Computational Molecular Biology 2026, Vol.16, No.4, 267-279
http://bioscipublisher.com/index.php/cmb

Research Insight Open Access

Prediction Model Construction for Bayberry Fruit Quality Based on Environmental and Cultivation Data

Weiping Wu[1,2]
1 Hangzhou Yinghe Jiatian Technology Co., Ltd., Hangzhou, 310056, Zhejiang, China
2 Zhejiang Agronomist College, Hangzhou, 310021, Zhejiang, China
Corresponding email: wwp729@163.com

Computational Molecular Biology, 2026, Vol.16, No.4 doi: 10.5376/cmb.2026.16.0019
Received: 22 Jun., 2026
Accepted: 30 Jul, 2026
Published: 15 Aug., 2026

Preferred citation for this article:
Wu W.P., 2026, Prediction model construction for bayberry fruit quality based on environmental and cultivation data, Computational Molecular Biology, 16(4): 267-279 (doi: 10.5376/cmb.2026.16.0019)

Abstract Bayberry (*Myrica rubra*) is an economically important fruit crop, and its commercial value is largely determined by quality attributes such as soluble solids content, acidity, firmness, color, and bioactive compounds. However, fruit quality formation is strongly influenced by complex interactions between environmental conditions and cultivation practices, making accurate prediction challenging under variable production environments. This review focuses on the construction of prediction models for bayberry fruit quality based on environmental and cultivation data. The mechanisms underlying quality formation are first summarized, including the regulation of temperature, precipitation, solar radiation, soil conditions, irrigation, and nutrient management on fruit development and biochemical accumulation. Subsequently, approaches for data acquisition, feature engineering, and integration of multi-source agricultural information are discussed. Various prediction methods, including statistical models, machine learning algorithms, and deep learning approaches, are evaluated with emphasis on model optimization, validation strategies, and interpretability analysis. A case study framework is presented to demonstrate the application of predictive models in orchard monitoring, quality forecasting, and precision management. Furthermore, the integration of prediction models with digital agriculture technologies, such as IoT sensors, remote sensing, and intelligent decision-support systems, is explored. Future research should emphasize multi-dimensional data integration, explainable artificial intelligence, and region-specific model adaptation to improve prediction accuracy and practical applicability. These advances will provide theoretical foundations and technical support for intelligent management and sustainable development of high-quality bayberry production.

Keywords Bayberry; Fruit quality prediction; Environmental factors; Machine learning; Precision horticulture

1 Introduction

Chinese bayberry (*Myrica rubra*) is a commercially important subtropical fruit crop in China, with cultivation extending back more than 2000 years and major production concentrated in Zhejiang and Jiangsu, where fruit quality strongly affects market value, consumer acceptance, and downstream processing potential. Bayberry fruit is valued for its attractive color, distinctive flavor, and rich phytochemical composition, especially soluble sugars, organic acids, phenolics, and anthocyanins, but it is also highly perishable, with very short storage life after harvest, making accurate preharvest quality prediction especially important for harvest scheduling, grading, marketing, and postharvest management (Mo et al., 2024). In modern horticultural production, this need is becoming more urgent because producers must make timely cultivation and marketing decisions under variable weather conditions while meeting increasing demands for stable fruit appearance, taste, and nutritional quality. A predictive framework that estimates fruit quality before harvest could therefore improve production efficiency, reduce uncertainty, and support more precise orchard management for a crop whose commercial window is narrow and whose quality deteriorates rapidly once harvested.

Bayberry fruit quality is formed through the combined action of genotype, environment, and cultivation practice, and available studies show that environmental regulation is particularly strong during fruit development and ripening. Fruit quality development is broadly modulated by environmental factors through signal transduction pathways linking environmental perception to the regulation of enzymes involved in pigmentation, sugar metabolism, organic acid accumulation, and antioxidant formation (Zhai et al., 2026). In bayberry specifically, light exposure has a pronounced effect on coloration and internal quality, as reduced light transmission during development sharply decreased sugars, organic acids, flavonoids, vitamin C, and anthocyanins, while also impairing normal coloration (Yang et al., 2025). These findings indicate that quality traits such as sweetness, acidity balance, and peel color are not isolated outcomes, but integrated physiological responses to orchard microclimate. As a result, environmental observations such as light, temperature, rainfall, and humidity, together with developmental timing, are biologically meaningful predictors for quality modeling in bayberry production systems.

Cultivation management further reshapes the orchard environment and can substantially alter bayberry fruit quality, which makes management data as important as meteorological data for predictive purposes. In southern China, bayberry harvest often coincides with rainy, humid, and low-sunshine conditions that reduce fruit quality and economic returns, whereas greenhouse cultivation can avoid rain damage, modify the fruiting microenvironment, and significantly improve fruit weight, soluble solids, sugar-acid ratio, and sucrose accumulation relative to open-field production (Wu et al., 2021). Likewise, orchard floor management influences belowground conditions linked to fruit performance: three years of accompanying ryegrass improved bayberry fruit quality parameters and was associated with shifts in soil physicochemical properties, rhizosphere microbial community structure, and metabolite profiles (Li et al., 2023). Together, these results show that fruit quality emerges from a multi-factor system in which cultivation mode, canopy and soil management, and local environmental conditions jointly shape the final commercial phenotype. This complexity also explains why single-factor evaluation is insufficient for modern decision-making and why multivariable prediction models are needed.

Against this background, constructing a bayberry fruit quality prediction model based on environmental and cultivation data is both scientifically justified and practically valuable. Recent horticultural research shows that machine-learning approaches can successfully predict fruit quality traits from large, multi-season datasets integrating cultivar, location, soil, and meteorological information, with XGBoost performing particularly well for tomato quality prediction and highlighting the importance of model interpretability and environmental feature effects. Broader reviews likewise indicate that machine-learning methods are increasingly effective in horticultural prediction and classification tasks, while also emphasizing the need to test multiple algorithms and data combinations to identify robust models for specific production systems. Therefore, the objective of this study is to integrate key environmental variables and cultivation-management indicators to construct an accurate and interpretable prediction model for bayberry fruit quality. The research framework is expected to include the acquisition of orchard environmental and management data, the measurement of representative fruit quality traits, feature screening and model training, comparison of candidate algorithms, and evaluation of predictive performance, thereby providing a technical basis for precision cultivation and quality-oriented management of bayberry orchards.

