteristics help explain the increasing adoption of ML in agricultural yield prediction [1], although the reviewed evidence does not allow the contribution of each model family to be isolated.

Third, environmental variability affects not only accuracy within a given season but also the transferability of models across locations and years. A model fitted to a single production environment risks learning site-specific correlations that do not hold elsewhere. The hybrid approaches identified in this review address this problem by anchoring data-driven learning in crop process knowledge. Von Bloh *et al.* [7] showed that incorporating agricultural knowledge into ML can improve the representation of the relationship between environmental conditions and yield, and Chen *et al.* [14] coupled ML with a biophysical model to predict apple maturity dates across several growing regions in China using phenological and meteorological data from multiple locations and years. A plausible interpretation is that exposure to multi-site and multi-season variability, combined with process-based constraints, pushes models toward relationships that are biologically meaningful and therefore more likely to hold under new conditions. This property is particularly important for high-value crops, for which inaccurate yield or harvest-time forecasts can affect fruit quality and market value.

Environmental variability also provides the conceptual link between yield and harvest timing, which the reviewed literature has largely treated separately [2,8]. The same drivers that govern the rate of crop development, and therefore the timing of maturity, also influence fruit set, fruit growth, and biomass accumulation, and therefore the quantity harvested. Yield and harvest date are consequently coupled outcomes of shared environmental processes rather than independent targets. Modeling them separately discards this shared information and may produce mutually inconsistent forecasts, for example a high predicted yield at a harvest date that the predicted crop development does not support. This coupling, more than the mere availability of diverse data, is the strongest argument for an integrated, multi-output framework.

Digital sensing technologies strengthen this argument by increasing the temporal and spatial resolution of the environmental information available to models. Remote sensing and SAR data provide spatially extensive information on crop and field conditions [4,10], non-destructive imaging captures fruit-level indicators of quality and maturity [12,13], and IoT-enabled frameworks such as that of Mumtaz *et al.* [15] allow updated measurements to be incorporated during the growing season. He *et al.* [3] similarly summarized the wide range of direct and

This work is licensed under a Creative Commons Attribution 4.0 International license

indirect technologies available for fruit yield estimation. Together, these developments move ML from retrospective estimation toward in-season decision support, although sensor calibration, missing observations, and the cost of data acquisition may constrain their adoption in practice.

These interpretations should be read with caution. The reviewed studies differ in crops, regions, data sources, validation procedures, and performance metrics, and this review did not pool effect sizes or formally appraise study quality. The consistent direction of the evidence therefore supports the conclusion that environmental information is beneficial, but it does not quantify how large the benefit is. In addition, studies reporting successful predictions are more likely to be published, which may overstate the benefit of integrating environmental variables.

Two research gaps follow from this synthesis. First, most studies focus on a single prediction objective, either yield estimation or harvest timing and maturity prediction: yield-oriented studies emphasize production, weather, and remote sensing variables, whereas harvest-oriented studies emphasize temporal, spectral, imaging, or maturity-related characteristics [2,8]. Second, the use of multiple information sources is not consistently integrated, as some studies emphasize remote sensing while others focus on imaging, weather observations, crop characteristics, or IoT data [3,4]. Addressing both gaps requires frameworks in which environmental, temporal, crop-development, and production variables are modeled jointly. Such an approach would connect production forecasting with harvest management rather than treating them as independent tasks, and it may provide a more comprehensive basis for precision horticulture and more efficient agricultural planning.

5. Conclusions

This systematic literature review shows that ML has considerable potential for predicting crop yield, harvest timing, and maturity in high-value horticultural crops. The fifteen reviewed studies applied ML to various crops and data sources, including weather observations, remote sensing, SAR, hyperspectral imaging, crop characteristics, temporal information, and IoT-generated data [4].

The findings indicate that environmental variability is an important consideration in horticultural prediction because environmental conditions drive both crop productivity and crop development. Advanced ML and deep learning approaches provide opportunities to capture the nonlinear and cumulative relationships between environmental conditions and crop outcomes [6]. However, most existing studies continue to examine yield prediction and harvest timing as separate objectives, which limits their ability to represent the interdependence of the two outcomes.

This work is licensed under a Creative Commons Attribution 4.0 International license

From a practical perspective, integrating environmental, crop-development, temporal, and production information into a unified predictive framework could support more comprehensive decision-making in horticultural production. Such a framework may assist producers and agricultural stakeholders in improving production planning, estimating expected yield, and determining appropriate harvest periods based on changing crop and environmental conditions [15].

Overall, the literature supports the continued development of ML-based approaches for precision horticulture. Future integration of environmental monitoring, crop development information, remote sensing, and advanced predictive algorithms may contribute to more adaptive and data-driven horticultural production systems.

This study is limited by its reliance on previously published literature, the small number of included studies (fifteen), the restriction to peer-reviewed articles published between 2022 and 2026, and the differences in datasets, crops, environmental conditions, and ML approaches across the reviewed studies. In addition, no formal quality appraisal or quantitative pooling of model performance was conducted. Therefore, the findings should be interpreted as a synthesis of existing evidence rather than a direct comparison of model performance.

