

Scientific Paper

DOI: <http://dx.doi.org/10.1590/1809-4430-Eng.Agric.v45e20250036/2025>

AN IMPROVED REGNET METHOD FOR DETECTING CHERRY FRUIT APPEARANCE QUALITY

Guohui Wang^{1*}, Wenchi Xu¹, Xi Chen¹, Jiwen Yang¹

^{1*}Corresponding author. Xi'an Technological University/Xi'an, China.

E-mail: booler@126.com | ORCID ID: <https://orcid.org/0000-0002-9631-0307>

KEYWORDS

RegNet, cherry
appearance,
convolutional neural
network, feature
extraction, fruit
classification.

ABSTRACT

Cherries are popular with consumers due to their high nutritional value. Their appearance characteristics are essential criteria for consumer purchasing and are closely related to ripeness classification. However, the existing deep learning methods for cherry appearance quality detection face a trade-off between accuracy and computational efficiency, while manual sorting remains labor-intensive and costly. To address these issues, we introduce an improved regular network (RegNet), aiming to accurately detect the visual quality of fruit. This is achieved by using different combinations of deep features and classifier models. The method involves extracting deep features from 12 RegNet models of varying parameter scales and combining these features with nine classifiers to assess cherry appearance quality. The proposed method achieves an accuracy of 98.7% in detecting the quality of cherry fruit, with a significant reduction in the training time. Experiments demonstrate that our method can rapidly and accurately evaluate cherry quality, offering a practical solution for automatic sorting systems.

INTRODUCTION

Cherries are rich in vitamins A and C, which enhance the immune system and have anti-inflammatory effects (Faenza et al., 2020). With China's rapid agricultural development, waste generation has become a major issue, especially when fruit quality fails to meet consumer expectations (Sanches et al., 2008). As cherry production expands, the demand for sorting grows. Currently, appearance-based sorting relies on labor, impacting efficiency and marketability (Neitzel et al., 2025). High-performance models are urgently needed to enhance ripeness and disease detection, accelerating cherries' market entry (Yang & Wang, 2023).

Recently, technologies such as machine vision and deep learning have been increasingly integrated into controlling and detecting the exterior quality of fruit, replacing traditional manual inspection methods (Bai et al., 2024; Barrios-Rodríguez et al., 2021; Wang et al., 2022). Deep learning has shown remarkable capabilities across various complex visual fields. For instance, it can enhance the image quality in challenging environments like underwater settings (Wang et al., 2025) and classify objects from difficult

data sources, such as low-resolution and imbalanced sonar images (Chen et al., 2024). Compared to conventional techniques, machine vision and deep learning offer high accuracy, fast detection, and lower costs while enabling non-destructive inspection adaptable to different cherry varieties (Liu et al., 2024; Wang et al., 2024). Numerous studies have explored deep learning for fruit quality assessment. Momeny et al. (2020) combined max-pooling and average-pooling to enhance a neural network's generalization ability. The technique was employed for cherry appearance recognition and classification. Song et al. (2023) and Zheng et al. (2022) both made significant contributions to fruit appearance quality detection using the Swin Transformer. Song et al. applied the Swin Transformer with a classifier for cherry sorting, advancing cherry-sorting machine development. Zheng et al. used a pre-trained Swin Transformer to extract class tokens, which were then processed by a multi-layer perceptron to recognize strawberries. Liu et al. (2023) utilized Raman spectroscopy combined with an improved convolutional neural network-convolutional block attention module model, achieving a 98.28% accuracy in identifying rice growth periods in Heilongjiang Province. Dong & Wang (2023) proposed a crop disease and pest identification model

¹ Xi'an Technological University/Xi'an, China.

Area Editor: Gizele Ingrid Gadotti

Received in: 3-8-2025

Accepted in: 6-23-2025



combining adaptive chaotic particle swarm optimization (ACPSO) with a support vector machine (SVM), achieving a 95.08% recognition accuracy through optimized feature extraction. Zhu et al. (2021) proposed an innovative method for carrot appearance detection based on convolutional neural networks (CNNs) and SVMs, contributing to the advancement of carrot-sorting machines. Kheiralipour & Pormah (2017) proposed a grading system for different cucumber morphologies, employing image processing techniques and artificial neural networks to optimize profitability. The method demonstrated a 97.1% accuracy in identifying cucumber morphologies. Wang et al. (2023) proposed a strawberry appearance identification method, ResNeXt-SVM, combining the ResNeXt network and an SVM. It achieved an accuracy of 98.56%, 1.42% higher than the original ResNeXt model. He et al. (2022) proposed a grape powdery mildew grading model, which effectively prevented model overfitting and improved the classification accuracy by enhancing the residual structure of ResNet-50. Overall, these advancements in visual deep learning reveal key trends: pursuing powerful, computation efficient feature extraction; integrating deep features with task specific classifiers effectively; and developing strategies to handle complex data features, boosting model robustness and performance.

The existing deep learning methods have delivered excellent performance on various tasks. However, they face challenges in processing large-scale data and usually require a long training time. For cherry quality detection, balancing high accuracy and computational efficiency is still a vital issue. Thus, this paper presents an improved regular network (RegNet)-based method for cherry appearance quality inspection. By integrating the feature extraction of RegNet models and multiple classifiers' strengths, it aims to enhance cherry inspection. Our study makes the following main contributions:

1. We introduce a hybrid framework, using 12 different RegNet models as deep feature extractors. The extracted features are used by nine classifiers for final classification, which is a departure from traditional models.

2. This study systematically evaluates the performance of various RegNet feature extractor and classifier combinations. Through comprehensive assessment, the combinations that achieve the best balance of accuracy and efficiency in cherry quality detection are identified.

3. Our optimal combination (RegNet_y_8gf + MLP) reaches a 98.7% accuracy on the cherry quality detection dataset.

4. Our proposed method significantly reduces the model training time by performing feature extraction only once, eliminating the need for repeated parameter optimization throughout the training process.

MATERIAL AND METHODS

Dataset and experimental environment

The dataset comprises 4,669 images of cherries captured by HUAWEI Nova 9 mobile phones at the Bai Lu Yuan cherry orchard in Xi'an. The cherry images were classified with precision into five independent categories, corresponding to different degrees of ripeness and states of cherry disease. The characteristics of the various categories of cherries can be observed in the example images provided in Figure 1. The number of unripe cherries was 1,198, and these exhibited a green epidermis (Figure 1a), indicating that their growth process was not yet complete. The number of half-ripe cherries was 1,002; they were characterized by a surface color showing a mixture of red and yellow (Figure 1b). Additionally, irregular crosses were observed on the surfaces of these cherries, indicating the onset of ripening. The number of ripe cherries was 1,047; they exhibited a wax-red surface (Figure 1c), representing a pivotal stage in the cherry growth cycle. The number of overripe cherries was 1,162. This type of cherry has surpassed the optimal ripening stage, exhibiting a slightly waxy black surface (Figure 1d), indicative of overripeness. Additionally, 255 instances were identified as diseased cherries, exhibiting visible signs of deterioration such as the wrinkling or rotting of the skin (Figure 1e). This distinct presentation differentiates them from the other cherry categories.