2 Biological Mechanisms Underlying Bayberry Fruit Quality Formation

2.1 Physiological processes regulating fruit growth and quality development

Bayberry fruit quality develops through coordinated changes in appearance, flavor, texture, and phytochemical composition during fruit growth and ripening. Fruit quality is a complex trait defined by external attributes such as size, shape, and color together with internal traits including sugar, acidity, and amino acid levels, and these traits shift rapidly as bayberry progresses toward maturity (Zhang et al., 2024). Across ripening stages, full-ripe fruit shows the highest fruit weight, total soluble solids, sugar content, and anthocyanin accumulation, while firmness declines and the volatile profile changes substantially, indicating that quality formation is a dynamic developmental process rather than a static endpoint.

At the molecular level, bayberry ripening is controlled by large transcriptional and metabolic reprogramming networks. RNA-seq analyses detected more than 3,600 differentially expressed genes during ripening, with upregulation of all anthocyanin biosynthetic genes alongside color development and important changes in carbohydrate and acid metabolism associated with genes such as sucrose phosphate synthase and glutamate decarboxylas. Integrated transcriptomic-metabolomic work further identified flavonoid biosynthesis and starch-sucrose metabolism as core pathways in ripening, showing that fruit quality emerges from coordinated gene-metabolite interactions rather than from single biochemical traits alone (Sun et al., 2024).

2.2 Effects of environmental conditions on bayberry fruit quality

Environmental conditions are major drivers of fruit quality formation because they regulate the metabolic pathways underlying sugar accumulation, acid turnover, pigmentation, and antioxidant synthesis. More broadly in fruit crops, environmental modulation operates through signaling processes that connect external cues to the transcriptional or post-transcriptional regulation of quality-related enzymes (Zhai et al., 2026). In bayberry, light is one of the clearest examples: restricting light during development sharply reduced sucrose, glucose, fructose, organic acids, flavonoids, vitamin C, and total anthocyanins while increasing acidity, showing that orchard radiation conditions directly shape both visual and eating quality (Yang et al., 2025).

Light effects are mediated by specific pigment-regulatory networks, which helps explain why environmental data can support mechanistic prediction of fruit quality. Fruit bagging strongly inhibited anthocyanin accumulation and suppressed the expression of anthocyanin biosynthetic genes together with the transcription factor MrMYB1, while newer omics work identified HY5 as a key regulator in light-mediated anthocyanin accumulation (Yang et al., 2025). Temperature also alters quality trajectories after harvest: lower storage temperatures slowed respiration and ethylene production, preserved firmness, total soluble solids, and titratable acidity, and delayed the decline of citric acid and sucrose, indicating that thermal conditions strongly influence quality maintenance and postharvest expression of ripening physiology.

2.3 Influence of cultivation management on quality variation

Cultivation management modifies the microenvironment in which bayberry fruit develops, so management variables can become direct determinants of final quality variation. Greenhouse cultivation improves fruit weight, size, soluble solids, and sugar-acid ratio compared with open-field production, and these effects appear to be linked to enhanced sucrose accumulation and changes in sucrose-phosphate synthase and acid invertase activities (Wu et al., 2021). Supplemental LED lighting can further improve quality under protected cultivation, increasing fruit weight, soluble solids, and vitamin C while reducing organic acid content in responsive cultivars, although the magnitude of response differs by genotype.

Soil and orchard-floor management also contribute to quality variation by altering nutrient status, rhizosphere biology, and fruit-associated microbial communities. Accompanying ryegrass increased bayberry fruit sugar, vitamin C, and flavonoid contents while lowering titratable acidity, and these changes were associated with shifts in soil physicochemical properties, microbial structure, and metabolite composition (Li et al., 2023). Likewise, insect- and rain-proof netting increased fruit diameter, weight, edible rate, and Brix/acid ratio, while also changing bacterial communities in ways that likely reduced sugar consumption and disease pressure, showing that management affects quality through both physical protection and biological regulation (Yu et al., 2021).

3 Data Acquisition and Feature Engineering for Bayberry Quality Prediction

3.1 Collection of environmental and climatic data

Environmental and climatic data should be collected as core explanatory variables because fruit prediction models in orchard systems commonly depend on both on-site measurements and broader-area interpolated weather records. In perennial fruit production, predictive datasets can combine local station data, orchard sensor grids, and external databases, while cumulative or stage-specific environmental features often improve model performance by better matching plant development windows (Niedbała et al., 2022). For bayberry quality prediction, this supports the routine acquisition of air temperature, precipitation, relative humidity, solar radiation, and derived agroclimatic indices across flowering, expansion, coloring, and harvest stages, rather than relying only on single-date observations. Historical evidence from fruit systems also shows that harvest date, soluble solids, and other quality traits become more interpretable when linked with long-term meteorological records and bioclimatic indicators collected across contrasting production environments.

The environmental dataset should also preserve temporal resolution and spatial identity, because fruit quality varies by orchard block, production season, and storage or harvest environment. Orchard research defines a management block by shared genotype, planting date, irrigation scheduling, management history, and harvest unit, which means climate variables should be linked to these same units during database construction (Anderson et al., 2021). In bayberry and related fruit studies, temperature has repeatedly shown strong explanatory value for quality change: different storage temperatures altered flavor deterioration trajectories in ‘Dongkui’ fruit, and meteorological indices were the strongest predictors in lowbush blueberry modeling, reinforcing the importance of temperature-based features and time-integrated weather variables in predictive design (Figure 1) (Gao et al., 2024).

