Academic Journal of Science and Technology
ISSN: 2771-3032 | Vol. 20, No. 1, 2026

Research on Grading Method of Korla Fragrant Pears Based on Gradient Boosting Decision Tree

Shuting Xu, Shuo Yang, Mengyuan Tao, Yixin Gao, Zhenghao Cui, Yingchao Wang *

Xinjiang Institute of Science and Technology, Korla, Xinjiang, 841000, China

*** Corresponding author:** Yingchao Wang

Abstract: To address the problems of traditional Korla fragrant pear grading relying on manual experience, strong subjectivity, low efficiency and insufficient grading accuracy, an intelligent grading method for Korla fragrant pears based on Gradient Boosting Decision Tree (GBDT) was proposed. Taking Korla fragrant pears as the research object, the grading indicators were determined in accordance with Xinjiang local standards and group standards, and four core features including single fruit weight, peeled fruit firmness, soluble solid content and fruit surface defect degree were selected as the model input to construct a GBDT model for pear grading. A total of 360 pear sample data were collected on site, and after data preprocessing and feature optimization, the model was trained, verified and parameter-tuned. Comparative experiments were conducted with traditional machine learning algorithms such as Random Forest (RF) and Support Vector Machine (SVM). The results showed that the grading accuracy of the constructed GBDT grading model on the test set reached 95.83%, which was higher than that of the RF model (91.67%) and the SVM model (88.33%). Moreover, it performed better in terms of recall rate and precision rate for the grading of super grade, first grade and second grade fragrant pears, and could effectively distinguish Korla fragrant pears of different grades. This method realizes the objective, accurate and efficient grading of Korla fragrant pears, and provides technical support and theoretical reference for the intelligent sorting of the fragrant pear industry.

Keywords: Korla Fragrant Pear; Gradient Boosting Decision Tree; Intelligent Grading.

1. Introduction

As a characteristic fruit product of Xinjiang, Korla fragrant pears are renowned at home and abroad for their unique quality of thin peel, fine flesh, crispness, juiciness, sweetness and delicate fragrance, and are one of the core pillars of Xinjiang's fruit industry. The grade of fragrant pears directly determines their market value, circulation channels and economic benefits, and scientific and reasonable grading is a key link to enhance the competitiveness of the fragrant pear industry and standardize market order. At present, the grading of Korla fragrant pears mainly adopts a combination of manual grading and traditional mechanical grading. Manual grading relies on the visual and tactile experience of operators, which has problems such as non-uniform grading standards, strong subjectivity, low efficiency, high labor intensity, and is prone to missing and misgrading. Traditional mechanical grading mostly only targets a single physical indicator such as single fruit weight, ignoring core quality indicators such as soluble solids and fruit firmness, with limited grading accuracy and failure to fully reflect the comprehensive quality level of fragrant pears. With the wide application of machine learning, computer vision and other technologies in the agricultural field, intelligent grading methods based on algorithm models have become a research hotspot due to their advantages of objectivity, high efficiency and high precision. Among them, the Gradient Boosting Decision Tree (GBDT) algorithm has the advantages of strong fitting ability for nonlinear data, good anti-overfitting performance, and no need for complex feature preprocessing, and has shown good application potential in the field of agricultural product grading. However, in existing studies: Xu et al. [1] proposed an improved LB-YOLO model to realize the detection of loquat damage areas and the estimation of damage volume ratio, providing a new method for non-destructive testing of fruits; Wanthong et al. [2] constructed a real-time sorting and grading system for Thai export fruits combining computer vision and AI, optimized the radiation dose adjustment process, and reduced energy consumption and costs, but this study is aimed at Thai export fruits, which is different from the grading demand of Korla fragrant pears and difficult to apply directly; Zhang et al. [3] built a multi-fruit automatic grading system based on machine learning, realizing the accurate grading of four kinds of fruits such as apples and pomegranates, and providing a general idea for the grading of multi-variety fruits; Wang et al. [4] improved the YOLOv8s lightweight model to solve the problems of foliage occlusion and slight grade differences in persimmon orchard grading, providing technical reference for fruit grading in orchard scenarios; Nikam et al. [5] fused deep CNN features with ensemble learning to construct a pitaya quality grading and maturity detection model, improving grading accuracy, and its feature fusion strategy can be used for reference in fragrant pear grading; Arwatchananukul et al. [6] used YOLOv8 and image processing technology to realize the automatic extraction of mango geometric features and optimize the quality evaluation process, providing a reference for the automatic detection of fruit surface defects and geometric features of fragrant pears; Li et al. [7] constructed a wolfberry damage degree grading model based on electrical characteristics, providing a new perspective for fruit damage detection; Ito et al. [8] developed a pear grading support system based on edge computing, effectively improving grading efficiency and providing ideas for the efficiency optimization of fragrant pear grading; Long et al. [9] improved the detection accuracy of litchi soluble solid and anthocyanin content through spectral normalization and multi-feature fusion methods, and realized litchi grading, whose multi-feature fusion idea can be used to improve the grading accuracy of fragrant pears. However, in existing