FIGURE 1. Cherry dataset categories.

The resolution of the acquired cherry images was adjusted by modifying them uniformly to 224×224 pixels. The lower-resolution images were also filled in to ensure that all images were suitable for model training and analysis. Subsequently, to guarantee accurate and reliable model performance evaluation, the entire dataset of 5,711 images was divided into three subsets at a 6:2:2 ratio: 3,429 images for training, 1,141 for testing, and 1,141 for validation.

Data preprocessing

The performance of deep learning models relies on a

sufficient data volume for training CNNs. Due to the limited number of images within each disease category, data augmentation techniques were applied to expand the dataset. Seven distinct techniques were used, including image brightness and contrast adjustments, random rotation, mirror flipping, affine transformation, dislocation transformation, and HSV (hue, saturation, value) enhancement. These seven data enhancement methods generated 1,302 images of disease categories, and the results are presented in Figure 2. The data distribution of each category of cherry images is subsequently presented in Table 1.

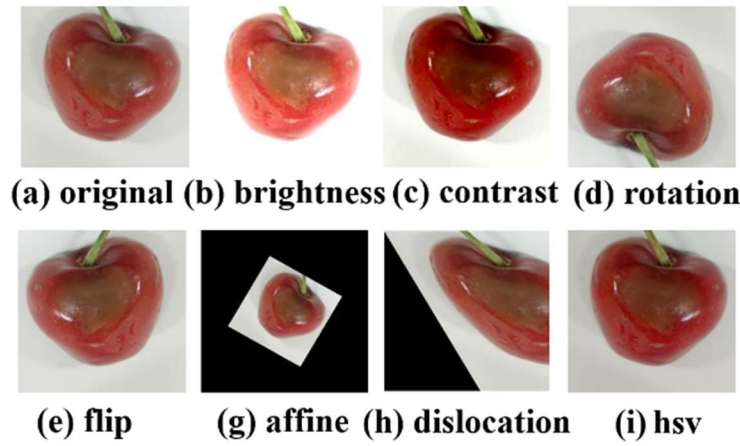


FIGURE 2. Data preprocessing.

TABLE 1. Data distribution.

Item	Resolution	Training Set	Validation Set	Test Set	Total
Unripe	224×224	718	240	240	1198
Half-ripe	224×224	602	200	200	1002
Ripe	224×224	629	209	209	1047
Overripe	224×224	698	232	232	1162
Disease	224×224	782	260	260	1302
All	224×224	3429	1141	1141	5711

Cherry ripeness and disease detection model

RegNet model

Radosavovic et al. (2020) put forth the RegNet structure, which employs neural architecture search (NAS) (Chitty-Venkata et al., 2023). Conventional approaches to NAS are based on a single network instance and are constrained by limitations in flexibility, generalization, and interpretability. However, RegNet's objective is to investigate the design of the space and identify network design principles rather than merely locating a set of parameters (Li et al., 2021). Figure 3 illustrates the RegNet architecture. The main parameters include the number of blocks d_i in each RegStage, the number of channels w_i in the output matrix in

each RegStage, and the width g of each Group in a Block. By adjusting these parameters, RegNet can reduce the number of parameters in the model while optimizing the performance, thus increasing the training speed. The RegNet design space network mainly comprises an input layer (Stem), a backbone layer (Body), and an output layer (RegHead), with its structure presented in Table 2. The input layer is a standard convolutional layer. The backbone layer represents the primary structure of the network and comprises four stages, as illustrated in Figure 3c. In each stage, the height and width of the input feature matrices are reduced by a factor of two. The output layer is a conventional classification structure comprising a global average pooling (GAP) layer and a fully connected (FC) layer.

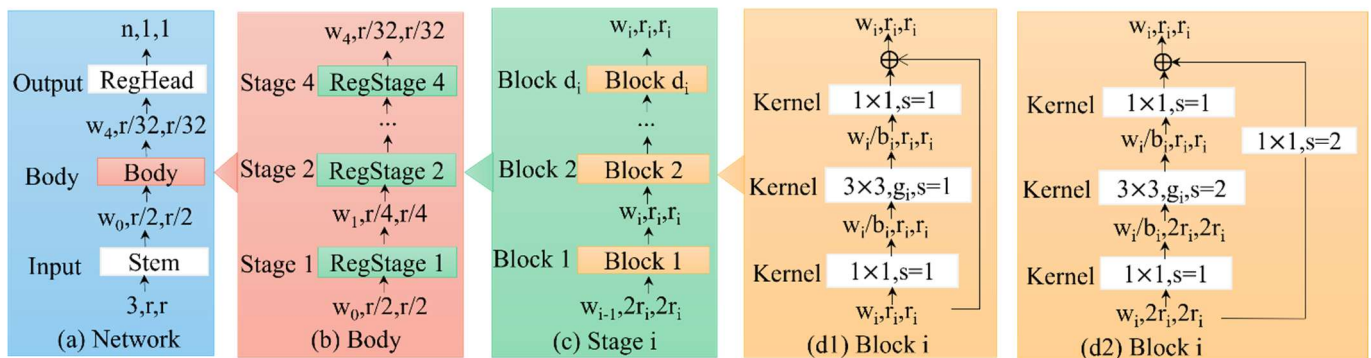


FIGURE 3. RegNet structure visualization.

TABLE 2. RegNet structure.

Layer	Input Size	Output Size	Type
Steam	3×224×224	32×224×224	3×3 Conv2d
RegStage 1	32×112×112	48×112×112	Block
RegStage 2	48×56×56	104×56×56	Block
RegStage 3	104×28×28	208×28×28	Block×4
RegStage 4	208×14×14	440×14×14	Block×7
RegHead	440×7×7	1×1×19	7×7 GAP, Dropout, FC

Proposed method

Cherry appearance quality detection requires the identification of different ripeness stages (unripe, half-ripe, ripe, overripe), as well as the presence of diseases. However, cherry ripening is a continuous process, and the changes in color, firmness, texture, and luster between different ripeness stages are gradual, lacking clear boundaries, which can lead to visual ambiguities. Beyond the diseases mentioned previously, cherries may also exhibit various other types of defects, such as cracks, bruises, and insect bites, each with distinct visual manifestations, posing significant challenges to appearance quality detection. To overcome these challenges and achieve a better balance between accuracy and computational efficiency, we introduce a novel hybrid method that combines deep feature extraction with diverse classification strategies.