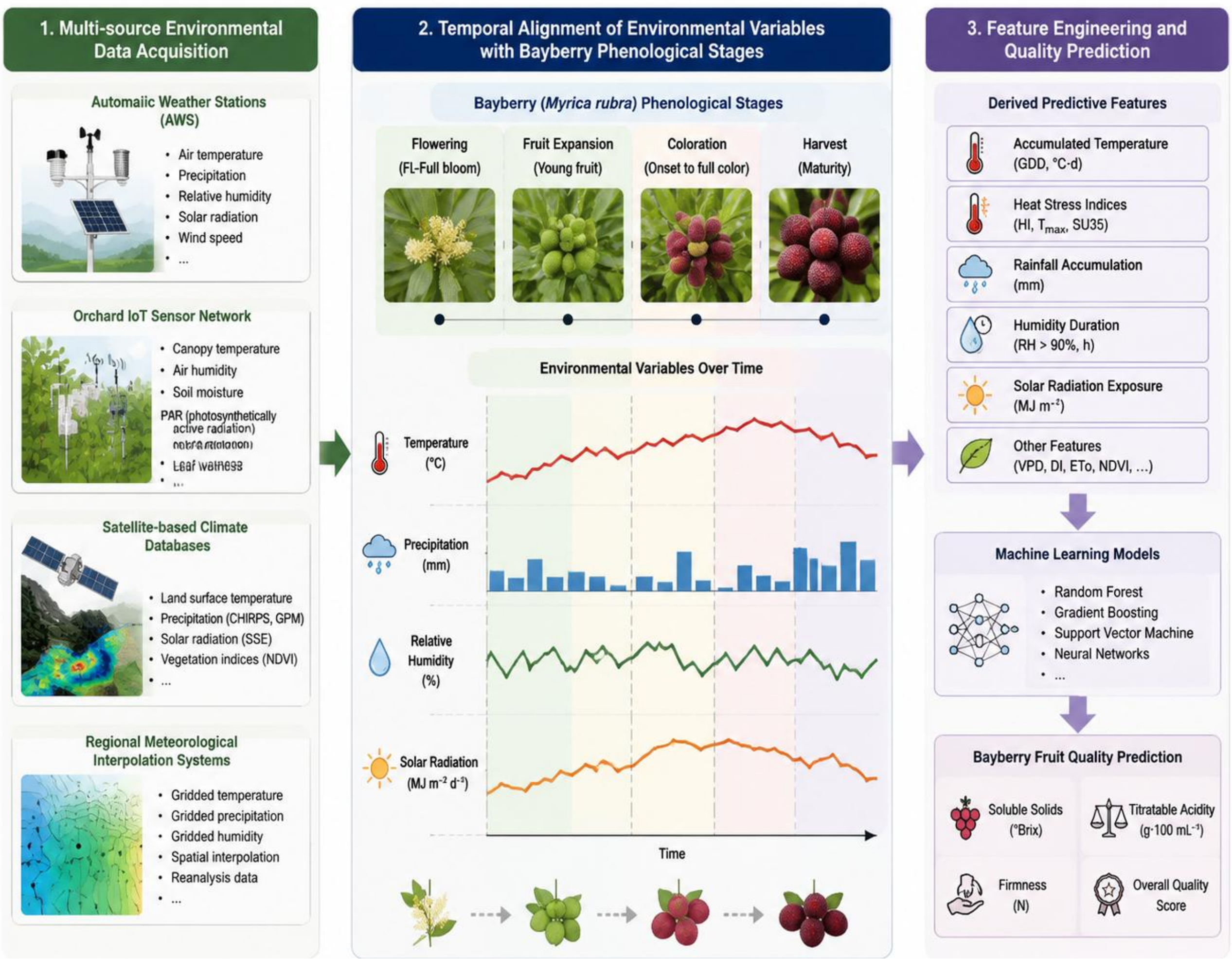

Figure 1 Integration of multi-source environmental data for bayberry fruit quality prediction across different phenological stages

3.2 Collection of cultivation and orchard management data

Cultivation and orchard management data should be collected with the same rigor as weather data because management modifies the biological context in which fruit quality forms. In orchard systems, a management block is distinguished not only by genotype but also by irrigation scheduling, management history, and harvest organization, indicating that records of cultivar, tree age, planting density, irrigation regime, fertilization, pruning, and harvest date are fundamental structured inputs for modeling (Anderson et al., 2021). A modern orchard database should therefore capture these variables in a standardized digital framework that supports storage, integration, and traceability across in-orchard and postharvest measurements, reducing the fragmentation that often limits model transferability (Williams et al., 2023).

For bayberry specifically, management records should also include ecological and protective practices that can reshape fruit quality through soil, microbial, and canopy effects. A field study on accompanying ryegrass documented the need to record treatment history, sampling position, soil depth, and replicate-tree structure when evaluating quality responses, because the practice altered fruit economic traits together with rhizosphere physicochemical and microbiological properties (Li et al., 2023). Postharvest and fruit-associated biological conditions are also relevant management-linked data sources: microbiome composition differed across bayberry cultivars and orchards, and bacterial variation was orchard-dependent, suggesting that orchard identity, cultivar, and handling environment should be retained as categorical variables when constructing a quality prediction dataset (Xie et al., 2025).

3.3 Fruit quality indicators and data preprocessing

Fruit quality indicators should cover both external and internal traits so that the target variables reflect the multidimensional nature of bayberry quality. Bayberry quality includes external traits such as shape, size, and color as well as internal traits such as sugar, acidity, and amino acid composition, while recent nondestructive studies have shown that weight and soluble solids content can be predicted effectively from hyperspectral image descriptors and spectra across batches and packaging states (Zhang et al., 2024). Additional indicators can include maturity color indices, since image-based maturity modeling in bayberry used the a*/b* ratio as ground truth and showed that fused color and texture features can support accurate maturity estimation, making such indicators useful either as targets or engineered predictors (Kai et al., 2021; Zheng et al., 2025).

Data preprocessing should address heterogeneity, noise, and multimodal measurement structure before model training. Prior studies in fruit prediction used normalization, missing-data and outlier analysis, cumulative feature generation, feature selection, and dimensionality reduction such as PCA to reduce noise and improve predictive robustness (Niedbała et al., 2022). For bayberry quality sensing, preprocessing is especially important because different wavebands, absorbance transformations, and calibration-transfer strategies can materially affect prediction accuracy: Vis/NIR and NIR studies used spectral preprocessing before PLS modeling, and hyperspectral work used slope-bias correction to reduce deviation across batches and acquisition modes (Yuan et al., 2025).

4 Construction of Machine Learning-Based Bayberry Fruit Quality Prediction Models

4.1 Selection of statistical and machine learning algorithms

The choice of algorithm should match the multidimensional and nonlinear nature of bayberry quality data. Classical linear models remain useful as baselines because they are simple, transparent, and effective for some targets, as shown by successful multiple linear regression for bayberry weight prediction and by the continued use of linear regression benchmarks in multimodel fruit-quality studies (Yuan et al., 2025). However, fruit quality prediction often involves nonlinear interactions among environmental, cultivation, and phenotypic variables, which is why tree-based and kernel-based methods such as Random Forest, Support Vector Machine, Gradient Boosting, XGBoost, LightGBM, and CatBoost are appropriate candidate models for bayberry quality prediction.

