Print ISSN: 3104-932X
Online ISSN: 3104-9338

Design of an Intelligent Tea-Picking Robot Based on STM32 and Machine Vision

Chensongyi Guo, Lei Lin, Bingkun Dan, Junyan Xu*

Wuyi University, Wuyishan, Fujian 354300, China

Received: Oct 15, 2025

Revised: Nov 6, 2025

Accepted: Nov 17, 2025

Published online: November 19, 2026

To appear in: *International Journal of Advanced AI Applications*, Vol. 3, No. 1 (January 2027)

*Corresponding Author: Junyan Xu (360128019@qq.com)

Abstract. To improve the recognition rate of high-quality tea shoots and picking efficiency while reducing labor costs, this study develops an intelligent tea-picking robot for a tea plantation in Wuyishan, China. Based on the operational requirements of the plantation, the main functional modules of the robot are identified and its overall system layout is established. After completing the mechanical design and three-dimensional modeling, the STM32 is selected as the main controller. Hardware components are selected and circuit designs are developed for the machine-vision module, robotic-arm picking module, and infrared sensing module. The basic system functions are then implemented through software programming. Finally, YOLOv8 is used to train a dataset for tea-shoot quality recognition and to evaluate system performance. The test results indicate that the recognition performance generally meets the design requirements. This work provides a foundation for the subsequent automatic picking of high-quality tea shoots by the robot.

Keywords: *STM32; machine vision; YOLOv8; intelligent tea-picking robot; host computer*

1. Introduction

In 2025, China's tea-plantation area exceeded 50 million mu, and tea production reached 3.92 million tonnes, ranking first in the world [1]. Therefore, identifying tea quality during harvesting and improving the efficiency of picking high-quality tea have considerable economic value.

In recent years, universities and enterprises in China have conducted extensive research on tea-quality identification using machine-vision technology. For example, in 2023, Wang Mengni, Gu Jinan, Wang Huajia, and colleagues at Jiangsu University improved the accuracy of tender tea-shoot recognition using an enhanced YOLOv5 algorithm [2]. In 2024, Liang Xiao

© 2027 The Author(s). Published by Hong Kong Dawn Clarity Press Limited. This is an open access article distributed under the terms of the Creative Commons Attribution 4.0 International Licence (CC BY 4.0), https://creativecommons.org/licenses/by/4.0/.

of South China Agricultural University constructed a tea-shoot dataset and developed a yield-assessment model based on image recognition, achieving satisfactory accuracy [3]. In 2025, Liao Shuzhen and Zeng Jing of Heyuan Polytechnic established datasets for one-bud-and-one-leaf and one-bud-and-two-leaves categories and combined them with Mosaic data augmentation, reporting satisfactory mean accuracy [4]. In 2026, Zhang Yaoxin of Chongqing University of Technology used binocular vision and an improved YOLOv11s-seg model to achieve high-precision segmentation and pose estimation of tea shoots [5]. These studies mainly use machine vision to recognize tea shoots and have improved recognition accuracy. However, although recognition performance has improved, the subsequent picking process still relies largely on manual labor, resulting in low picking efficiency and high costs. To address these limitations, this study develops a robot that automatically picks tea based on machine-vision recognition, integrating recognition and picking into a single system.

2. Overall System Layout of the Automatic Tea-Picking Robot

Table 1. Mapping Between Functions and System Modules

Function	System control	Locomotion	Picking	Tea recognition	Wireless communication	Line tracking and obstacle avoidance	Human–machine interaction
Module	Main control module	Motor-drive module	Robotic-arm picking module	Machine-vision module	Bluetooth module	Infrared-sensing module	Host-computer module

The overall system layout is shown in Figure 1.

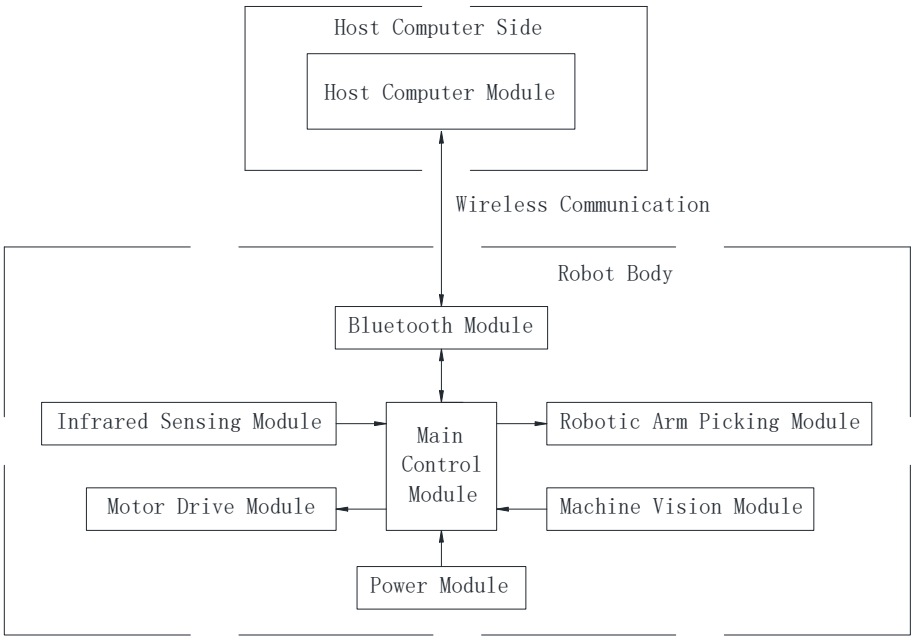


Figure 1. Overall system layout

© 2027 The Author(s). Published by Hong Kong Dawn Clarity Press Limited. This is an open access article distributed under the terms of the Creative Commons Attribution 4.0 International Licence (CC BY 4.0), https://creativecommons.org/licenses/by/4.0/.

The robot developed in this study is intended primarily for operation in tea plantations in Wuyishan. The terrain is mainly flat, and the soil is predominantly sandstone. Accordingly, the robot is designed to perform several functions, including locomotion, tea picking, tea-shoot recognition, wireless communication, and line tracking and obstacle avoidance. The corresponding modules are listed in Table 1.

3. Hardware Design

3.1. Construction of the Robot's Three-Dimensional Model

The three-dimensional model of the tea-picking robot was designed primarily according to the terrain in which it operates and the types of tea to be picked. The design process comprised the following steps: (1) Core components were designed according to the robot's functional requirements. For example, because the operating area is mainly flat but contains considerable loose gravel, a tracked running gear was selected. This mechanism can adapt to terrain in a variety of complex environments [6]. (2) The dimensions, shapes, and mating features of the core components were designed according to their relationships and assembly requirements. For example, when two components are joined by an interference fit, the corresponding mating recess and protruding structure are designed, and the final model is determined according to the interference-fit dimensions. (3) The completed component models are assembled into an integrated structure. The main components include the tracked running gear, picking blade head, support plate, robotic-arm links, and collection funnel. Their three-dimensional models are shown in Figures 2–7.