Our method's core involves employing the RegNet architecture as the primary feature extractor. While traditional NAS focuses on identifying a single optimal network instance, RegNet is dedicated to discerning network design principles that balance model performance and computational cost. By adjusting the parameter scales of RegNet models, their performance can be optimized under various computational budgets, effectively reducing the model size and increasing the processing speed.

We utilize 12 RegNet network models of different parameter scales to comprehensively capture rich deep features from cherry images. Concurrently, considering that different classification algorithms possess unique principles and advantages, potentially yielding complementary or divergent predictions even with identical features (Patil & Burkpalli, 2025), we did not directly input the features into a fixed classification layer. Instead, we combined the extracted deep features with nine distinct classifiers: the SVM (Zhang et al., 2004; Cui et al., 2024), extremely randomized trees (ET), random forest (RF) (Murmur & Kumar, 2024), k-nearest neighbor (KNN) (Elasan & Keskin, 2019; Miranda et al., 2021), multi-layer perceptron (MLP) (Cahyo et al., 2023), Gaussian naive Bayes (GNB) (Manickam, 2025), linear discriminant analysis (LDA) (Suesse et al., 2023), XGBoost, and decision trees (DT) (Fonseca et al., 2025). Each classifier possesses a unique decision-making mechanism, aiming to balance accuracy and computational efficiency through diverse approaches. Through this hybrid framework, we can fully leverage the advantages of each component and achieve the best comprehensive performance on the task. The entire network structure is represented as a feature extraction module, followed by multiple optional classifiers, as shown in Figure 4.

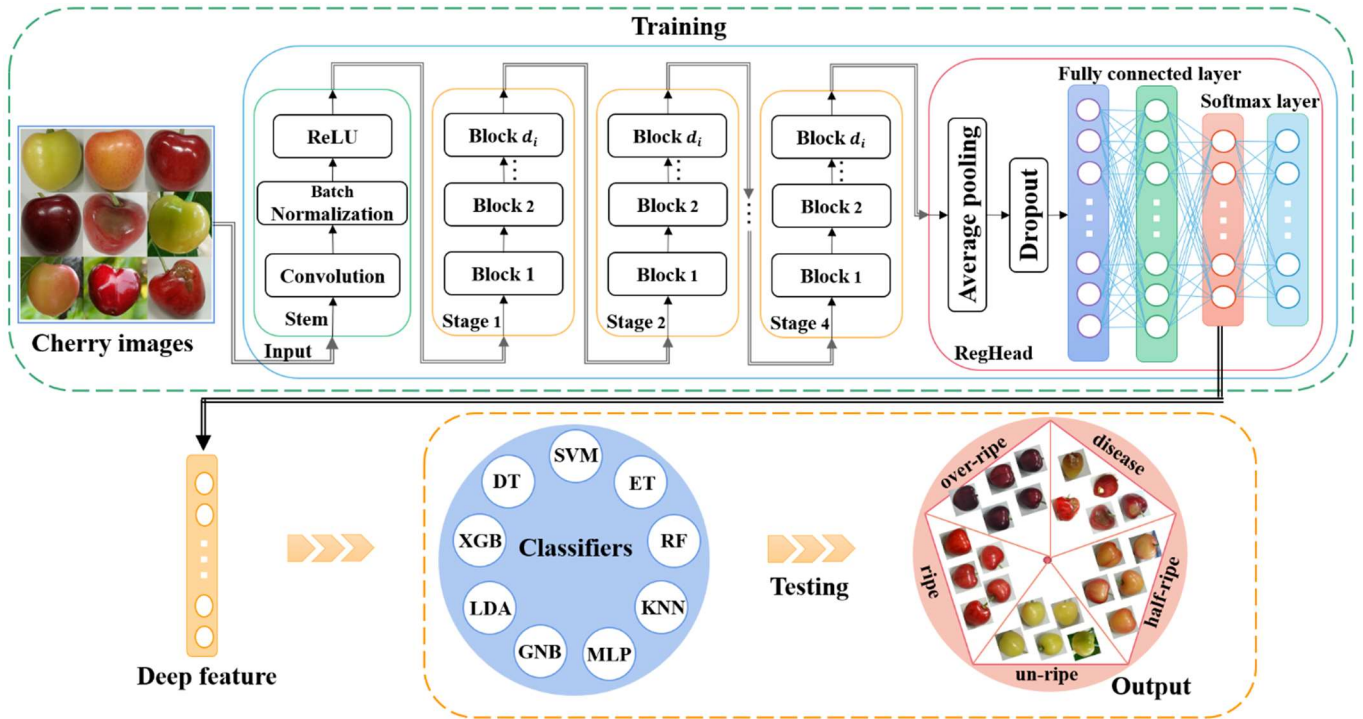


FIGURE 4. Overall network structure.

The experimental process comprises three stages, and its overall flow is illustrated in Figure 5:

(1) The collected cherry image dataset undergoes preprocessing, including image screening and enhancement.

(2) Pre-trained RegNet models are used to extract features from the entire dataset. Subsequently, the extracted features are individually fed into nine different classifiers for

training and classification. The key objective of this stage is to evaluate the performance of all RegNet models combined with all classifiers to identify the optimal combination.

(3) The recognition results of our proposed optimal method are comprehensively compared and evaluated against those of individual RegNet models and various transfer learning models.