Comparative evidence indicates that no single algorithm is universally best across all fruit-quality tasks, so model construction should begin with parallel testing of several statistical and machine learning approaches rather than committing to one method in advance (Ribeiro et al., 2022). At the same time, boosting and ensemble methods often perform especially well in structured agricultural datasets: Gradient Boosting achieved the lowest error in pear quality evaluation, while stacking regression outperformed other techniques in wild blueberry prediction, supporting the inclusion of ensemble learners as strong candidates in bayberry model development (Ping et al., 2025).

4.2 Model training, optimization, and validation strategies

Model training should follow a standardized pipeline that includes preprocessing, feature reduction, calibration, tuning, and independent evaluation. Studies in fruit-quality and crop-prediction modeling consistently apply normalization, feature selection, and dimensionality reduction before prediction, and several frameworks explicitly separate preprocessing, feature selection, and prediction into sequential phases to improve robustness and reduce redundancy (Abdel-Salam et al., 2024; Ping et al., 2025). For bayberry datasets that combine environmental records, management variables, and fruit traits, this suggests first cleaning missing values, scaling continuous features, encoding categorical management variables, and then constructing reduced feature sets for subsequent model training (Jewan et al., 2024).

Hyperparameter optimization and validation are central because fruit datasets are often limited in size and prone to overfitting. Ten-fold cross-validation has been used to optimize multimodal grape models, and Bayesian optimization has improved both parameter search and final predictive performance in large-dimensional quality prediction systems and fruit-classification pipelines (Jewan et al., 2024). Where sample size is constrained, repeated experiments or repeated resampling can further stabilize performance estimates, as shown by blueberry sensory modeling repeated 20 times, while data augmentation or model transfer may be needed for image or spectral submodels that must generalize across batches, orchards, or packaging conditions (Yuan et al., 2025).

4.3 Model interpretation and identification of key predictors

Interpretability is essential because the practical value of a bayberry quality model depends not only on predictive accuracy but also on its ability to reveal which environmental and cultivation variables drive quality variation. Recent work in apple quality assessment used LIME and SHAP to explain model decisions, while grapevine prediction studies explicitly ranked predictor importance after validation, showing that modern fruit-quality modeling can combine strong performance with biologically meaningful interpretation (Chowdhury et al., 2024; Jewan et al., 2024). In a bayberry context, this interpretive step can identify whether temperature, humidity, light exposure, cultivar, irrigation, harvest timing, or canopy-related features are the dominant contributors to soluble solids, acidity, size, or maturity class (Zheng et al., 2025).

Feature-importance analysis also supports model simplification by showing when a smaller subset of predictors retains most of the predictive signal. Random Forest and XGBoost have been used to select the most relevant physicochemical features in wine-quality modeling, and Random Forest combined with Bayesian optimization identified fewer but critical features while preserving satisfactory forecasting accuracy and computational efficiency (Wang et al., 2023). For bayberry quality prediction, this means the final framework should not stop at reporting accuracy metrics alone, but should also extract a compact set of key predictors that can guide orchard monitoring, reduce data-collection costs, and improve the interpretability and transferability of the model.

5 Case Study: Development and Application of a Bayberry Fruit Quality Prediction Model

5.1 Experimental orchard design and data collection

The case-study orchard should be defined as a management block with consistent genotype, planting date, irrigation scheduling, management history, and harvest unit so that environmental and cultivation variables correspond to a biologically coherent prediction unit (Anderson et al., 2021). To improve dataset representativeness, sampling should span multiple orchards, cultivars, and collection dates, as bayberry quality prediction studies have already drawn fruit from different regions and batches, and bayberry biochemical calibration studies likewise used different cultivars from different areas in China (Yuan et al., 2025). Within each orchard, fruit sampling should follow a structured biological design with tree-level replication and canopy-position control, because bayberry microbiome work successfully used six biological replicates per orchard-cultivar combination and collected fruits from the outer canopy in each cardinal direction at harvest maturity (Xie et al., 2025).

The data collection system should integrate environmental observations, management logs, and fruit phenotypes in a standardized digital framework, because orchard research increasingly depends on comprehensive data ecosystems for storage, integration, processing, and traceability (Williams et al., 2023). For a bayberry case study, the recorded predictors should include time, orchard identity, temperature, humidity, and management operations such as fertilization, pruning, spraying, thinning, or mulching, since orchard-level machine-learning systems in other fruit crops have successfully combined these variables for fruit-quality prediction (Kim and Heo, 2024). Response variables should include both destructive and nondestructive quality traits, such as weight, soluble solids, titratable acidity, organic acids, maturity color, and image or spectral descriptors, because bayberry studies have built predictive datasets around these same endpoints (Figure 2).