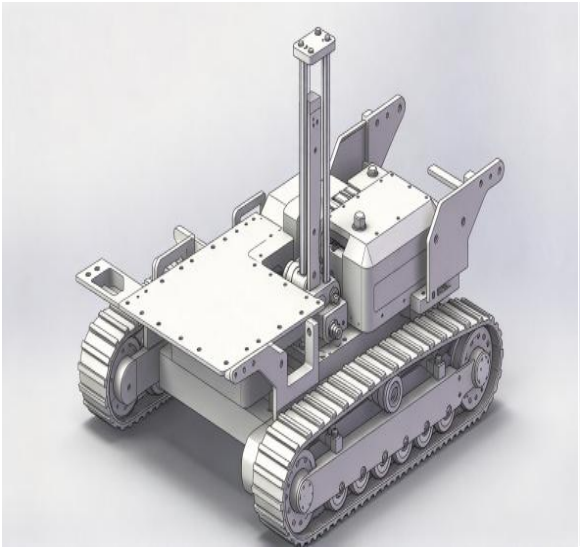


Figure 2. Tracked running gear



Figure 3. Picking blade head

© 2027 The Author(s). Published by Hong Kong Dawn Clarity Press Limited. This is an open access article distributed under the terms of the Creative Commons Attribution 4.0 International Licence (CC BY 4.0), https://creativecommons.org/licenses/by/4.0/.

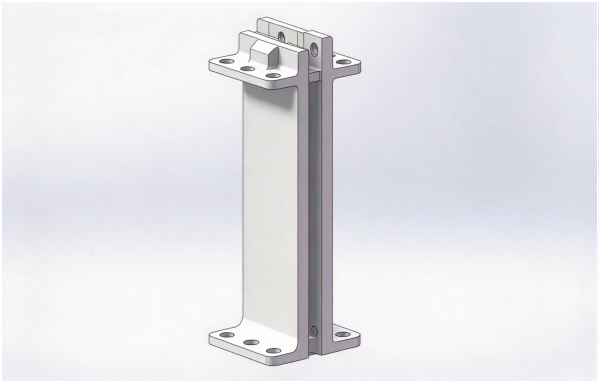


Figure 4. Support plate

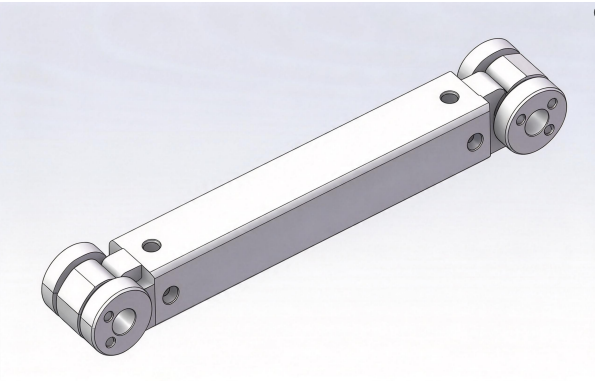


Figure 5. Robotic-arm link

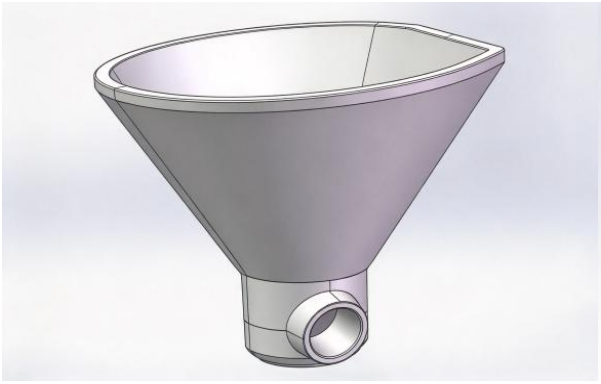


Figure 6. Collection funnel

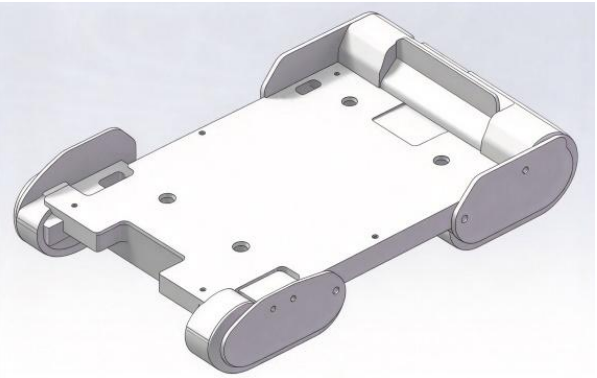


Figure 7. Chassis base plate

The overall assembly of the tea-picking robot is shown in Figure 8.

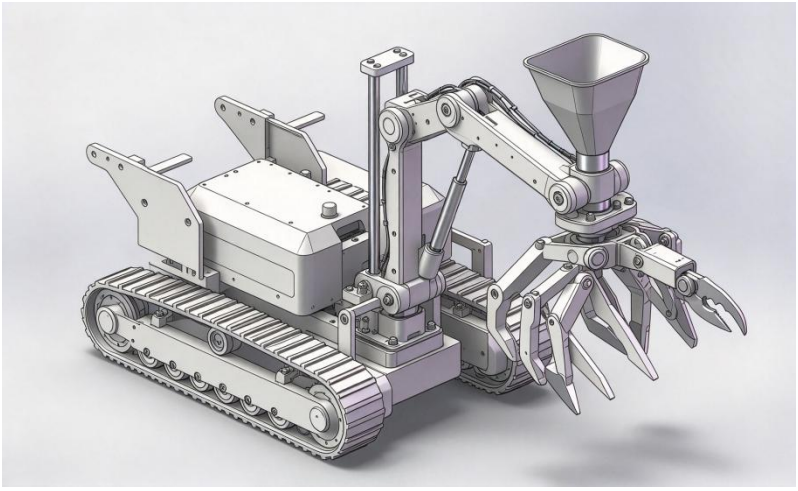


Figure 8. Overall assembly of the tea-picking robot

3.2. Design of the Main Controller

The main controller of the tea-plantation robot must drive the motor module, receive tea-shoot grading information from the machine-vision system, and control the robotic arm.

© 2027 The Author(s). Published by Hong Kong Dawn Clarity Press Limited. This is an open access article distributed under the terms of the Creative Commons Attribution 4.0 International Licence (CC BY 4.0), https://creativecommons.org/licenses/by/4.0/.