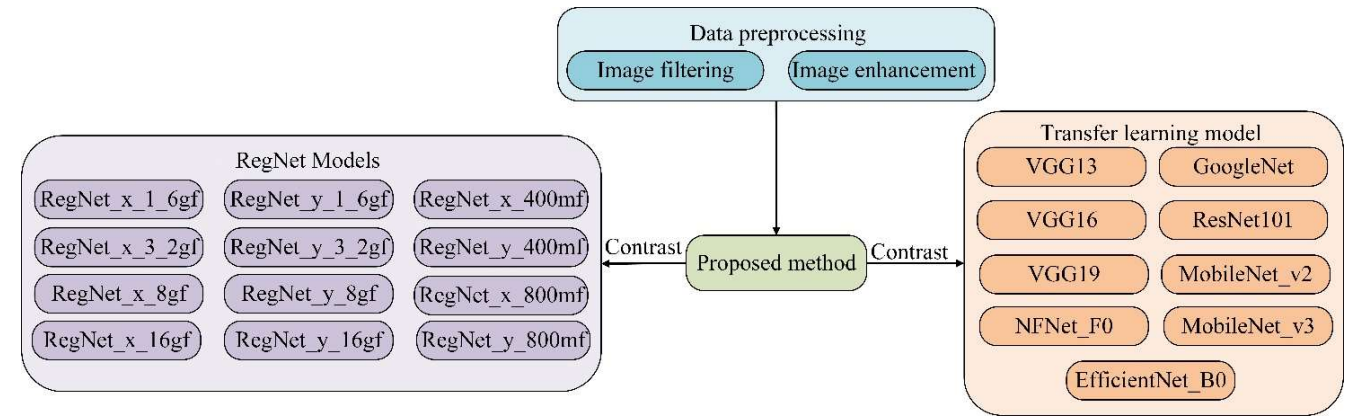


FIGURE 5. Experiment flowchart.

EXPERIMENTAL RESULTS AND ANALYSIS

To evaluate the performance of different model and classifier combinations, the results of the combinations are measured by six evaluation metrics, namely the accuracy, precision, recall, F1-score, false positive rate (FPR), and confusion matrix, as shown in Table 3. The accuracy offers a

general measure, and the precision and recall are critical for minimizing false rejections and missed defects in quality control. The F1-score effectively balances precision and recall, and a low FPR is vital for the sorting efficiency. Together, these metrics robustly assess the model's ability to accurately categorize cherries across diverse stages and defect types, ensuring both effectiveness and practicality.

TABLE 3. Model evaluation criteria and definitions.

Criterion	Definition
Accuracy	$\frac{TP + TN}{TP + TN + FP + FN}$
Precision	$\frac{TP}{TP + FP}$
Recall	$\frac{TP}{TP + FN}$
F1-score	$\frac{2 \times TP}{2 \times TP + FP + FN}$
FPR	$\frac{FP}{TN + FP}$

All experiments were conducted on a dedicated personal computer platform, utilizing independent training, validation, and test sets. The hardware environment comprised two NVIDIA RTX 2080 Ti GPUs, along with an Intel Core i9-9900X CPU running at 3.50 GHz. For the software stack, the operating system was Ubuntu 18.04, with optimized GPU computation supported by CUDA 10.1 and cuDNN 8.04. All CNN training was performed using Python

3.7 and PyTorch 1.8.1, while the classifiers were implemented via the scikit-learn 1.0.2 library.

Here, the "training time" represents the total time it takes for each model to complete the training process, excluding the pre-training time. Our method includes the training time of the RegNet feature extractor and the classification time of the classifier. Unlike large-scale learning models that are trained from scratch, our method adopts the efficient RegNet architecture as the feature extractor.

Evaluation of RegNet and classifier performance

The performance evaluation of different combinations is primarily based on accuracy. In this study, we use features extracted from 12 distinct RegNet models and combine them with nine different classifiers to conduct experiments and assess the effectiveness of these combinations. The accuracy data for the various model combinations are presented in Table 4. A detailed analysis of the table reveals the following key points: first, the combination of RegNet_y_8gf with the MLP achieves the highest accuracy of 98.7%. Second, the RegNet_x_3_2gf and LDA combination shows the second highest accuracy at 98.4%, while the RegNet_y_8gf and

SVM combination also performs remarkably well, with an accuracy of 98.3%.

Several other combinations achieve an accuracy of 98.2%. To gain deeper insight into the performance of the various classifiers, we calculate the mean accuracy of each classifier using the 12 model features. The final row of the table shows that the SVM has the highest average accuracy at 97.9%, followed by LDA at 97.7%. Although the MLP has a slightly lower accuracy than these two classifiers, it still performs commendably, with an accuracy of 97.6%. In summary, the classification performances of the SVM, LDA, and the MLP are significantly better than those of the other classifiers in this experiment, providing a strong foundation for future research.

TABLE 4. Accuracy of RegNet models with different classifiers on test set.

Model	SVM	ET	RF	KNN	MLP	GNB	LDA	XGB	DT	Average
RegNet_x_1_6gf	97.9%	94.1%	90.5%	95.9%	97.6%	82.8%	98.2%	96.5%	83.5%	93.0%
RegNet_x_8gf	98.2%	93.5%	90.3%	93.5%	97.5%	75.9%	97.9%	95.9%	82.6%	91.7%
RegNet_x_3_2gf	97.8%	94.7%	90.4%	92.3%	97.5%	81.1%	98.4%	96.8%	82.2%	92.4%
RegNet_x_16gf	97.8%	94.8%	90.1%	93.6%	97.3%	75.5%	97.8%	96.5%	78.4%	91.3%
RegNet_x_400mf	97.4%	93.9%	89.6%	92.2%	97.8%	80.0%	96.5%	96.5%	81.2%	91.7%
RegNet_x_800mf	97.6%	93.1%	89.6%	92.8%	96.7%	78.7%	97.8%	95.8%	81.3%	91.5%
RegNet_y_1_6gf	98.0%	95.7%	92.4%	94.7%	97.6%	85.7%	97.9%	97.4%	81.4%	93.4%
RegNet_y_8gf	98.3%	94.9%	91.0%	93.9%	98.7%	78.7%	98.2%	96.0%	82.8%	92.5%
RegNet_y_3_2gf	98.2%	96.1%	91.5%	94.6%	97.9%	83.9%	98.1%	97.2%	85.5%	93.7%
RegNet_y_16gf	98.2%	93.3%	90.3%	92.0%	98.0%	79.1%	96.8%	96.4%	79.6%	91.5%
RegNet_y_400mf	97.0%	93.9%	89.6%	91.4%	96.8%	81.8%	96.7%	96.1%	80.2%	91.5%
RegNet_y_800mf	97.8%	95.0%	91.3%	94.2%	98.1%	83.8%	98.0%	96.2%	81.5%	92.8%
Average	97.9%	94.4%	90.6%	93.4%	97.6%	80.6%	97.7%	96.4%	81.7%	92.3%

Significant differences exist in the feature extraction time for different networks, influenced by the network structure. Table 5 shows the feature extraction time of the 12 networks in this experiment; RegNet_x_400mf has the shortest feature extraction time of 8.51 s, while RegNet_y_16gf has the longest feature extraction time of 58.258 s. The time taken to classify cherry images using different classifiers varies depending on the classifier algorithm and feature data. Table 6 shows the prediction times

of different classifiers; among the nine classifiers, the MLP has the highest average prediction time, and KNN has the shortest. There is a slight difference in the prediction times for the same classifier for different RegNet models, except for the MLP classifier. To calculate the training time of the RegNet model combined with classifiers, the feature extraction time is added to the prediction time of the classifiers. The training time is 65.05 s for RegNet_y_8gf + MLP, 21.46 s for RegNet_x_3_2gf + LDA, and 43.70 s for RegNet_y_8gf + SVM.