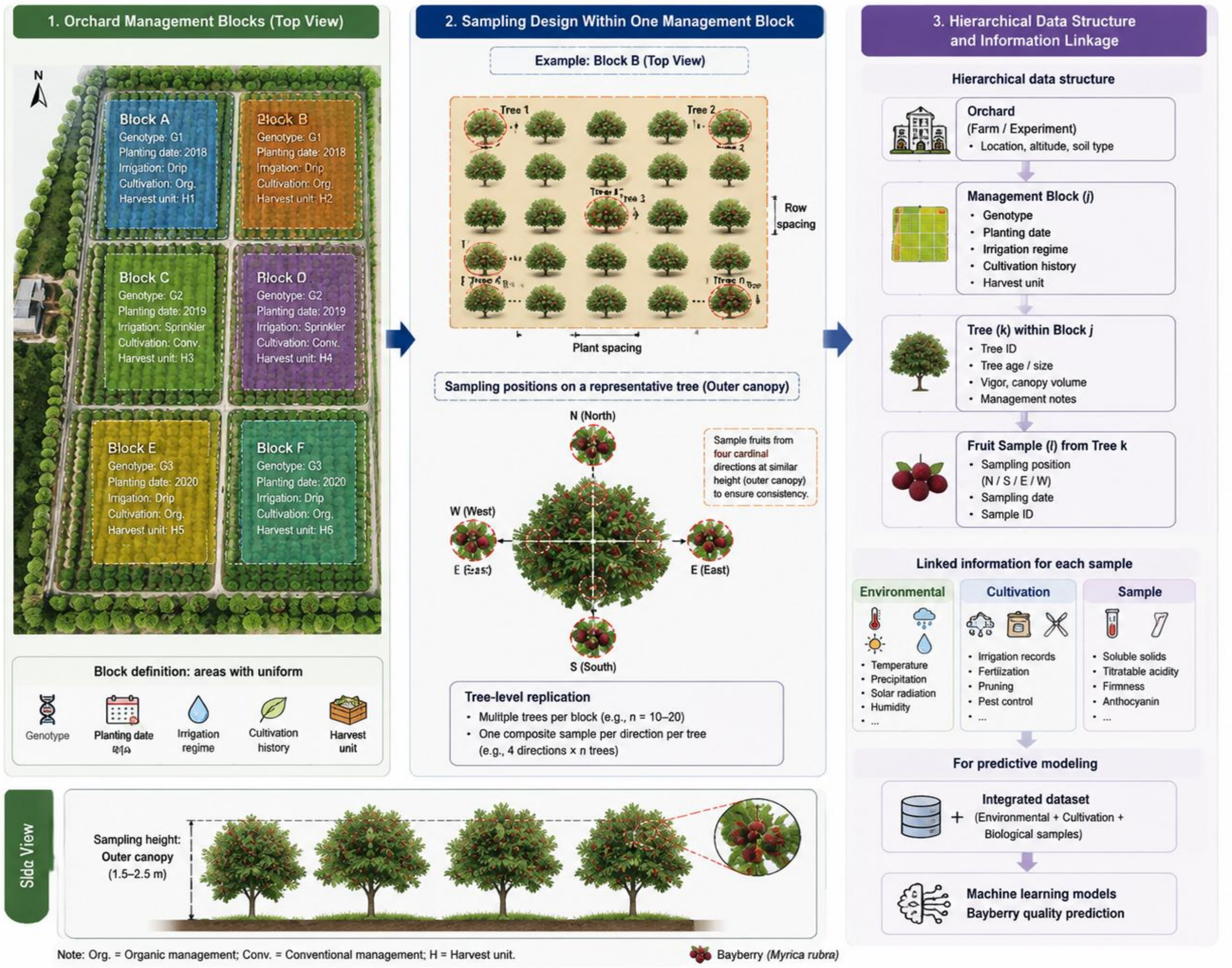

Figure 2 Hierarchical sampling design and data integration framework for bayberry (Myrica rubra) orchard management and yield prediction

5.2 Model performance evaluation and prediction accuracy analysis

Model performance should be evaluated with separate training and testing sets and reported using complementary metrics rather than a single accuracy value. Image-based fruit-quality studies have used a 70:30 train-test split and assessed models with R2, MAE, and RMSE, while hyperspectral studies of bayberry similarly reported RMSEP as a key measure for weight and soluble solids prediction (Júnior et al., 2025). In a bayberry case study, several algorithms should be compared in parallel, because fruit-quality prediction performance varies with target trait and data structure, and large comparative studies routinely test linear, kernel, tree-based, and neural models side by side. This is especially important because some bayberry targets are easier to predict than others: weight prediction reached RMSEP below 0.6 and soluble solids prediction below 0.8 in hyperspectral work, whereas acid-related traits showed more moderate accuracy in earlier NIR calibration (Yuan et al., 2025).

Validation should also test robustness across batches, orchards, and acquisition conditions, because transferability is a practical bottleneck for deployment. In bayberry, slope-bias correction improved predictions across different batches and packaging modes, and orchard-level fruit modeling in mandarin showed that fruit size was predicted substantially better than sugar content, implying that some quality traits require finer-scale or additional predictors (Yuan et al., 2025; Kim and Heo, 2024). A strong case study would therefore include external validation beyond the original orchard and discuss data limitations explicitly, since small and homogeneous datasets reduce generalizability and can miss interannual, regional, and management-driven variation. Where labeled data are scarce, spectral-fusion and self-supervised strategies are promising because they improved downstream prediction with far fewer labeled samples in fruit chemistry modeling (Zhang et al., 2024).

5.3 Application of prediction models for precision orchard management

Once validated, the prediction model can support precision orchard management by converting preharvest data into actionable estimates of fruit maturity, harvest timing, and marketable quality. In orchard systems more broadly, management and marketing require forecasts of expected size, quality, and timing, and technologies relevant to harvest timing already include proximal sensing of fruit maturity attributes (Anderson et al., 2021). For bayberry specifically, machine-vision maturity models have direct operational value because they can classify maturity with high accuracy and are explicitly positioned for intelligent harvesting and automated orchard systems (Zheng et al., 2025). Comparable bayberry maturity work based on fused color and texture features also achieved fast inference and strong test accuracy, which supports on-device deployment in smart orchards (Kai et al., 2021).

The broader management value of the model lies in targeting interventions at the orchard or sub-orchard scale instead of treating the whole farm uniformly. Digital-twin work in mandarins showed that orchard-level variables can predict fruit quality and that data-driven systems can advance micro-precision and individualized agriculture (Kim and Heo, 2024). In bayberry production, such a framework could link predicted quality to selective harvesting, sorting, irrigation adjustment, and fertilization or canopy-management decisions, while also integrating visual detection modules for fruit counting and complex-orchard monitoring. Even so, large-scale commercialization still depends on better standardization of agronomic practices and data collection under complex orchard environments.

6 Integration of Prediction Models with Precision Bayberry Production Systems

6.1 Digital agriculture technologies for quality monitoring

Precision bayberry production increasingly depends on non-destructive digital technologies that can monitor fruit quality in real time or near real time. Hyperspectral imaging has already been used in bayberry to predict fruit weight and soluble solids content across fruits from different regions and dates, while a low-cost portable Si-sensor device achieved 94.74% and 97.14% detection accuracy for sugar and acidity, respectively, showing that both high-dimensional imaging and low-cost sensing can support field-oriented quality monitoring (Yuan et al., 2025). These technologies are well aligned with the broader digitalization needs of the bayberry industry, where low-cost information sensing has been identified as a bottleneck for digital agriculture adoption and portable quality detectors are especially relevant for small-scale production systems (Wang et al., 2023).

Imaging and machine-vision systems extend quality monitoring beyond internal traits to visible maturity, fruit localization, and harvest readiness. A cascaded instance-segmentation and regression framework predicted bayberry maturity with a mean absolute error of 3.946% and 95.51% accuracy for three maturity classes, while an improved YOLOv7-tiny model reached 97.6% recall and was deployed on a mobile harvesting robot, indicating that in-field visual monitoring can be integrated directly with automated harvesting workflows (Zheng et al., 2025). More broadly, smart agriculture studies show that in-situ sensor systems can track fruit growth and health status with strong correlations to chemical indicators, and AI-based fruit-monitoring reviews indicate that multimodal sensing frameworks are becoming central for scalable, real-time quality evaluation in precision farming (Colaco and Kamat, 2025).