studies, there are few studies on the application of the GBDT algorithm in the grading of Korla fragrant pears, and the grading requirements of Xinjiang local standards and group standards are not fully combined. Compared with the above advanced fruit grading studies carried out in sequence, fragrant pear grading still has gaps in the automatic detection of indicators, model precision optimization, multi-feature fusion and other aspects, and the practicality and pertinence of the grading model are insufficient.

2. Data Collection and Preprocessing

2.1. Sample Collection

The experimental samples were collected from the core producing areas of Korla City, Xinjiang (Luntai County and Yuli County). Korla fragrant pears with consistent maturity and no serious damage were selected as experimental samples, with a total of 360 samples collected, including 120 super grade, 120 first grade and 120 second grade fragrant pears. Sample collection followed GB/T 8855-2018, combined with the sampling requirements of relevant standards, and random sampling was conducted among fragrant pears of different batches and picking times to ensure the representativeness and diversity of the samples.

2.2. Data Preprocessing

The collected original data may have problems such as missing values and outliers, which will affect the training effect and grading accuracy of the model. Therefore, it is necessary to preprocess the data set, and the specific steps are as follows:

2.2.1. Outlier Processing

The box plot method was used to detect outliers. The quartiles (Q1, Q3) and interquartile range (IQR = Q3 - Q1) of each characteristic index were calculated, and the data beyond the range of [Q1 - 1.5×IQR, Q3 + 1.5×IQR] were determined as outliers. After detection, there were 8 outliers in the 360 sample data (mainly abnormal single fruit weight and fruit firmness), and the median of the characteristic index was used to replace the outliers to avoid the interference of outliers on model training.

2.2.2. Missing Value Processing

Through the investigation of the data set, it was found that some samples had missing soluble solid content (5 samples in total), and the average value of soluble solid content of fragrant pears of the same grade was used for filling to ensure the integrity of the data set.

2.2.3. Feature Standardization

Due to the different dimensions of the four core features (e.g., the unit of single fruit weight is g, and the unit of fruit firmness is kg/cm²), the model will assign unbalanced weights to different features, which affects the model training effect. The Z-score standardization method was used to standardize the feature data, converting all feature data into standardized data with a mean of 0 and a standard deviation of 1. The standardization formula is $x^{'}=\frac{x-\mu}{\sigma}$, where μ is the feature mean and σ is the feature standard deviation.

2.2.4. Data Set Division

The preprocessed data set was randomly divided into a training set and a test set at a ratio of 7:3. The training set contained 252 samples (72 super grade, 72 first grade, 108 second grade) for model training and parameter tuning; the test set contained 108 samples (48 super grade, 48 first grade, 12 second grade) for model performance verification.

3. Introduction to GBDT

Gradient Boosting Decision Tree (GBDT) is a supervised learning algorithm based on ensemble learning, which belongs to the Boosting algorithm family. Its core idea is to train a series of weak learners (usually decision trees) through iteration. Each new weak learner is committed to correcting the prediction error of the previous learner, and finally the prediction results of all weak learners are weighted and combined to form a high-precision strong learner.