TABLE 5. Feature extraction time (s) for 12 RegNet models on test set.

Model	Time (s)
RegNet_x_1_6gf	13.82
RegNet_x_8gf	28.55
RegNet_x_3_2gf	21.01
RegNet_x_16gf	50.55
RegNet_x_400mf	8.51
RegNet_x_800mf	11.02
RegNet_y_1_6gf	18.75
RegNet_y_8gf	42.38
RegNet_y_3_2gf	23.63
RegNet_y_16gf	58.25
RegNet_y_400mf	10.61
RegNet_y_800mf	11.97
RegNet_x_1_6gf	13.82
RegNet_x_8gf	28.55
RegNet_x_3_2gf	21.01

TABLE 6. Prediction time (s) for different classifiers on test set.

Model	SVM	ET	RF	KNN	MLP	GNB	LDA	XGB	DT	Average
RegNet_x_1_6gf	1.36	0.63	0.75	0.002	3.72	0.02	0.59	19.87	4.65	3.51
RegNet_x_8gf	1.34	0.53	0.78	0.003	3.74	0.02	0.59	19.34	4.63	3.44
RegNet_x_3_2gf	1.88	0.5	0.77	0.008	4.15	0.04	0.45	15.76	4.89	3.17
RegNet_x_16gf	1.46	0.64	0.78	0.008	7.36	0.04	0.42	16.91	4.07	3.52
RegNet_x_400mf	1.94	0.59	1.33	0.005	7.94	0.03	0.79	17.43	7.67	4.19
RegNet_x_800mf	1.33	0.63	0.82	0.003	2.92	0.02	0.57	19.68	4.82	3.42
RegNet_y_1_6gf	1.35	0.52	0.75	0.003	13.59	0.02	0.57	17.24	4.53	4.29
RegNet_y_8gf	1.32	0.53	0.77	0.003	22.67	0.02	0.55	18.36	4.62	5.43
RegNet_y_3_2gf	1.19	0.63	0.79	0.002	13.69	0.02	0.56	18.49	4.97	4.48
RegNet_y_16gf	1.38	0.62	0.76	0.003	23.37	0.02	0.59	17.60	3.98	5.37
RegNet_y_400mf	1.35	0.63	0.78	0.003	22.49	0.02	0.57	20.23	5.27	5.70
RegNet_y_800mf	1.28	0.62	0.77	0.003	23.22	0.02	0.56	18.96	4.55	5.55
Average	1.43	0.59	0.82	0.003	12.41	0.02	0.57	18.32	4.89	4.34

Analysis of high-performing combinations

From Table 4, the three combinations with the highest accuracy are RegNet_y_8gf + MLP, RegNet_x_3_2gf + LDA, and RegNet_y_8gf + SVM, which achieve accuracies of 98.7%, 98.4%, and 98.3%, respectively. Five other combinations achieve a 98.2% accuracy, namely RegNet_x_1_6gf + LDA, RegNet_x_8gf + SVM, RegNet_y_8gf + LDA, RegNet_y_3_2gf + SVM, and RegNet_y_16gf + SVM.

Table 7 provides a comprehensive overview of various evaluation metrics. It is evident that the combinations ranking highest in accuracy also exhibit superior performance across other critical metrics, such as precision and recall. Among these top performers, RegNet_y_8gf + MLP stands out. This combination achieves a precision of 98.68%, a recall of 98.64%, the highest F1-score, and the lowest FPR, positioning it as the most effective model among the three highest-accuracy combinations.

TABLE 7. Evaluation metrics for top combinations on test set.

Model + Classifier	Precision	Recall	F1-Score	FPR
RegNet_y_8gf + MLP	98.68%	98.64%	98.62%	0.0032
RegNet_x_3_2gf + LDA	98.40%	98.32%	98.34%	0.0038
RegNet_y_8gf + SVM	98.22%	98.38%	99.24%	0.0040
RegNet_x_8gf + SVM	98.14%	98.24%	98.18%	0.0042
RegNet_x_1_6gf + LDA	98.20%	98.17%	98.09%	0.0045
RegNet_y_3_2gf + SVM	98.10%	98.12%	98.08%	0.0044
RegNet_y_16gf + SVM	98.14%	98.12%	98.10%	0.0046
RegNet_y_8gf + LDA	98.05%	98.22%	98.11%	0.0045

Figure 6 presents the confusion matrices for the three most accurate combinations. The RegNet_y_8gf + MLP combination demonstrates superior performance in recognizing the half-ripe, overripe, and ripe classes compared to RegNet_y_8gf + SVM and RegNet_x_3_2gf + LDA.

Specifically, this combination achieves remarkable accuracy rates of 97.7% for diseases, 98.0% for half-ripe, 100.0% for overripe, 99.0% for ripe, and 98.7% for unripe. These results highlight the exceptional recognition capabilities of the RegNet_y_8gf + MLP combination across all categories.

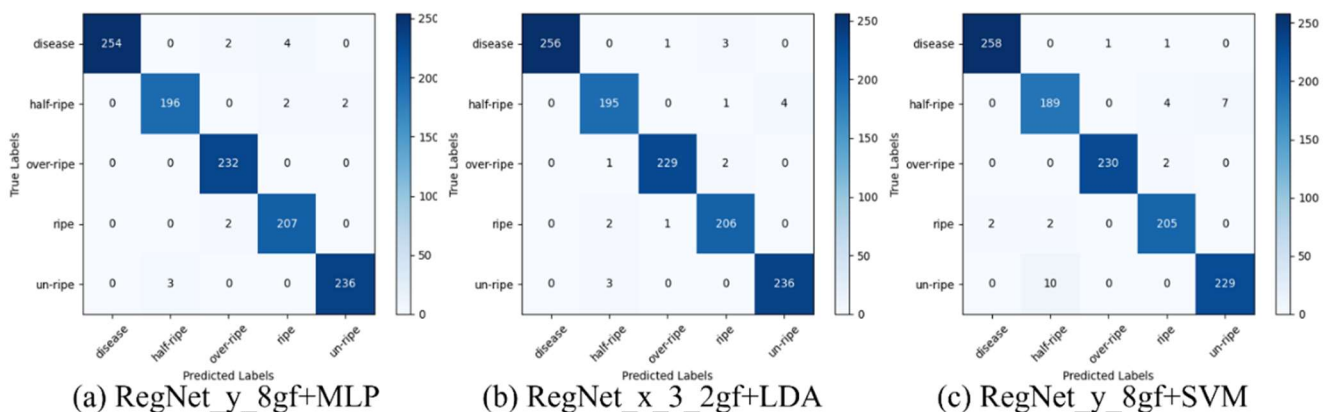


FIGURE 6. Confusion matrices for top combinations.