6.2 Dynamic prediction and decision-support systems

Prediction models become decision-support tools when they transform continuous sensor and orchard data into timely estimates of maturity, harvest timing, and expected quality. Reviews of precision agriculture for fruit maturity emphasize that pre-harvest quality estimation is essential for harvest timing, storage planning, and profitability, and that high spatiotemporal resolution sensing systems are needed because in-field maturity indicators vary substantially over time and space (Islam et al., 2024). In bayberry, this logic is already supported by quality-prediction studies using rapid sensing: an electronic-nose model predicted bayberry quality with 95% validation accuracy, and postharvest flavor-deterioration work showed that e-nose sensor responses plus ANN classification could distinguish fruit that had become off-flavored under specific storage temperature-time combinations (Gao et al., 2024).

Effective decision support also requires the system to present predictions in an operationally usable form. A flexible Vis/NIR wireless sensing system for banana integrated predictive models into an MCU and displayed predicted quality parameters and quality levels through both internet-connected and local interfaces, illustrating a practical pathway for embedding fruit-quality prediction into real-time digital dashboards (Wang et al., 2023). In orchard-scale applications, digital-twin research in mandarin showed that orchard-level weather and management variables can be used for fruit-quality prediction and argued that such systems support micro-precision and individualized agriculture, which is directly relevant to bayberry orchards where prediction outputs could guide selective harvest, grading, and management adjustments (Kai et al., 2021).

6.3 Challenges in model transferability and practical application

A major challenge in practical deployment is that models often lose accuracy when they are transferred across batches, orchards, cultivars, or measurement modes. In bayberry hyperspectral work, slope-bias correction was needed to improve prediction for other batches and packaged fruits, and the authors showed that this approach reduced deviations across different batches and acquisition modes, highlighting that calibration transfer is not optional but often necessary for robust field application (Yuan et al., 2025). Generalization challenges also arise in orchard vision systems: a bayberry detection model for complex orchards had to explicitly address small targets, occlusion, and dense distribution, and its value was strengthened by external generalization tests on other small-target and dense-occlusion datasets.

Transferability is further constrained by biological and management heterogeneity that prediction models must absorb without overfitting. Bayberry fruit quality is a complex trait involving external and internal attributes and multiple quality-associated genes, while orchard studies show that practices such as accompanying ryegrass can significantly alter fruit quality, soil properties, microbial communities, and metabolites, meaning that models trained in one genotype-management context may not transport cleanly to another (Zhang et al., 2024). Practical deployment also remains limited by real orchard conditions: segmentation accuracy declines under variable lighting and occlusion in natural environments, and reviews of fruit-production automation note that inconsistent horticultural management and complex field environments remain major barriers to large-scale commercialization (Júnior et al., 2025).

7 Future Perspectives and Research Directions

7.1 Multi-source data integration for improved prediction accuracy

Future bayberry prediction models should move beyond single-source inputs and combine environmental, spectral, imaging, and management data, because multimodal fusion consistently improves the robustness and accuracy of fruit-quality assessment. Reviews of precision fruit sensing indicate that combining image and sensor modalities typically yields measurable gains over unimodal systems, and recent fruit studies have shown that multi-source fusion can markedly improve prediction of internal quality traits under non-destructive conditions (Kalprajsinh et al., 2026). This direction is especially relevant for bayberry, where maturity, soluble solids, and acidity are influenced by both visible traits and internal biochemical variation. A practical future framework would therefore integrate orchard weather, canopy images, Vis/NIR spectra, and cultivation records into one prediction pipeline rather than modeling each information source separately (Colaco and Kamat, 2025).

Data fusion will also need to address two current bottlenecks: limited labeled samples and poor transfer across varieties or orchards. Multi-source spectral fusion combined with self-supervised learning improved downstream prediction while using very small labeled datasets, and balanced multi-source transfer learning reduced negative transfer and improved R2 by 36.4% over single-crop training in data-scarce settings (Zhang et al., 2024). For bayberry, these strategies are promising because new cultivars or newly instrumented orchards rarely have enough historical quality records for stable supervised modeling. Future work should therefore prioritize shared pretraining datasets, source-target balancing strategies, and cross-orchard calibration protocols that make prediction models usable earlier in the development cycle (Kim and Sun, 2026).

7.2 Advanced artificial intelligence approaches in fruit quality prediction

Advanced AI approaches in bayberry prediction should increasingly rely on deep learning architectures that can learn directly from high-dimensional and heterogeneous data. In fruit-quality research, deep learning has improved robustness relative to conventional chemometrics, and optimized deep fusion models such as WOA-CNN-LSTM have shown clear advantages for handling complex high-dimensional inputs in multi-source quality prediction. This is important for bayberry because future prediction targets will likely include not only soluble solids or acidity, but also maturity stage, disease risk, and postharvest behavior from combined sensory streams. Bayberry-specific maturity research already shows that lightweight segmentation plus fused deep, color, and texture features can achieve accurate in-field maturity prediction while remaining computationally efficient (Zheng et al., 2025).

The next step is not only higher accuracy, but also more interpretable and adaptable AI. Explainable AI methods such as LIME and SHAP are already being used to visualize model decisions in fruit-quality prediction, while XAI-guided wavelength selection has shown that transparent feature reduction can improve efficiency and enable spatial visualization of internal quality variation (Chowdhury et al., 2024). For bayberry systems, interpretable AI would help identify whether weather, pigmentation, canopy texture, or cultivation factors are the main drivers of a predicted quality outcome, which is essential for management uptake. At the same time, transfer learning, multi-task learning, and knowledge-guided model design appear especially promising for improving cross-scenario robustness beyond narrowly tuned datasets.

7.3 Towards sustainable and intelligent bayberry production

Future bayberry production systems should connect prediction models directly to sustainable orchard management decisions rather than treating prediction as an isolated analytical task. AI in agriculture is increasingly positioned as decision support for real-time monitoring, predictive analytics, and resource optimization, and this broader shift aligns well with fruit systems that need finer control over harvest timing, inputs, and quality consistency. In bayberry, protected and ecological cultivation practices already show strong quality effects: greenhouse cultivation improves fruit weight, width, and soluble solids, while insect- and rain-proof nets improve edible rate and the Brix/acid ratio while reducing pest damage and sugar-consuming bacteria (Wu et al., 2021). Prediction systems that detect these responses early could help growers adjust protected cultivation, harvest windows, and orchard interventions more precisely.