Therefore, the STM32F407ZGT6 is selected as the main control chip. It has a clock frequency of 168 MHz, 1 MB of Flash memory, and an integrated floating-point unit (FPU) [7]. The chip also provides multiple serial interfaces. These enable it to perform multi-axis robotic-arm motion calculations while simultaneously controlling the tracked chassis and acquiring visual signals, making it well suited to the robot's operating requirements. The minimum system of the main controller is shown in Figure 9.

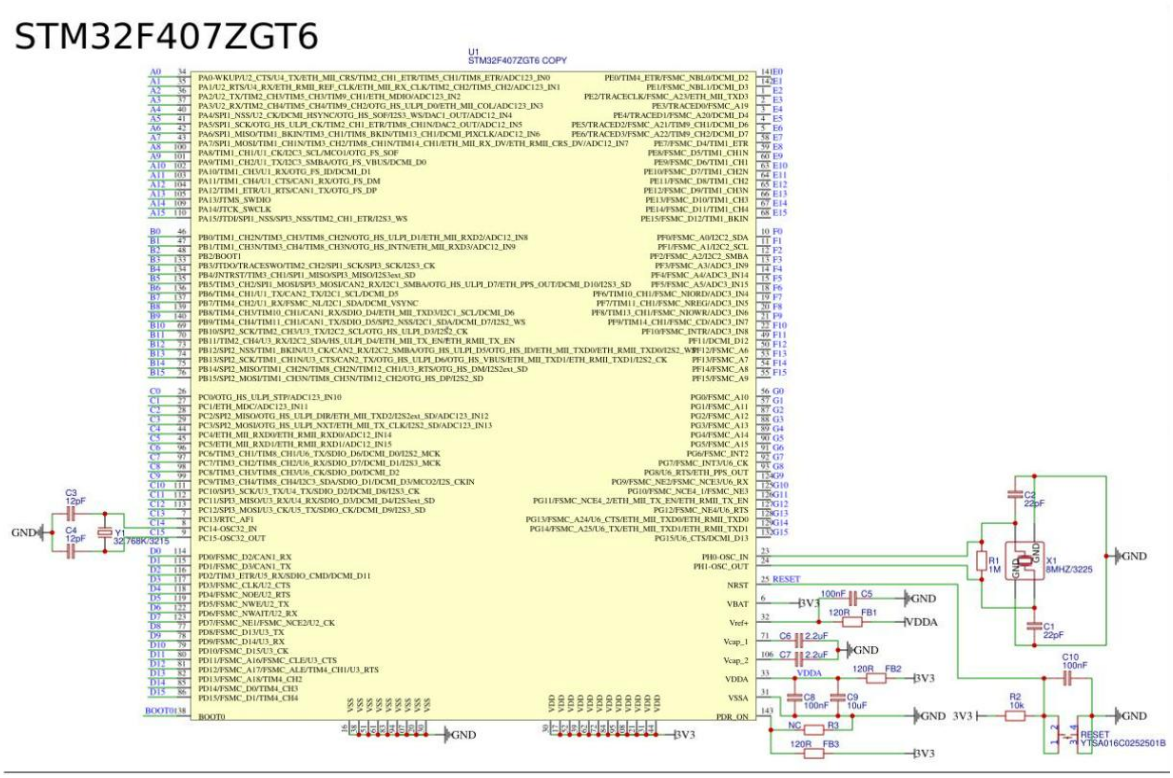


Figure 9. Minimum system circuit

3.3. Design of the Motor-Drive Module

The motor-drive module is designed primarily to ensure that the robot can move quickly and stably throughout the tea plantation. The selected drive motor is the JGB37-520 DC geared motor for three main reasons: (1) High torque. Its rated torque range is 0.015–1.1 N·m, enabling the robot to travel quickly and stably along a preset route in the plantation. (2) Integrated Hall-effect encoder. Closed-loop speed control can keep the two track wheels rotating at the same speed and prevent the robot from deviating from its path. (3) Simple installation. No additional reduction mechanism is required; the motor can be connected directly to the drive wheel [8]. The circuit schematic of the motor-drive module is shown in Figure 10.

© 2027 The Author(s). Published by Hong Kong Dawn Clarity Press Limited. This is an open access article distributed under the terms of the Creative Commons Attribution 4.0 International Licence (CC BY 4.0), https://creativecommons.org/licenses/by/4.0/.

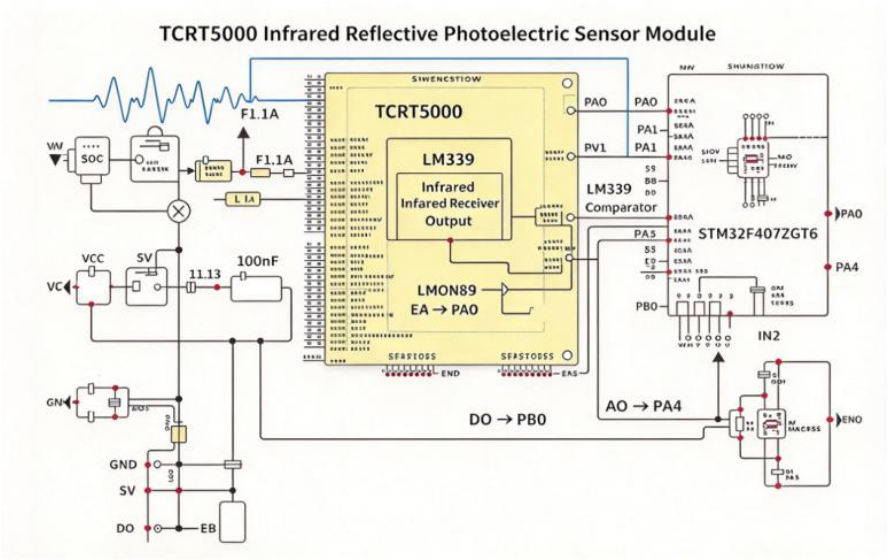


Figure 10. Motor-drive module circuit schematic

3.4. Design of the Infrared-Sensing Module

The TCRT5000 reflective infrared photoelectric sensor is selected for the infrared obstacle-avoidance module. The sensor integrates an infrared emitter and receiver, operates at 3.3–5 V, and has a detection range of 1–8 cm. Its detection sensitivity can be adjusted using a potentiometer. The module outputs a digital-level signal that can be connected directly to a GPIO pin of the main controller [9]. When a tea bush or other obstacle is detected ahead, the sensor output changes logic level. After detecting the obstacle signal, the main controller instructs the tracked chassis to slow down, stop, or turn, thereby providing obstacle avoidance in the tea-plantation environment. The circuit schematic of the TCRT5000 reflective infrared photoelectric sensor module is shown in Figure 11.