Given a training data set $D=\{(x_1,y_1),(x_2,y_2),\ldots,(x_n,y_n)\}$, where $x_i \in X \subseteq \mathbb{R}^d$ is the input feature vector (the four core grading indicators of fragrant pears in this paper), $y_i \in Y$ is the output label (the grade of fragrant pears in this paper), and the loss function is $L(y,F(x))$. The specific process of the GBDT algorithm is as follows:

1. Initialize the model: Initialize the strong learner with a constant predicted value to minimize the loss function, i.e., $F_0(x)=\arg\min_{\gamma} \sum_{i=1}^{n} L(y_i,\gamma)$;

2. Iteratively train weak learners: For $m=1,2,\ldots,M$ (M is the number of weak learners):

(1) Calculate the negative gradient (pseudo-residual) of the current model: $r_{im}=-\left(\frac{\partial L(y_i,F(x_i))}{\partial F(x_i)}\right)_{F(x)=F_{m-1}(x)}$, where r_{im} is the pseudo-residual of the i-th sample in the m-th iteration;

(2) Fit the pseudo-residual r_{im} with a decision tree to obtain the m-th weak learner $f_m(x)$;

(3) Determine the learning rate η through line search to minimize the loss function $\sum_{i=1}^{n} L(y_i,F_{m-1}(x_i)+\eta f_m(x_i))$;

(4) Update the strong learner: $F_m(x)=F_{m-1}(x)+\eta f_m(x)$;

3. Output the final model: $F(x)=F_M(x)=\sum_{m=1}^{M} \eta f_m(x)$ to complete the prediction of fragrant pear grades.

4. Experimental Results and Analysis

4.1. Model Training Results

The optimized GBDT model, RF model and SVM model were respectively trained on the training set, and the results of 5-fold cross-validation are shown in Table 1. It can be seen from Table 1 that the cross-validation accuracy of the GBDT model on the training set was 97.62%, which was higher than that of the RF model (94.44%) and the SVM model (90.08%), indicating that the GBDT model had a better fitting effect on the training set and could better learn the mapping relationship between fragrant pear grading indicators and grades.

Table 1. Model Training Results

Model	Cross-validation Accuracy (%)	Cross-validation F1-score
GBDT	97.62	0.975
RF	94.44	0.943
SVM	90.08	0.898

4.2. Model Test Results

The three trained models were used for prediction on the test set, and various evaluation indicators were calculated. The test results are shown in Table 2.

Table 2. Test Results of Each Model

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-score
GBDT	95.83	95.91	95.76	0.958
RF	91.67	91.72	91.60	0.917
SVM	88.33	88.45	88.21	0.883

It can be seen from Table 2 that all evaluation indicators of the GBDT model on the test set were superior to those of the RF model and the SVM model. The accuracy reached 95.83%, which was 4.16 percentage points higher than the RF model and 7.5 percentage points higher than the SVM model; the precision, recall and F1-score were also at a high level, indicating that the GBDT model had good generalization ability and grading accuracy, and could accurately distinguish Korla fragrant pears of different grades.

4.3. Analysis of Grading Results of Fragrant Pears of Different Grades

To further analyze the grading effect of the GBDT model on fragrant pears of different grades, the precision, recall and F1-score of the model for super grade, first grade and second grade fragrant pears were calculated respectively, and the results are shown in Table 3.

Table 3. Grading Results of Fragrant Pears of Different Grades

Grade	Precision (%)	Recall (%)	F1-score
Super grade	97.92	95.83	0.969
First grade	95.83	97.92	0.969
Second grade	93.75	93.75	0.938

It can be seen from Table 3 that the GBDT model had a better grading effect on super grade and first grade fragrant pears, with the precision and recall both reaching more than 95% and the F1-score being 0.969; the grading accuracy for second grade fragrant pears was relatively slightly lower, but the precision, recall and F1-score all reached more than 93%. The reason analysis shows that some indicators of second grade fragrant pears (such as single fruit weight and fruit firmness) overlap with those of first grade fragrant pears to a certain extent, and the judgment of fruit surface defect degree has slight subjectivity, leading to some second grade fragrant pears being misjudged as first grade ones, but the overall grading effect still meets the actual production needs.