Longer term, intelligent bayberry production will likely depend on integrating biological, ecological, and digital information into unified management platforms. Bayberry research is beginning to expand from phenotype monitoring toward genomic, transcriptomic, and microbiome resources, while reviews of Myrica emphasize the value of centralized databases and the need for more sustainable development pathways for the crop. Microbiome-oriented evidence also suggests that orchard-specific bacterial composition and cultivar-dependent fungal diversity are relevant to quality and postharvest outcomes, which means future intelligent systems may need to incorporate microbial indicators alongside climate and management variables (Xie et al., 2025). Overall, the strongest future direction is a bayberry production model that is simultaneously predictive, explainable, and sustainability-oriented.

8 Conclusions

Recent advances show that bayberry quality prediction has moved from conceptual possibility to measurable performance across multiple quality targets. Hyperspectral imaging predicted bayberry weight with RMSEP below 0.6 and soluble solids content with RMSEP below 0.8, while image-based maturity modeling reached 95.51% accuracy for three maturity classes, demonstrating that both internal and external quality traits can now be estimated effectively using data-driven methods. These results also confirm that prediction targets in bayberry are multidimensional rather than singular, because fruit quality includes size, color, sugar, acidity, and other internal traits with direct commercial relevance.

The broader methodological advance is the successful linkage of fruit phenotypes to environmental and cultivation signals in precision systems. Reviews across fruit crops show that pre-harvest fruit quality and maturity estimation is now built on high-spatiotemporal-resolution sensing, and orchard-scale sensor networks have already linked soil and environmental measurements to harvest-time physicochemical fruit traits in commercial production. At the same time, fruit modeling studies indicate that accurate prediction increasingly depends on combining climate conditions, management strategies, and plant or fruit growth variables, which directly supports the environmental-and-cultivation framework proposed for bayberry. These prediction advances matter because they fit directly into the larger transition toward precision horticulture. Precision horticulture aims to increase crop yield and quality while reducing environmental impact, and it uses sensors, robotics, and data-driven decision making to manage spatial and temporal variability more efficiently. In fruit systems, this shift is especially valuable because predictive monitoring enables growers to adjust orchard management during the season, improving the balance between quality, yield, and input use before harvest rather than reacting after quality losses have already occurred.

For bayberry, the practical contribution is the possibility of more sustainable and intelligent production decisions at fine scale. Precision horticulture reviews identify sensor-based monitoring, IoT, AI, and decision-support systems as key tools for improving irrigation, fertigation, protected cultivation, and resource-use efficiency, while smart horticulture frameworks argue that precise perception, smart operation, and smart control are central to future horticultural systems. Because bayberry quality responds strongly to cultivation environment, including light exposure and protected production conditions, prediction models can help translate these management effects into earlier and more targeted actions for quality improvement and sustainability. The main remaining challenge is that strong local model performance does not guarantee reliable transfer across orchards, batches, cultivars, or sensing conditions. In bayberry itself, model transfer required slope-bias correction to reduce prediction deviations across batches and acquisition modes, and broader remote-sensing research still identifies robustness and transferability of fruit-quality models as active unresolved problems. Additional barriers include data accessibility, algorithmic bias, implementation cost, and training requirements, all of which still limit practical AI integration in horticultural stress and management systems.

The clearest future opportunity is to integrate multi-source sensing with biological and digital intelligence in a more standardized pipeline. Multi-omics, high-throughput phenotyping, and machine learning are reshaping fruit agriculture, but progress is still limited by heterogeneous datasets and the gap between discovery and field deployment; similarly, berry-crop programs emphasize open data platforms and FAIR-style data availability to support durable, reusable improvement workflows. For bayberry specifically, future systems should combine environmental monitoring, cultivation records, fruit phenotyping, and emerging genetic or microbiome information, because fruit quality is a complex multigenic trait and orchard-dependent microbiome variation also contributes to quality and postharvest behavior.

Acknowledgments

I extend my sincere gratitude to the anonymous reviewers for their valuable and insightful comments, which have greatly strengthened this paper.

Conflict of Interest Disclosure

The author affirms that this research was conducted without any commercial or financial relationships that could be construed as a potential conflict of interest.

References

Abdel-Salam M., Kumar N., and Mahajan S., 2024, A proposed framework for crop yield prediction using hybrid feature selection approach and optimized machine learning, Neural Computing and Applications, 36(33): 20723-20750.
https://doi.org/10.1007/s00521-024-10226-x

Anderson N., Walsh K., and Wulfsohn D., 2021, Technologies for forecasting tree fruit load and harvest timing-From ground, sky and time, Agronomy, 11(7): 1409.
https://doi.org/10.3390/agronomy11071409

Chowdhury R., Das R., Ananna F. B. F., Saha A., Nawar S., and Hosen M. H., 2024, Unveiling predictive factors in apple quality: Leveraging LIME, SHAP, and the synergy of machine learning models and artificial neural networks, 2024 6th International Conference on Electrical Engineering and Information and Communication Technology (ICEEICT), 2024: 1026-1031.
https://doi.org/10.1109/ICEEICT62016.2024.10534426

Colaco L. and Kamat P., 2025, Artificial intelligence advances for cashew fruit maturity and quality detection: A systematic review on models, sensors, and farming applications, Journal of Big Data, 12(1): 250.
https://doi.org/10.1186/s40537-025-01296-2

Gao J., Zheng X., Jiang A., Rong J., Yue W., Cao J., and Sun C., 2024, Characterization of flavor quality deterioration of postharvest Chinese bayberry (*Myrica rubra* cv. Dongkui) at different storage temperatures, Journal of Food Composition and Analysis, 130: 106146.
https://doi.org/10.1016/j.jfca.2024.106146

Islam M., Bijjahalli S., Fahey T., Gardi A., Sabatini R., and Lamb D., 2024, Destructive and non-destructive measurement approaches and the application of AI models in precision agriculture: A review, Precision Agriculture, 25(3): 1127-1180.
https://doi.org/10.1007/s11119-024-10112-5

Jewan S., Gautam D., Sparkes D., Singh A., Billa L., Cogato A., Murchie E. H., and Pagay V., 2024, Integrating hyperspectral, thermal, and ground data with machine learning algorithms enhances the prediction of grapevine yield and berry composition, Remote Sensing, 16(23): 4539.
https://doi.org/10.3390/rs16234539

Júnior M. R. B., Santos R. G. D., De Azevedo Sales L., Vargas R. B. S., Deltsidis A., and De Oliveira L. P., 2025, Image-based and ML-driven analysis for assessing blueberry fruit quality, Heliyon, 11(3): e42288.
https://doi.org/10.1016/j.heliyon.2025.e42288