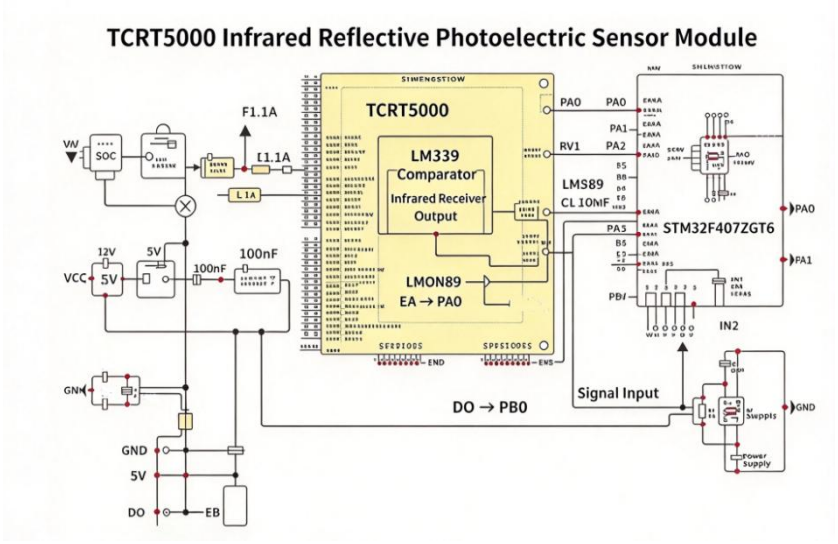


Figure 11. Infrared-sensing module circuit schematic

© 2027 The Author(s). Published by Hong Kong Dawn Clarity Press Limited. This is an open access article distributed under the terms of the Creative Commons Attribution 4.0 International Licence (CC BY 4.0), https://creativecommons.org/licenses/by/4.0/.

3.5. Design of the Robotic-Arm Picking Module

The robotic-arm picking module designed in this study uses a four-degree-of-freedom serial configuration. A combined cutting-and-gripping mechanism is mounted at the end of the actuator. The joints are driven by MG996R metal-gear servos. Five servos are used in total: four joint servos and one picking-actuator servo. The servos have a rated voltage of 5 V and an output torque of 13 kg·cm. Their angles are controlled by 50 Hz PWM signals generated by the timers of the main controller [10]. The servo circuit of the picking module is shown in Figure 12.

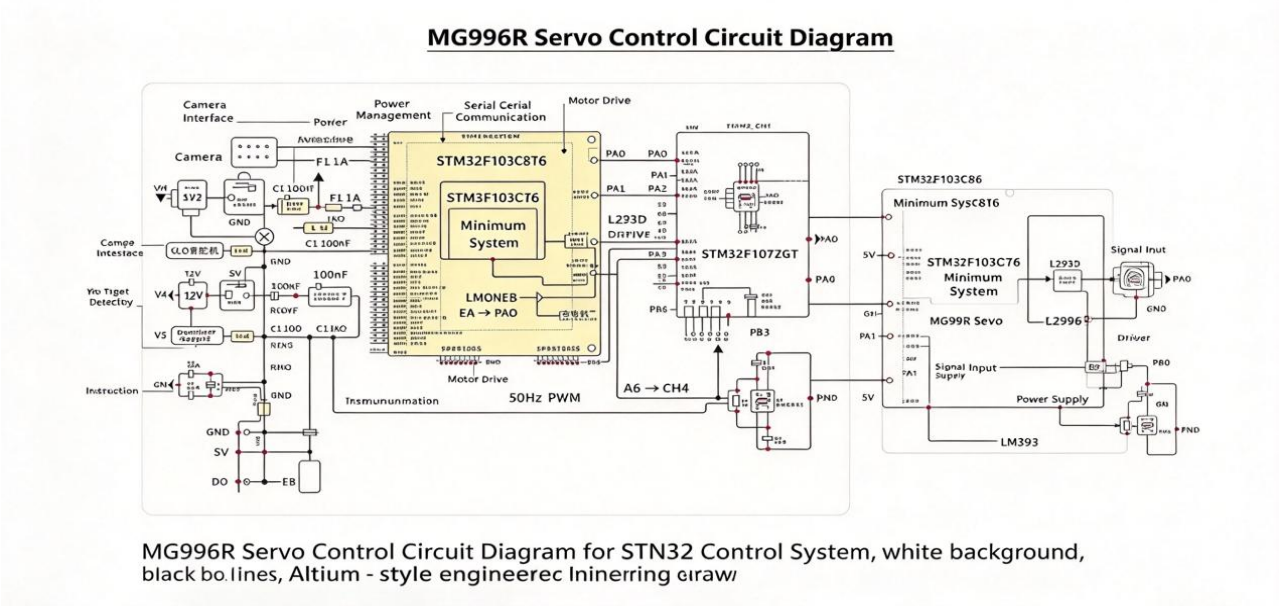


Figure 12. Servo circuit of the picking module

4. Software Design

4.1. Overall Software Architecture

The software adopts a distributed, coordinated architecture consisting of a host computer and a lower-level controller. The host computer uses a YOLO algorithm to acquire and preprocess tea images and identify tea-shoot quality. After selecting high-quality shoots, it sends picking commands to the lower-level controller through a serial port. The lower-level controller, based on the STM32, manages the robot's operating states. After receiving target information, it controls chassis movement, infrared obstacle avoidance, and the robotic-arm picking sequence, including shoot alignment, cutting and gripping, transfer and release, and mechanism reset. The host computer and lower-level controller communicate through a custom serial protocol with error checking, and a communication timeout-protection mechanism is implemented to ensure stable and safe operation in the complex plantation environment. The overall software

© 2027 The Author(s). Published by Hong Kong Dawn Clarity Press Limited. This is an open access article distributed under the terms of the Creative Commons Attribution 4.0 International Licence (CC BY 4.0), https://creativecommons.org/licenses/by/4.0/.

workflow is shown in Figure 13.