To evaluate the performance stability of the RegNet_y_8gf+MLP model and its propensity to overfit across different data subsets, we conducted K-fold cross-validation (K = 5) on the combined training and validation sets, which comprised a total of 4,570 images. During each iteration, the 4,570 images were evenly partitioned into five sub-folds. Four of these sub-folds were utilized as training data to fit the model parameters, while the remaining sub-fold functioned as the

validation set for performance assessment. After completing the K-fold cross-validation, we analyzed the model's performance across the five iterations on the different data subsets. This analysis enabled us to directly assess the model's robustness to variations in data partitioning and to evaluate its generalization capabilities. Ultimately, the RegNet_y_8gf + MLP model was tested on a separate test set consisting of 1,141 images, and the evaluation outcomes are detailed in Table 8.

TABLE 8. K-fold cross-validation results for RegNet_y_8gf + MLP.

Method	Precision	Recall	F1-Score	FPR
RegNet_y_8gf+MLP (k=1)	98.67%	98.63%	98.65%	0.0034
RegNet_y_8gf+MLP (k=2)	98.80%	98.75%	98.77%	0.0030
RegNet_y_8gf+MLP (k=3)	98.35%	98.32%	98.33%	0.0042
RegNet_y_8gf+MLP (k=4)	98.75%	98.70%	98.72%	0.0032
RegNet_y_8gf+MLP (k=5)	98.40%	98.35%	98.37%	0.0040
RegNet_y_8gf+MLP (Average)	98.59%	98.55%	98.57%	0.0036

Comparative analysis of optimal combinations and transfer learning models

To comprehensively evaluate the competitiveness of the method proposed in this paper, we benchmark its optimal combinations against advanced methods in the field of fruit quality assessment. We selected nine transfer learning models as comparison objects, including representative deep learning models such as EfficientNet_B0, NFNet_F0, and MobileNet_v3.

Figure 7 illustrates the accuracy comparison of the three optimal combinations against various transfer learning models (including EfficientNet_B0, NFNet_F0, and MobileNet_v3). Among them, RegNet_y_8gf + MLP, RegNet_x_3_2gf + LDA, and RegNet_y_8gf + SVM

achieved the highest accuracies of 98.70%, 98.40%, and 98.30%, respectively. Specifically, RegNet_y_8gf + MLP stands out with the highest accuracy of 98.70%. Its accuracy significantly surpasses those of all of the compared models; for instance, it is higher than the accuracies of GoogleNet, the VGG series (VGG13, VGG16, VGG19), and MobileNet_v2 by differences ranging from 3.80 to 13.00 percentage points. Even when compared to high-performing transfer models like ResNet101 (97.60%), EfficientNet_B0 (96.40%), and NFNet_F0 (97.20%), RegNet_y_8gf + MLP maintains its lead, surpassing them by 1.10, 2.30, and 1.50 percentage points, respectively. Furthermore, its accuracy is 2.00 percentage points higher than that of the baseline model RegNet_y_8gf (96.70%).

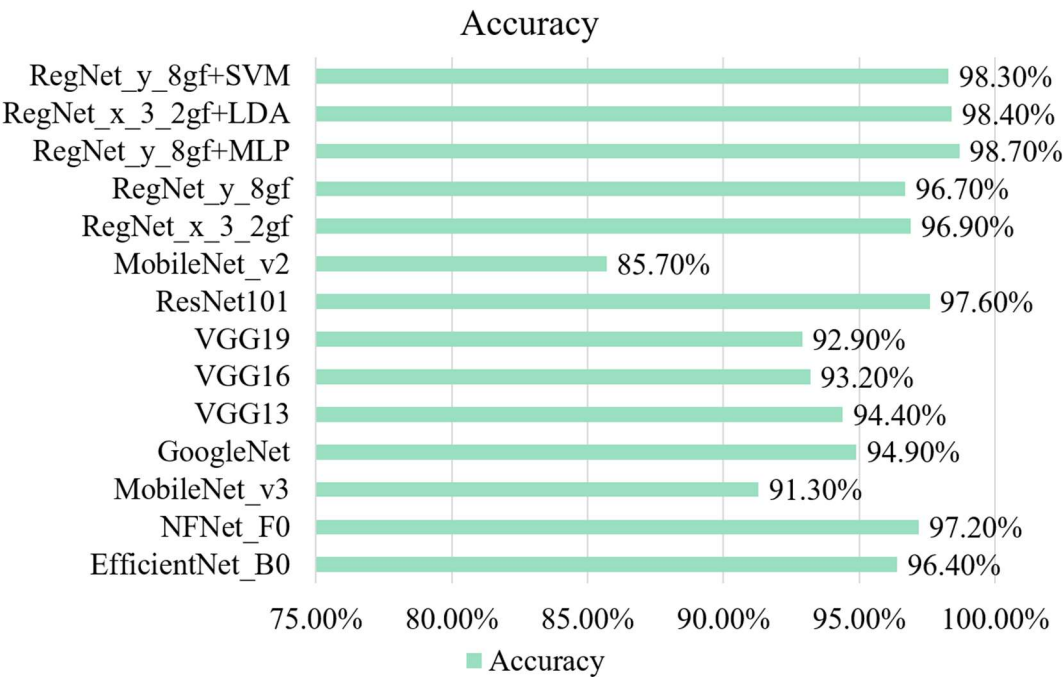


FIGURE 7. Accuracy comparison: optimal combinations and transfer learning models.

Figure 8 illustrates the training time comparison between the three optimal combinations and seven migration learning models. The results highlight that our proposed optimal combinations require substantially less training time than the migration learning models. The RegNet_y_8gf + MLP combination reduces the training

time by 1,408.88 s compared to RegNet_y_8gf alone; RegNet_x_3_2gf + LDA reduces it by 580.69 s compared to RegNet_x_3_2gf; and RegNet_y_8gf + SVM shows a reduction of 1430.18 s compared to RegNet_y_8gf. This modular strategy achieves a high accuracy with a significantly reduced training burden.