Kai H., Huan L., Zeyu J., Tianlun H., Zaili C., and Nan W., 2021, Bayberry maturity estimation algorithm based on multi-feature fusion, 2021 IEEE International Conference on Artificial Intelligence and Computer Applications (ICAICA), 2021: 514-518.
https://doi.org/10.1109/ICAICA52286.2021.9498084

Kalprajsinh R., Chaudhari N., and Satapathy S., 2026, Multimodal data fusion of image and sensor data for precision fruit quality evaluation, 2026 IEEE International Conference on Interdisciplinary Approaches in Technology and Management for Social Innovation (IATMSI), 4: 1-6.
https://doi.org/10.1109/IATMSI68868.2026.11465697

Kim S. and Heo S., 2024, An agricultural digital twin for mandarins demonstrates the potential for individualized agriculture, Nature Communications, 15(1): 1561.
https://doi.org/10.1038/s41467-024-45725-x

Kim Y. and Sun E.-H., 2026, Balanced multi-source transfer learning for fruit quality prediction with limited data, IEEE Access, 2026: 87992-88008.
https://doi.org/10.1109/ACCESS.2026.3697751

Lei H., Huang K., Jiao Z., Tang Y., Zhong Z., and Cai Y., 2022, Bayberry segmentation in a complex environment based on a multi-module convolutional neural network, Applied Soft Computing, 119: 108556.
https://doi.org/10.1016/j.asoc.2022.108556

Li C., Li G., Qi X., Yu Z., Abdallah Y., Ogunyemi S. O., Zhang S., Ren H., Mohany M., Al-Rejaie S. S., Li B., and Liu E., 2023, The effects of accompanying ryegrass on bayberry trees by change of soil property, rhizosphere microbial community structure, and metabolites, Plants, 12(21): 3669.
https://doi.org/10.3390/plants12213669

Mo J., Rashwan A. K., Osman A. I., Eletmany M. R., and Chen W., 2024, Potential of Chinese bayberry (*Myrica rubra* Sieb. et Zucc.) fruit, kernel, and pomace as promising functional ingredients for the development of food products: A comprehensive review, Food and Bioprocess Technology, 17(11): 3506-3524.
https://doi.org/10.1007/s11947-023-03313-9

Niedbała G., Kurek J., Świderski B., Wojciechowski T., Antoniuk I., and Bobran K., 2022, Prediction of blueberry (*Vaccinium corymbosum* L.) yield based on artificial intelligence methods, Agriculture, 12(12): 2089.
https://doi.org/10.3390/agriculture12122089

Ping D., Wu Q.-Y., Huang T., Li S., and Wu W., 2025, Research on intelligent evaluation of pear fruit quality based on machine learning, 2025 IEEE 3rd International Conference on Computer, Vision and Intelligent Technology (ICCVIT), 2025: 1-5.
https://doi.org/10.1109/ICCVIT67848.2025.11391429

Ribeiro M. N., Carvalho I., Ferreira D., and Pinheiro A. C. M., 2022, A comparison of machine learning algorithms for predicting consumer responses based on physical, chemical, and physical-chemical data of fruits, Journal of Sensory Studies, 37(3): e12738.
https://doi.org/10.1111/joss.12738

Sun Z., Tian H., Hu D., Yang J., Xie L., Xu H., and Ying Y., 2024, Integrating deep learning and data fusion for enhanced oranges soluble solids content prediction using machine vision and Vis/NIR spectroscopy, Food Chemistry, 464(1): 141488.
https://doi.org/10.1016/j.foodchem.2024.141488

Wang J., Wu W., Tian S., He Y., Huang Y., Wang F., and Zhang Y., 2023, Non-destructive determination of bayberry sugar and acidity by hyperspectral remote sensing of Si-sensor and low-cost portable instrument development, Sensors, 23(24): 9822.
https://doi.org/10.3390/s23249822

Williams S., Baniya A. A., Islam M. S., and Murphy K., 2023, A data ecosystem for orchard research and early fruit traceability, Horticulturae, 9(9): 1013.
https://doi.org/10.3390/horticulturae9091013

Wu B.-P., Zhang C., Gao Y.-B., Zheng W., and Xu K., 2021, Changes in sugar accumulation and related enzyme activities of red bayberry (*Myrica rubra*) in greenhouse cultivation, Horticulturae, 7(11): 429.
https://doi.org/10.3390/horticulturae7110429

Xie Z., Zheng S., Yang S., Tang Y., Yu H., Chang Y., Zhu Y., Zhan X., Zeng G., and Chen H., 2025, Comparative analysis of fruit microbiota and metabolites in two bayberry cultivars: Implications for fruit quality and postharvest disease control, Food Chemistry: X, 29: 102773.
https://doi.org/10.1016/j.fochx.2025.102773

Yang H., Sun L., Qi Y., Li Z., Lei K., Cheng F., Wu Y., Ying Z., Lei Y., Ahmed T., Yu Z., Qi X., and Zhang S., 2025, Integrated transcriptomic and metabolomic analysis reveals light-induced modulation of anthocyanin biosynthesis in Chinese bayberry (*Myrica rubra*), Fruit Research, 5(1): 4.
https://doi.org/10.48130/frures-0025-0004

Yuan L.-M., Fu X., Zuo X., Jiang Q., Ji H., Chen X., Jiang C., Xie Z., and Chen X., 2025, Non-destructive assessment of bayberry quality using hyperspectral imaging analysis: From individual to template-packaged product via model transfer, Food Chemistry, 2025: 146965.
https://doi.org/10.1016/j.foodchem.2025.146965

Zhai C., Li Y., Li J., Guan P., Jin J., and Jia W., 2026, Signaling mechanisms governing the environmental modulation of fruit quality development, Horticulture Research, 13(4): uhag005.
https://doi.org/10.1093/hr/uhag005

Zhang L., Liu J., Wei Y., An D., and Ning X., 2024, Self-supervised learning-based multi-source spectral fusion for fruit quality evaluation: A case study in mango fruit ripeness prediction, Information Fusion, 117: 102814.
https://doi.org/10.1016/j.inffus.2024.102814

Zheng H., Sun L., Wang Y., Yang H., and Zhang S., 2025, Image-based detection of Chinese bayberry (*Myrica rubra*) maturity using cascaded instance segmentation and multi-feature regression, Horticulturae, 11(10): 1166.
https://doi.org/10.3390/horticulturae11101166