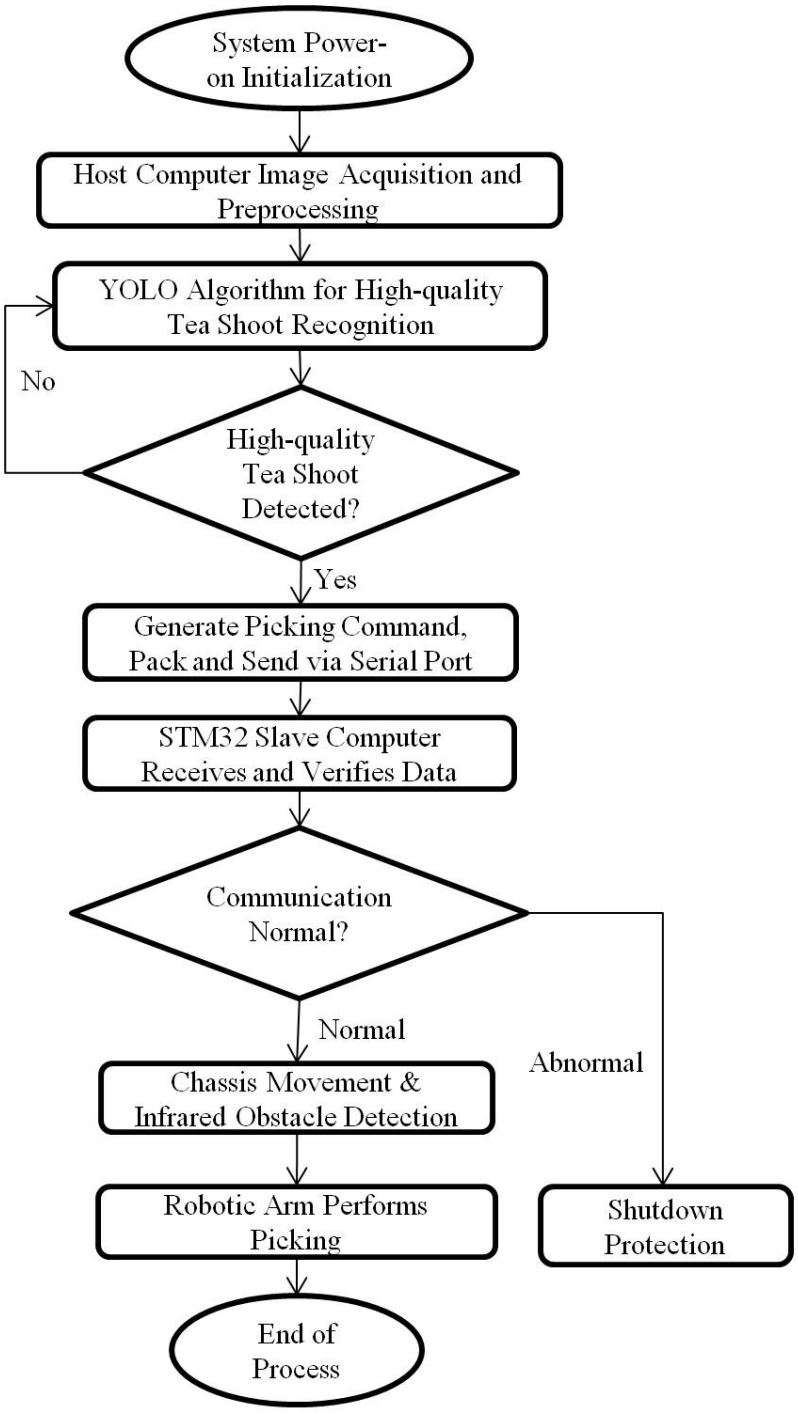


Figure 13. Overall software workflow

4.2. Design of the Machine-Vision Tea-Recognition System

4.2.1. Tea-Shoot Grading and Definition of the Recognition Task

The system classifies the recognition targets into four categories: tender shoots, mature shoots, one bud with one leaf, and one bud with two leaves. The morphological criteria and

© 2027 The Author(s). Published by Hong Kong Dawn Clarity Press Limited. This is an open access article distributed under the terms of the Creative Commons Attribution 4.0 International Licence (CC BY 4.0), https://creativecommons.org/licenses/by/4.0/.

corresponding picking decisions for each category are presented in Table 2.

Table 2. Tea-Shoot Categories and Corresponding Picking Decisions

No.	Category	Morphological criteria	Picking decision
No.	Category	Morphological criteria	Picking decision
0	Tender shoot	Plump shoot; leaves not yet unfolded and tightly enclosed	Priority picking
1	Mature shoot	Elongated shoot with unfolded, hardened leaves	Do not pick; leave for further growth
2	One bud with one leaf	One bud accompanied by one newly unfolded leaf	Pick
3	One bud with two leaves	One bud accompanied by two unfolded leaves	Pick

Because tea grade directly affects the subsequent processing quality and purchase price of fresh leaves, mixing tender shoots, one-bud-and-one-leaf shoots, and one-bud-and-two-leaves shoots during harvesting would increase the burden of subsequent sorting. Therefore, this system uses four fine-grained categories. Mature shoots are the only non-picking category included in training, allowing the model to learn not only what should be picked but also what should not be picked.

4.2.2. Dataset Construction and Annotation

The quality of training data largely determines the upper limit of recognition accuracy. The images used in this system come from two sources. Approximately 60% of the samples are taken from publicly available datasets of tea diseases and tea shoots. Most of these images are close-up photographs against plain backgrounds, with clearly defined target contours; they help compensate for the limited sample size of any single collection channel. The remaining approximately 40% of the samples were captured by the researchers using mobile phones in tea plantations. These images cover a range of natural conditions, including front lighting, backlighting, cloudy weather, and mutual occlusion among branches and leaves, thereby improving the model's adaptability to real operating environments. During dataset preparation, all images from both sources were converted to JPG format and renumbered according to the original dataset contents.

All samples were randomly divided into training and validation sets at an approximate ratio of 8:2. Annotation was performed using an annotation tool. Each target was enclosed in a bounding box, and the annotation was saved as a TXT file with the same name as the corresponding image. Each line records the class index, bounding-box center coordinates, width, and height in that order. All coordinates were normalized by the image dimensions to support consistent resizing during training. The annotation statistics are presented in Table 3.

© 2027 The Author(s). Published by Hong Kong Dawn Clarity Press Limited. This is an open access article distributed under the terms of the Creative Commons Attribution 4.0 International Licence (CC BY 4.0), https://creativecommons.org/licenses/by/4.0/.

Table 3. Statistics of Annotated Dataset Bounding Boxes

Class index	Category	Percentage of total
0	Tender shoot	25.8%
1	Mature shoot	14.4%
2	One bud with one leaf	27.0%
3	One bud with two leaves	32.9%
Total	—	100%

4.2.3. Feature Extraction and Model Selection

The lightweight YOLOv8n model from the YOLOv8 family is selected as the detection framework for the following reasons: (1) The framework uses the C2f module as its basic backbone unit, reducing the number of parameters while maintaining gradient flow, and employs the SPPF structure to enlarge the receptive field. (2) The neck uses the bidirectional fusion pathway of PAN-FPN to combine the localization accuracy of shallow features with the semantic information of deeper features. This is particularly important for distinguishing bud-and-leaf combinations with similar contours. (3) The decoupled detection head separates classification and regression into independent branches, reducing conflicts between the feature preferences of these two tasks. (4) The anchor-free design eliminates the need to cluster prior-box sizes, reducing the influence of manual parameter tuning on the results. The model contains approximately 3.1×10^6 parameters, and its inference time per image on the platform used in this study is approximately 4.3 ms, leaving sufficient processing capacity for real-time operation of the complete robot.