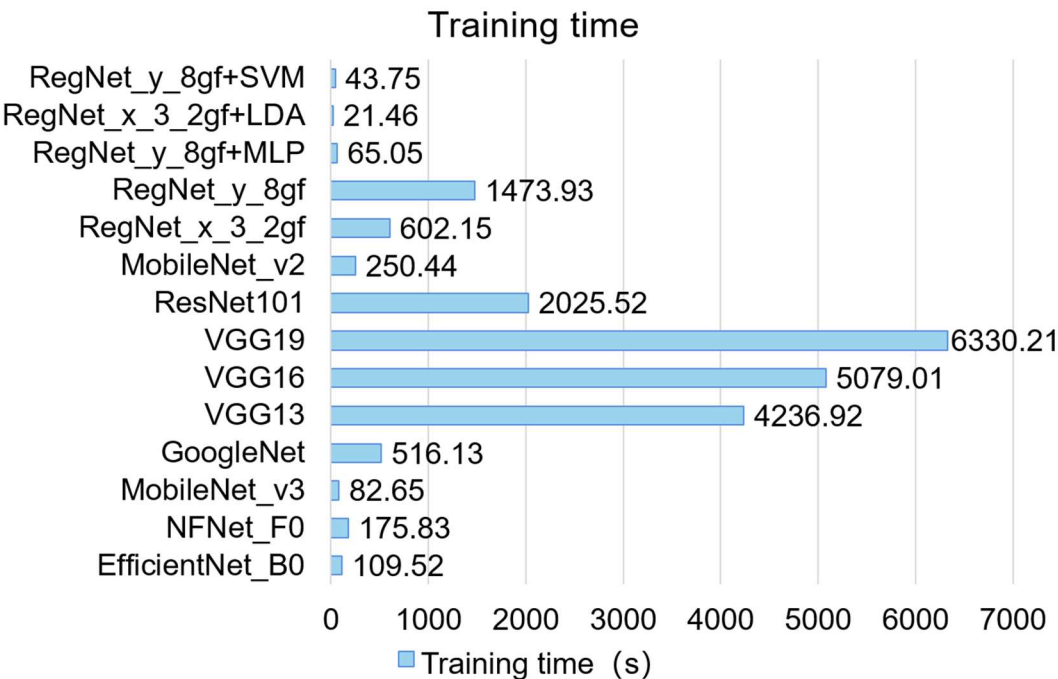


FIGURE 8. Training time comparison: optimal combinations and transfer learning models.

CONCLUSIONS

To address the challenge of balancing high accuracy and computational efficiency in cherry appearance quality recognition, we propose a novel method based on the RegNet model. This approach achieves a higher recognition accuracy and significantly reduces the training time.

In terms of accuracy, our hybrid model combinations consistently outperformed their standalone RegNet counterparts. Specifically, the RegNet_y_8gf + MLP combination improved the accuracy by 2.0% over the original RegNet_y_8gf, demonstrating strong robustness. The RegNet_x_3_2gf + LDA combination showed a 1.5% accuracy increase compared to the original RegNet_x_3_2gf. Beyond improved accuracy, our method also significantly reduced the training time. For instance, our proposed RegNet_y_8gf + MLP combination reduced the training time by 1,408.88 s compared to the standalone RegNet_y_8gf model. Similarly, the RegNet_x_3_2gf + LDA combination cut the training time by 580.69 s compared to RegNet_x_3_2gf. Among various hybrid model combinations, the RegNet_y_8gf + MLP combination performed best.

Moving forward, we will optimize the RegNet architecture to enhance its feature extraction capabilities and effectively reduce the processing time. Concurrently, to improve the model's generalization ability, we will diversify our datasets and extend its application to appearance recognition and detection for other fruits.

FUNDING

The work is partly supported by the Scientific Research Program funded by the Education Department of Shaanxi Provincial Government (Program No. 22JY025).

REFERENCES

Bai, Y., Yu, J., Yang, S., & Ning, J. (2024). An improved YOLO algorithm for detecting flowers and fruits on strawberry seedlings. *Biosystems Engineering*, 237, 1-12. <https://doi.org/10.1016/j.biosystemseng.2023.11.008>

Barrios-Rodríguez, Y., Collazos-Escobar, G. A., & Gutiérrez-Guzmán, N. (2021). ATR-FTIR for characterizing and differentiating dried and ground coffee cherry pulp of different varieties (Coffea Arabica L.). *Engenharia Agrícola*, 41, 70-77. <https://doi.org/10.1590/1809-4430-Eng.Agric.v41n1p70-77/2021>

Cahyo, A., Sudarsono, A., & Yuliana, M. Y. (2023). Multi-layer perceptron neural network implementation as train type classification. *Indonesian Journal of Computer Science*. <https://doi.org/10.33022/ijcs.v12i3.3204>

Chen, X., Tao, H., Zhou, H., Zhou P., & Deng Y. (2024). Hierarchical and progressive learning with key point sensitive loss for sonar image classification. *Multimedia Systems*, 30, 380. <https://doi.org/10.1007/s00530-024-01590-8>