Based on these findings, future research should empirically test and compare a limited set of ML architectures within an integrated framework for simultaneous yield and harvest timing prediction, using standardized datasets. Four groups of models appear to hold the greatest potential. First, gradient-boosted decision trees (e.g., XGBoost or LightGBM) and Random Forest are suitable as strong and interpretable baselines for tabular weather, soil, and crop data, and the contribution of each variable can be quantified using importance measures such as SHAP. Second, sequence models such as long short-term memory (LSTM) or gated recurrent unit (GRU) networks and Transformer-based temporal models are promising for season-long weather and crop-development time series, building on the RNN-based approach used for apple harvest-time prediction [8] and on the cumulative nature of environmental effects. Third, multi-task deep learning architectures with a shared encoder and separate output layers for yield (regression) and harvest date (regression or time-to-event formulation) directly address the separation of the two objectives identified in this review. Fourth, hybrid process-guided models that couple ML with crop growth or phenology models [7,14], and multimodal architectures that fuse an image or remote sensing branch (e.g., a convolutional neural network) with a temporal branch for weather and IoT data, are promising for improving transferability across locations and seasons. Model comparison should use spatial and temporal validation (e.g., leave-one-season-out and leave-one-site-out), common metrics such as RMSE, MAE, and R^2 for yield and MAE in days for harvest date, and uncertainty quantification to support decision-making.

This work is licensed under a Creative Commons Attribution 4.0 International license
Agricultural Power Journal, November **2026**, Vol 3, No 4

References

1. Shawon, S.M.; Ema, F.B.; Mahi, A.K.; Niha, F.L.; Zubair, H.T. Crop yield prediction using machine learning: An extensive and systematic literature review. Smart Agric. Technol. 2025, 10, 100718. https://doi.org/10.1016/j.atech.2024.100718

2. Kasimatis, C.N.; Psomakelis, E.; Katsenios, N.; Papatheodorou, M.; et al. Industrial tomato yield prediction using machine learning models. Smart Agric. Technol. 2025, 11, 100920. https://doi.org/10.1016/j.atech.2025.100920

3. He, L.; Fang, W.; Zhao, G.; Wu, Z.; Fu, L.; Li, R.; Majeed, Y.; Dhupia, J. Fruit yield prediction and estimation in orchards: A state-of-the-art comprehensive review for both direct and indirect methods. Comput. Electron. Agric. 2022, 195, 106812. https://doi.org/10.1016/j.compag.2022.106812

4. Torgbor, B.A.; Rahman, M.M.; Brinkhoff, J.; Sinha, P.; Robson, A. Integrating remote sensing and weather variables for mango yield prediction using a machine learning approach. Remote Sens. 2023, 15, 3075. https://doi.org/10.3390/rs15123075

5. Andrade, C.B.; Moura-Bueno, J.M.; Comin, J.J.; Brunetto, G. Grape yield prediction models: Approaching different machine learning algorithms. Horticulturae 2023, 9, 1294. https://doi.org/10.3390/horticulturae9121294

6. Zheng, S.; Gao, P.; Zhang, J.; Ma, Z.; Chen, S. A precise grape yield prediction method based on a modified DCNN model. Comput. Electron. Agric. 2024, 225, 109338. https://doi.org/10.1016/j.compag.2024.109338

7. von Bloh, M.; Lobell, D.; Asseng, S. Knowledge informed hybrid machine learning in agricultural yield prediction. Comput. Electron. Agric. 2024, 227, 109606. https://doi.org/10.1016/j.compag.2024.109606

8. Boechel, T.; Policarpo, L.M.; Ramos, G.d.O.; Righi, R.d.R.; Singh, D. Prediction of harvest time of apple trees: An RNN-based approach. Algorithms 2022, 15, 95. https://doi.org/10.3390/a15030095

9. Lin, Z.; Liu, W.; Wang, S.; Pan, J.; Fu, R.; Chen, T.; Lin, T. Strawberry harvest date prediction using multi-feature fusion deep learning in plant factory. Comput. Electron. Agric. 2025, 234, 110174. https://doi.org/10.1016/j.compag.2025.110174

This work is licensed under a Creative Commons Attribution 4.0 International license

10. Mimić, G.; Mishra, A.K.; Marković, M.; Živaljević, B.; Pavlović, D.; Marko, O. Machine learning-based harvest date detection and prediction using SAR data for the Vojvodina region (Serbia). Sensors 2025, 25, 2239. https://doi.org/10.3390/s25072239

11. Takahashi, M.; Kawasaki, Y.; Naito, H.; Lee, U.; Yoshi, K. Fruit size prediction of tomato cultivars using machine learning algorithms. Front. Plant Sci. 2025, 16, 1516255. https://doi.org/10.3389/fpls.2025.1516255

12. Lee, S.-D.; Gil, C.-S.; Lee, J.-H.; Jeong, H.-B.; Kim, J.-H.; Jang, Y.-A.; Kim, D.-Y.; Lee, W.-M.; Moon, J.-H. Internal quality prediction technology for 'Sulhyang' strawberry fruit using organic analysis and hyperspectral imaging. Spectrochim. Acta A Mol. Biomol. Spectrosc. 2024, 323, 124912. https://doi.org/10.1016/j.saa.2024.124912

13. Singh, R.; Nickhil, C.; Upendar, K.; Mishra, P.; Deka, S.C. Deep learning-based approach for classifying mandarin orange maturity. Sustain. Food Technol. 2025, 3, 2215–2225. https://doi.org/10.1039/D5FB00408J

14. Chen, R.; Wang, J.; Gong, Z.; Zhou, J.; Li, Y.; Bai, R.; Qu, Z.; Le, Z. Coupling machine learning with a biophysical model for maturity date prediction of apple fruit across China's apple planting regions. J. Sci. Food Agric. 2026. https://doi.org/10.1002/jsfa.70943

15. Mumtaz, A.; Mohd Noor, A.S.; Khan, M.R.; Bano, A.; Muhaimin, A.; Mohamad, M.S.; Yahya, C.A.C. IoT-enabled machine learning framework for strawberry yield prediction with mobile decision support. Procedia Comput. Sci. 2026, 280, 606–613. https://doi.org/10.1016/j.procs.2026.04.077