4.2.4. Training Strategy and Data Augmentation

Training was performed using PyTorch 2.8.0 and CUDA 12.8 on an RTX 5070 Laptop GPU with 8 GB of video memory. The framework version was Ultralytics 8.4.177. The input resolution was 640 × 640, and the batch size was set to 4. Mixed-precision training was disabled because of the memory constraints, in order to maintain numerical stability. The optimizer was estimated automatically by the framework according to the dataset size and batch size; AdamW was selected in practice. The initial learning rate, determined through automatic estimation, was 0.00125; the momentum was 0.937, and the weight decay was 0.0005. The learning rate was linearly warmed up during the first three epochs and then linearly decayed to 1% of its initial value. Training continued for 100 epochs.

Because the number of samples collected in the field was still limited, a relatively strong combination of data-augmentation techniques was enabled during training. Mosaic augmentation was applied with a probability of 1.0, randomly combining four images into one input image. Random rotation of ±15°, translation of 10%, and scaling of ±50% were also

© 2027 The Author(s). Published by Hong Kong Dawn Clarity Press Limited. This is an open access article distributed under the terms of the Creative Commons Attribution 4.0 International Licence (CC BY 4.0), https://creativecommons.org/licenses/by/4.0/.

applied to expose the model to more combinations of object scales and positions. Horizontal and vertical flips were each applied with a probability of 50%. Since tea shoots in their natural state do not have a fixed upright orientation, vertical flipping does not alter their morphological semantics. Enabling both types of flipping therefore expands the effective sample diversity to approximately four times the original without increasing data-collection costs. In addition, hue, saturation, and brightness in the HSV color space were slightly perturbed to reduce the effects of illumination changes at different times of day. To prevent synthetic images generated by Mosaic from interfering with convergence, Mosaic augmentation was automatically disabled during the final 10 epochs, allowing the model to finish training on images closer to the real data distribution.

4.2.5. Algorithmic Recognition Process

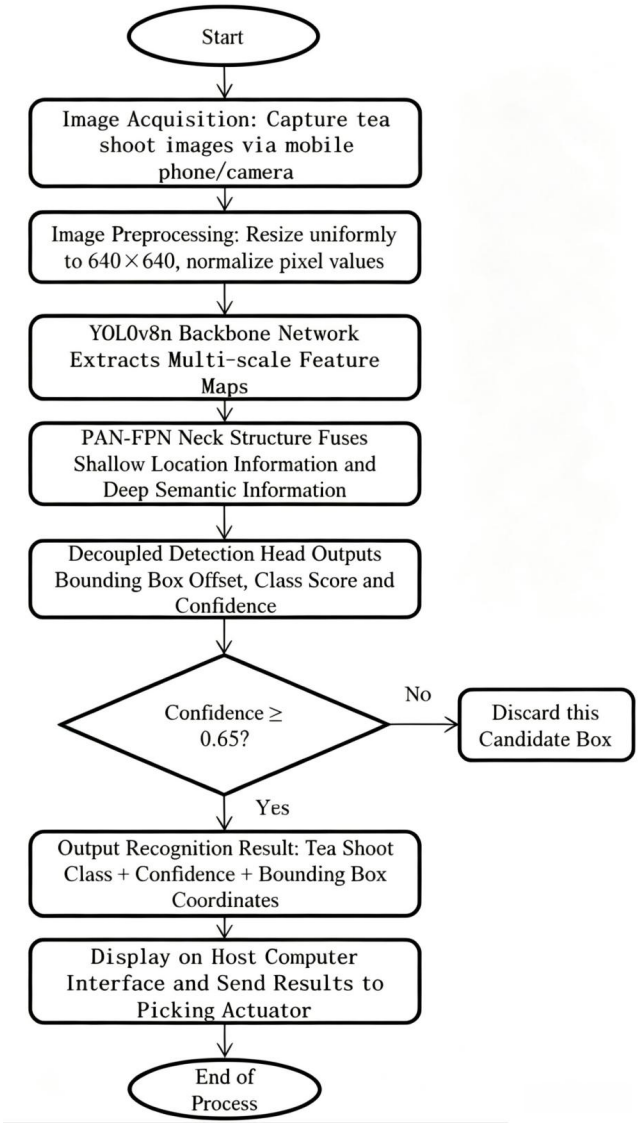


Figure 14. Algorithmic recognition workflow

© 2027 The Author(s). Published by Hong Kong Dawn Clarity Press Limited. This is an open access article distributed under the terms of the Creative Commons Attribution 4.0 International Licence (CC BY 4.0), https://creativecommons.org/licenses/by/4.0/.

The complete tea-shoot recognition process is shown in Figure 14. After an image is captured by a mobile phone or camera, it is uniformly resized to 640 × 640 and its pixel values are normalized to eliminate resolution differences. The image is then fed into the YOLOv8n backbone, where multi-scale convolution extracts feature maps. Bidirectional feature fusion is subsequently performed by the PAN-FPN neck. The decoupled detection head outputs bounding-box offsets, class scores, and confidence values from the fused features. Non-maximum suppression removes overlapping boxes, and candidate targets with confidence values of at least 0.25 are retained. Finally, the system outputs the category, confidence value, and bounding-box coordinates of each tea shoot. The recognition results are displayed in real time on the interface and are also sent by the host computer through a wireless link to the picking actuator, providing the basis for target localization and picking actions.