- Chitty-Venkata, K. T., Emani, M., Vishwanath, V., & Somani, A. K. (2023). Neural architecture search benchmarks: Insights and survey. *IEEE Access*, 11, 25217-25236. <https://doi.org/10.1109/ACCESS.2023.3253818>
- Cui, Y., Tang, B., Wu, G., Li, L., Zhang, X., Du, Z., & Zhao, W. (2024). Classification of dog breeds using convolutional neural network models and support vector machine. *Bioengineering*, 11(11), 1157. <https://doi.org/10.3390/bioengineering11111157>
- Dong, Z., & Wang, Y. (2023). Crop disease and pest identification technology based on ACPSo-SVM algorithm optimization. *Engenharia Agricola*, 43(5), e20230104. <https://doi.org/10.1590/1809-4430-Eng.Agric.v43n5e20230104/2023>
- Elasan, S., & Keskin, S. (2019). Classification of variables affecting birth weight by decision trees and k-nearest neighbor methods. *International Journal of Scientific and Technological Research*, 5, 112-119. <https://doi.org/10.7176/jstr/5-12-12>
- Faienza, M. F., Corbo, F., Carocci, A., Catalano, A., Clodoveo, M. L., Grano, M., & Portincasa, P. (2020). Novel insights in health-promoting properties of sweet cherries. *Journal of Functional Foods*, 69, 103945. <https://doi.org/10.1016/j.jff.2020.103945>
- Fonseca, C. S. D., Nhantumbo, B. G., Ferreira, Y. M., Silva, L. A. D., & Costa, A. G. (2025). Computational vision for tomato classification using a decision tree algorithm. *Engenharia Agricola*, 45, e20240124. <https://doi.org/10.1590/1809-4430-Eng.Agric.v45e20240124/2025>
- He, D., Wang, P., Niu, T., Mao Y., & Zhao Y. (2022). Classification model of grape downy mildew disease degree in field based on improved residual network. Nongye Jixie Xuebao/ Transactions of the Chinese Society for Agricultural Machinery. pp. 235-243. <http://dx.doi.org/10.6041/j.issn.1000-1298.2022.01.026>
- Kheiralipour, K., & Pormah, A. (2017). Introducing new shape features for classification of cucumber fruit based on image processing technique and artificial neural networks. *Journal of Food Process Engineering*, 40(6), e12558. <https://doi.org/10.1111/jfpe.12558>
- Li L, Zhang S, Wang B. (2021). Apple leaf disease identification with a small and imbalanced dataset based on lightweight convolutional networks. *Sensors*, 22(1), 173–173. <https://doi.org/10.3390/s22010173>
- Liu, M., Zhang, X., Wang, X., Wei, Z., & Cheng, X. (2024). Research on maize stem recognition based on machine vision. *Engenharia Agricola*, 44, e20240028. <https://doi.org/10.1590/1809-4430-Eng.Agric.v44e20240028/2024>
- Liu, R., Tan, F., & Ma, B. (2023). Intelligent identification of rice growth period (GP) based on Raman spectroscopy and improved CNN in Heilongjiang province of China. *Engenharia Agricola*, 43(6), e20230127. <https://doi.org/10.1590/1809-4430-Eng.Agric.v43n6e20230127/2023>
- Manickam, H. (2025). Apple Fruit Disease Identification Using a Novel Gray-Scale Segmentation Algorithm and Sorting Using a Naive Bayesian Classifier. *Indian Journal of Science and Technology*, 18(19), 1489-1497. <https://doi.org/10.17485/IJST/v18i19.1782>
- Miranda-Vega, J. E., Rivas-Lopez, M., & Fuentes, W. F. (2021). K-nearest neighbor classification for pattern recognition of a reference source light for machine vision system. *IEEE Sensors Journal*, 21(10), 11514–11521. <https://doi.org/10.1109/JSEN.2020.3024094>
- Momeny, M., Jahanbakhshi, A., Jafarnejad, K., & Zhang, Y. D. (2020). Accurate classification of cherry fruit using deep CNN based on hybrid pooling approach. *Postharvest Biology and Technology*, 166, 111204. <https://doi.org/10.1016/j.postharvbio.2020.111204>
- Murmu, A., & Kumar, P. (2024). DLRNet: deep learning with random forest network for classification and detection of malaria parasite in blood smear. *Multimedia Tools & Applications*, 83(23). <https://doi.org/10.1007/s11042-023-17866-6>
- Neitzel, G., Carpes, R., Monteiro, R. D. C., & Gadotti, G. I. (2025). Digital devices for grain classification: efficiency and accuracy in the food industry. *Engenharia Agricola*, 45(spe1), e20240183. <https://doi.org/10.1590/1809-4430-Eng.Agric.v45nespe120240183/2025>
- Patil, B. M., & Burkpalli, V. (2025). Performance evaluation of the classifiers based on features from cotton leaf images. *Multimedia Tools and Applications*, 1-24. <https://doi.org/10.1007/s11042-025-20670-z>
- Radosavovic, I., Kosaraju, R. P., Girshick, R., He, K., & Dollár, P. (2020). Designing network design spaces. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 10428–10436). <https://doi.org/10.1109/CVPR42600.2020.01044>
- Sanches, J., Durigan, J. F., & Durigan, M. F. (2008). Application of mechanical damages in avocados and their effects to the quality of the fruits. *Engenharia Agricola*, 28, 164-175. <https://doi.org/10.1590/S0100-69162008000100017>
- Song, K., Yang, J., & Wang, G. (2023). A Swin Transformer and MLP based method for identifying cherry ripeness and decay. *Frontiers in Physics*, 11, 1278898. <https://doi.org/10.3389/fphy.2023.1278898>
- Suesse, T., Brenning, A., & Grupp, V. (2023). Spatial linear discriminant analysis approaches for remote-sensing classification. *Spatial Statistics*, 57. <https://doi.org/10.1016/j.spasta.2023.100775>
- Wang, G., Zheng, H., & Li, X. (2023). ResNeXt-SVM: a novel strawberry appearance quality identification method based on ResNeXt network and support vector machine. *Journal of Food Measurement and Characterization*, 17(5), 4345-4356. <https://doi.org/10.1007/s11694-023-01959-9>
- Wang, G., Zheng, H., Zhang, X. (2022). A robust checkerboard corner detection method for camera calibration based on improved YOLOX. *Frontiers in Physics*, 9, 819019. <https://doi.org/10.3389/fphy.2021.819019>

- Wang, J., Liu, M., Du, Y., Zhao, M., Jia, H., Guo, Z., & Liu, Y. (2024). PG-YOLO: An efficient detection algorithm for pomegranate before fruit thinning. *Engineering Applications of Artificial Intelligence*, 134, 108700. <https://doi.org/10.1016/j.engappai.2024.108700>
- Wang, Z., Tao, H., Zhou, H., Deng Y & Zhou P. (2025). A content-style control network with style contrastive learning for underwater image enhancement. *Multimedia Systems* 31, 60. <https://doi.org/10.1007/s00530-024-01642-z>
- Yang, J., & Wang, G. (2023). Identifying cherry maturity and disease using different fusions of deep features and classifiers. *Journal of Food Measurement and Characterization*, 17, 5794–5805. <https://doi.org/10.1007/s11694-023-02091-4>
- Zhang, L., Zhou, W., & Jiao, L. (2004). Wavelet support vector machine. *IEEE Trans Syst. Man Cybern.* 34, 34–39. <https://doi:10.1109/tsmcb.2003.811113>
- Zheng, H., Wang, G., & Li, X. (2022). Swin-MLP: a strawberry appearance quality identification method by Swin Transformer and multi-layer perceptron. *Journal Food Measurement and Characterization*, 16, 2789. <https://doi.org/10.1007/s11694-022-01396-0>
- Zhu, H., Yang, L., Fei, J., Zhao, L., & Han, Z. (2021). Recognition of carrot appearance quality based on deep feature and support vector machine. *Computers and Electronics in Agriculture*, 186, 106185. <https://doi.org/10.1016/j.compag.2021.106185>