5. Conclusion

Aiming at the problems of strong subjectivity, low efficiency and insufficient accuracy of traditional Korla fragrant pear grading methods, this paper proposed an intelligent grading method for fragrant pears based on Gradient Boosting Decision Tree (GBDT). Through experimental research, the following conclusions were drawn: Combined with Xinjiang local standards and group standards, four core grading indicators including single fruit weight, peeled fruit firmness, soluble solid content and fruit surface defect degree were screened out, and the constructed grading indicator system was in line with the actual production needs and could fully reflect the comprehensive quality of Korla fragrant pears [1][3]; after parameter tuning, the constructed GBDT grading model achieved a grading accuracy of 95.83% on the test set, with precision, recall and F1-score all reaching more than 95%, which was significantly better than the RF model (91.67%) and the SVM model (88.33%), and had good grading accuracy and generalization ability; feature importance analysis showed that single fruit weight was the most critical feature affecting fragrant pear grading, followed by soluble solid content, peeled fruit firmness and fruit surface defect degree, and the results could provide a reference for the optimization of fragrant pear grading indicators and the improvement of detection efficiency; the Korla fragrant pear grading method based on GBDT realized the objective, accurate and efficient grading of fragrant pears, solved the drawbacks of traditional grading methods, and provided a feasible technical scheme for the intelligent sorting of the fragrant pear industry.

6. Discussion

In summary, combined with the experimental results of this study and existing relevant research, the Korla fragrant pear grading method based on GBDT effectively solves the drawbacks of strong subjectivity, low efficiency and insufficient accuracy of traditional grading. The accuracy of the constructed model on the test set reaches 95.83%, which is significantly better than the RF and SVM models and also higher than the accuracy of existing similar fruit grading studies, verifying the applicability and superiority of the GBDT algorithm in fragrant pear grading. In addition, the four selected core grading indicators are in line with Xinjiang local standards and industrial practice, taking into account the physical, quality and appearance characteristics of the fruit. The method has low cost and strong practicability, and is suitable for promotion in small and medium-sized orchards and processing enterprises. However, the research still has limitations: the sample size is limited and the sources are concentrated; the detection of indicators such as fruit surface defects relies on manual detection with subjectivity; deep learning technology and more appearance features are not integrated, and there is still room for optimization compared with existing advanced grading methods. In the follow-up, the research can be improved by expanding the sample coverage, introducing machine vision to realize automatic indicator detection, optimizing the model structure, etc., to further improve the scientificity and practicability of the grading method and help the intelligent upgrading of the Korla fragrant pear industry.

Acknowledgments

Funding: Industry-Education Integration and New Business Development Research Center Project (2025-KYJ D07).

References

[1] Xu C, Feng L, Huang X, et al. A lightweight network for non-destructive and precise prediction of loquat bruising volume based on RGB images[J]. Food Control,2026,184:112044.

[2] Wanthong I, On S T, Rodporn S, et al. A Computer Vision and AI-Based System for Real-Time Sizing and Grading of Thai Export Fruits[J]. AgriEngineering,2025,7(11):377.

[3] Zhang P, Li X. Automatic Fruit Grading System with High Adaptability Using Machine Learning Method[J]. Applied Sciences,2025,15(22):11866.

[4] Wang H, Jing Y. Persimmon Fruit Quality Grading Detection Based on an Improved YOLOv8s Lightweight Model[J]. Journal of Electronic Research and Application,2025,9(5):209-218.

[5] Nikam B S, Lande B S, Nagalkar J V, et al. Predictive Classification Model for Quality Grading and Maturity Detection of Dragon Fruit Using Fused Deep CNN Feature and Ensemble Learning[J]. Journal of Food Processing and Preservation,2025,2025(1):6938071.

[6] Arwatchananukul S, Wongsawat S, Chaiwong S, et al. Automated Physical Feature Extraction of Namdokmai Sithong Mangoes Using YOLOv8 and Image Processing Techniques[J]. AgriEngineering,2025,7(9):312.

[7] Li H J, Ma F L, Zhang W W, et al. Grading the damage degree of fresh Lycium barbarum L. fruits based on electrical characteristics[J]. Information Processing in Agriculture,2025, 12 (3):398-407.

[8] Ito R, Konuma S, Yamazaki T. Grading Support System for Pear Fruit Using Edge Computing[J]. Engineering Proceedings, 2025,107(1):45.

[9] Long T, Wei K, Huang B, et al. Improved detection of soluble solid and anthocyanin in litchi fruits by normalized VNIR-SWIR transmittance hyperspectral imaging and SPF-SSF-GSF-NLF fusing method[J]. Food Chemistry,2025,493(Pt 4):145987.