4.3. Design of the Host Computer

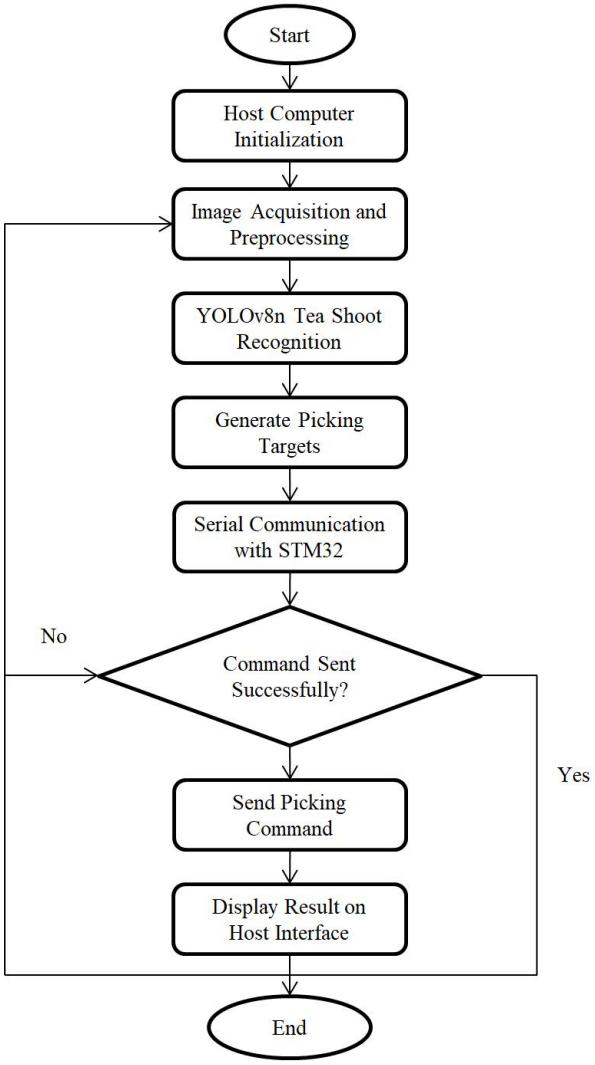


Figure 15. Host-computer design workflow

© 2027 The Author(s). Published by Hong Kong Dawn Clarity Press Limited. This is an open access article distributed under the terms of the Creative Commons Attribution 4.0 International Licence (CC BY 4.0), https://creativecommons.org/licenses/by/4.0/.

The host computer is implemented on a PC or mobile terminal. An HC-05 Bluetooth module provides bidirectional wireless communication with the robot's main controller, while Wi-Fi is used to extend the communication range and improve resistance to interference. The host-computer software consists of three parts: a display interface, a control interface, and a data storage and analysis module. The display interface receives and displays, in real time, tea images captured by the machine-vision module, quality-recognition results, temperature and humidity, illuminance, and the robot's operating status. The control interface sends start and stop commands, sets patrol routes, and adjusts recognition parameters and obstacle-avoidance thresholds. The data storage and analysis module stores and analyzes historical environmental data and recognition results, generates picking recommendations, and can transmit the criteria for selecting high-quality tea shoots to the tea-picking robot. The host computer therefore provides human–machine interaction, remote monitoring, and intelligent decision-making, supporting intelligent management of tea plantations. The programming workflow of the host computer is shown in Figure 15.

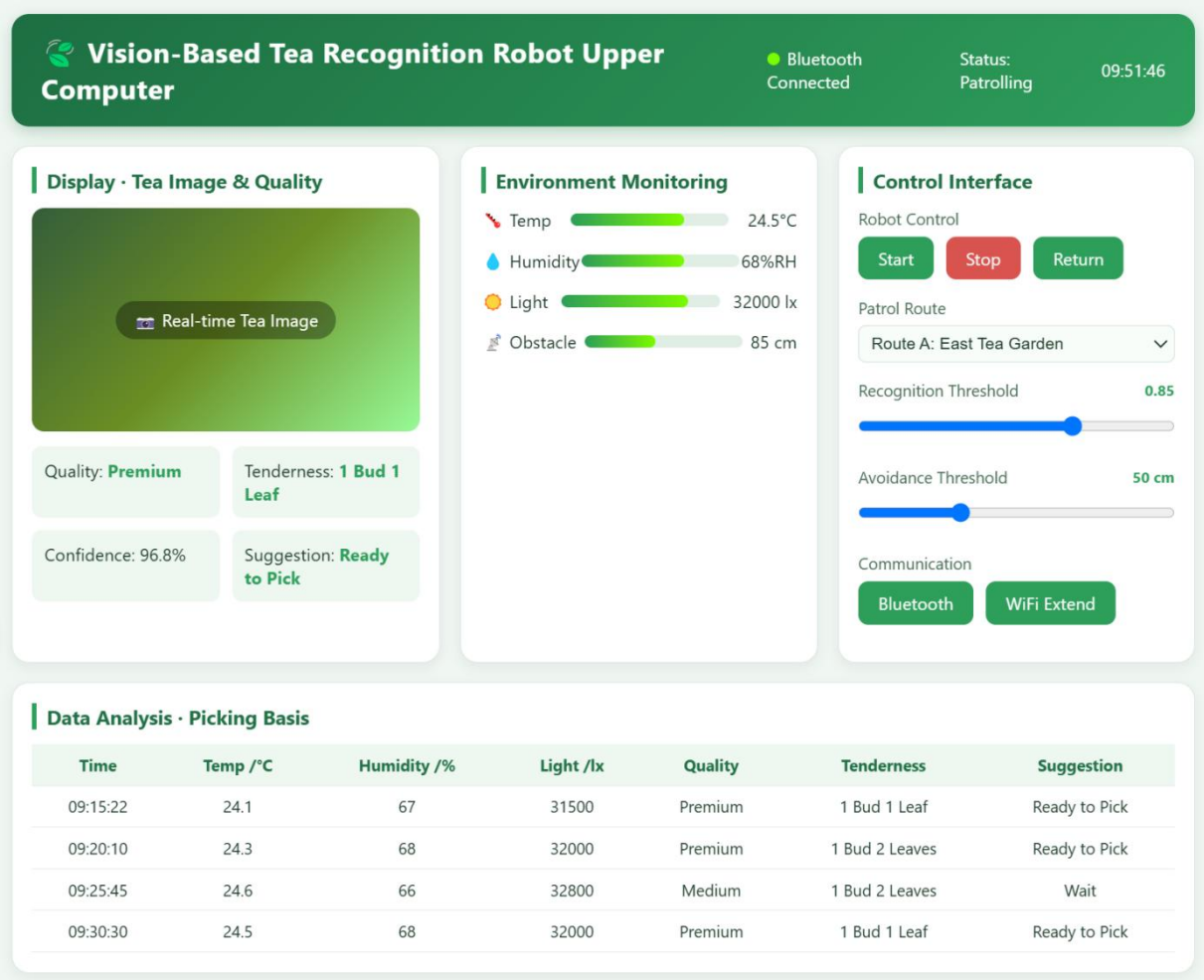


Figure 16. Host-computer user interface

© 2027 The Author(s). Published by Hong Kong Dawn Clarity Press Limited. This is an open access article distributed under the terms of the Creative Commons Attribution 4.0 International Licence (CC BY 4.0), https://creativecommons.org/licenses/by/4.0/.

5. Performance Evaluation of the Tea-Picking Robot

5.1. Recognition Results in Representative Scenarios

To demonstrate the system's recognition performance under different conditions, six representative test images were selected. The results are shown in Figures 17–22. The numbers beside the bounding boxes indicate the confidence values generated by the model.



Figure 17. Recognition of a tender shoot (confidence: 0.82)

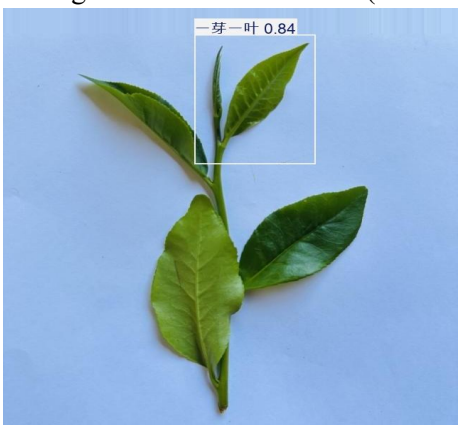


Figure 18. Recognition of one bud with one leaf (confidence: 0.84)



Figure 19. Recognition of one bud with two leaves (confidence: 0.90)

© 2027 The Author(s). Published by Hong Kong Dawn Clarity Press Limited. This is an open access article distributed under the terms of the Creative Commons Attribution 4.0 International Licence (CC BY 4.0), https://creativecommons.org/licenses/by/4.0/.



Figure 20. Recognition of a mature shoot (maximum confidence: 0.74)



Figure 21. Dense arrangement of harvested leaves, with multiple targets detected simultaneously



Figure 22. Field image from a tea plantation, showing recognition results against a complex background

As shown in Figures 17–19, under conditions with a plain background and a single target, the model reliably distinguishes the three categories of pickable shoots. Their confidence values are all above 0.8, and the bounding boxes fit the bud-and-leaf structures well, with no obvious localization errors. Figure 20 shows the recognition result for mature leaves. The model can classify expanded and hardened leaves as mature shoots; however, because the leaves overlap,

© 2027 The Author(s). Published by Hong Kong Dawn Clarity Press Limited. This is an open access article distributed under the terms of the Creative Commons Attribution 4.0 International Licence (CC BY 4.0), https://creativecommons.org/licenses/by/4.0/.

the confidence values for some targets are only around 0.3. Figure 21 shows a dense scene of fresh leaves piled together after picking, in which multiple targets are detected in the same image, indicating that the model has a degree of parallel multi-target detection capability. Figure 22 shows an actual field image from a tea plantation. Although the background contains branches, weeds, and distant buildings, and the illumination is uneven, the model can still detect the main bud-and-leaf targets.

5.2. Recognition Accuracy and Inference Speed

The overall metrics and class-specific performance on the validation set are presented in Table 4.

Table 4. Recognition Performance on the Validation Set

Category	Precision (P)	Recall (R)	mAP50	mAP50–95
Tender shoot	0.861	0.715	0.702	0.441
Mature shoot	0.772	0.626	0.614	0.362
One bud with one leaf	0.837	0.751	0.740	0.483
One bud with two leaves	0.764	0.735	0.728	0.476
All categories	0.808	0.706	0.696	0.441

As shown in Table 4, the model achieves an overall precision of 0.602, recall of 0.420, mAP50 of 0.433, and mAP50–95 of 0.274. Among the individual categories, one bud with two leaves performs best, with an mAP50 of 0.647, followed by one bud with one leaf at 0.557. Tender shoots have the highest precision (0.751), but their recall is only 0.269. This indicates that predictions classified as tender shoots are relatively reliable, although many tender shoots are missed. Mature shoots have the lowest precision and mAP50 among the four categories, at 0.310 and 0.215, respectively.

In terms of inference speed, the median values from repeated measurements show that preprocessing takes approximately 0.8 ms per image, network forward inference takes approximately 4.3 ms, and postprocessing takes approximately 1.0 ms. Including image reading and result rendering, end-to-end processing takes approximately 12.6 ms per image, corresponding to a throughput of about 79 frames per second. This is substantially higher than the typical video acquisition rate of 25–30 frames per second. Therefore, the vision module is unlikely to become a bottleneck in the picking cycle, and the complete robot can operate at the mechanical speed of its actuators without waiting for the recognition algorithm.

Overall, the visual recognition accuracy of the tea-picking robot generally meets the design requirements. Further training with an expanded dataset may improve recognition accuracy. In

© 2027 The Author(s). Published by Hong Kong Dawn Clarity Press Limited. This is an open access article distributed under the terms of the Creative Commons Attribution 4.0 International Licence (CC BY 4.0), https://creativecommons.org/licenses/by/4.0/.

addition, the algorithm's high processing speed should not interfere with the robot's normal operation.

6. Conclusion

To address low picking efficiency and high labor costs in tea plantations, this study developed an intelligent recognition-and-picking robot that integrates STM32 control with machine vision. The robot's mechanical structure and control system were designed according to the operating conditions of tea plantations. The YOLOv8 algorithm was used to identify tea quality, and the resulting confidence values provide a basis for the robot's picking decisions. The system is expected to reduce picking costs and has achieved its intended design objectives.

References

[1] You, W. (2026). China's tea trade: Current status, problems and suggestions. China Tea Processing, (03), 67–72.
[2] Wang, M., Gu, J., & Wang, H. (2023). Method for identifying tea buds based on an improved YOLOv5s model. Transactions of the Chinese Society of Agricultural Engineering, 39(12), 150–157.
[3] Liang, X. (2024). Research on identification, counting, and yield estimation of tea buds in natural environments. South China Agricultural University.
[4] Liao, S., & Zeng, J. (2025). Tea bud recognition based on YOLOv5s. Equipment Manufacturing Technology, (09), 31–34, 52.
[5] Zhang, Y. (2026). Research on recognition and localization methods for tea buds based on binocular vision. Chongqing University of Technology.
[6] Yu, K., Wang, X., & Xu, W. (2024). Design of a modular small-scale lawn-mowing robot. Modern Manufacturing Technology and Equipment, 60(02), 46–50.
[7] Qiu, X. (2025). A sound recognition and key-playing system based on STM32F407ZGT6. Shanxi Electronic Technology, (05), 1–4.
[8] Wei, X., & Tianhao, T. (2026). Electrical motors and drives (4th ed.)). China Machine Press.
[9] Zhang, H., Gu, X., & Zhang, Q. (2025). Design of a microcontroller-based automatic line-tracking and obstacle-avoidance car. Practical Electronics, 33(24), 72–76.
[10] Qi, X., Wang, J., & Yang, W. (2026). Bidirectional intelligent control method of ship steering gear based on sliding-mode active-disturbance rejection. Machine Design and Manufacturing Engineering, 55(10), 50–54.

© 2027 The Author(s). Published by Hong Kong Dawn Clarity Press Limited. This is an open access article distributed under the terms of the Creative Commons Attribution 4.0 International Licence (CC BY 4.0), https://creativecommons.org/licenses/by/4.0/.