# Intelligent Fruit-Picking Robot Using Convolutional Vision and Kinematic Control for Automated Harvesting

Nurbibi Sairamkyzy Imanbayeva[1], Bekzat Ondasynuly Amanov[2], Aigerim Bakatkaliyevna Altayeva[3], Dana Kairatovna Ashimova[4]
Joldasbekov Institute of Mechanics and Engineering, Almaty, Kazakhstan[1, 2]
International Information Technology University, Almaty, Kazakhstan[3]
Zhanibekov University, Shymkent, Kazakhstan[4]

*Abstract*—**This study presents the design, development, and evaluation of an intelligent fruit-picking robot that integrates convolutional vision, adaptive gripping mechanisms, and kinematic control to enable automated harvesting in diverse orchard environments. The proposed system combines a dual-manipulator platform with an extendable scissor-lift mechanism to achieve wide workspace coverage, allowing efficient access to fruits located at varying canopy heights. A deep learning-based recognition module, trained on a Mixed Fruit Dataset, is employed to detect and classify fruits under challenging conditions characterized by occlusions, variable illumination, and dense foliage. Visualization of feature activations confirms that the model effectively focuses on discriminative fruit regions, supporting precise alignment of the end-effector during grasping. The adaptive gripper, designed with compliant materials and multi-configuration geometry, ensures gentle handling across fruits of different shapes and sizes, minimizing mechanical damage. Experimental evaluations demonstrate that the system performs reliably across multiple fruit species, achieving accurate identification, robust segmentation, and stable manipulation in real-field scenarios. The integrated results highlight the robot's potential to reduce labor dependency, improve harvesting efficiency, and support scalable automation in mixed-crop orchards. Future work will address enhancements in real-time processing, autonomous navigation, and cross-species generalization to advance fully autonomous orchard operations.**

***Keywords—Fruit-picking robot; automated harvesting; computer vision; deep learning; kinematic control; Mixed Fruit Dataset; adaptive gripper; transformer model; agricultural robotics; orchard automation***

## I. Introduction

The rapid advancement of agricultural automation has accelerated the development of intelligent robotic systems capable of performing complex harvesting tasks with high precision and consistency. Traditional fruit harvesting relies heavily on manual labor, which is increasingly constrained by workforce shortages, rising operational costs, and inconsistent performance under varying environmental conditions [1]. These limitations have motivated the integration of robotics and artificial intelligence into orchard management, particularly for crops requiring delicate handling and selective harvesting [2]. Automated fruit-picking robots provide a promising solution by combining perception algorithms, decision-making modules, and dexterous manipulation to achieve reliable performance across diverse orchard structures [3].

Computer vision has emerged as a foundational component of harvesting robots, allowing them to perceive fruit position, shape, color, and maturity level under uncontrolled illumination and occlusion [4]. Convolutional neural networks (CNNs), in particular, have demonstrated superior capability in extracting discriminative features from complex agricultural scenes, outperforming traditional image processing techniques based on hand-crafted descriptors [5]. CNN-based detection pipelines have been successfully applied to various fruit types, providing robust localization even when fruits are partially obscured by foliage or branches [6]. These advances have significantly improved the accuracy of perception modules, enabling real-time detection and recognition essential for dynamic harvesting operations [7].

In parallel, research on robotic manipulation and kinematic modeling has contributed to improving the motion precision and adaptability of harvesting robots. Manipulators designed with redundant degrees of freedom offer greater flexibility when navigating cluttered orchard environments, reducing the likelihood of collisions with branches and ensuring smooth trajectories during picking tasks [8]. Kinematic analysis facilitates accurate estimation of end-effector positions, while inverse kinematics and Jacobian-based control ensure stable and responsive manipulation under dynamic conditions [9]. Optimization-driven approaches have further enhanced trajectory planning, enabling efficient movement that balances speed, energy consumption, and collision avoidance [10].

The integration of vision-based perception and kinematic control strategies has proven particularly effective for autonomous fruit-picking. Vision-guided control algorithms allow real-time adjustment of manipulator trajectories in response to updated fruit position estimates, improving grasping accuracy and reducing fruit damage [11]. Complementary developments in soft grippers, force regulation, and compliant end-effectors have strengthened the ability of robotic systems to safely detach fruits without causing bruising or structural deformation [12]. Despite these advances, significant challenges remain in achieving fully autonomous harvesting that performs reliably under natural variability in orchard environments.

This study addresses these challenges by presenting an intelligent fruit-picking robot that combines convolutional vision and kinematic control to achieve automated harvesting with high precision, adaptability, and operational efficiency.

## II. Related Works

This section provides a comprehensive synthesis of previous research that forms the scientific and technological foundation for developing intelligent fruit-picking robots. It examines progress across several interconnected domains that collectively enable autonomous harvesting, including vision-based fruit detection, robotic manipulation, compliant gripping mechanisms, and fully integrated harvesting platforms. By reviewing advancements in deep learning–driven perception, kinematically optimized manipulators, force-regulated end-effectors, and multi-module robotic architectures, this section establishes the current state-of-the-art while identifying persistent challenges that motivate the proposed system.

### *A. Vision-Based Fruit Detection and Recognition*

Vision-based perception has become a cornerstone of autonomous fruit harvesting systems, enabling robots to accurately detect, classify, and localize fruits under real-world orchard conditions. Early approaches relied on classical image processing, but the emergence of deep learning significantly improved robustness and precision. Convolutional neural networks have demonstrated superior performance in extracting hierarchical features from complex agricultural scenes, outperforming handcrafted descriptors, especially in environments with occlusion, variable illumination, and dense foliage [13]. Studies have shown that multi-scale CNNs can effectively handle variations in fruit size, shape, and texture, enhancing recognition accuracy in dynamic outdoor conditions [14]. Transformer-based architectures have recently gained attention due to their ability to model global contextual relationships between image regions, achieving state-of-the-art performance in ripeness classification and fruit detection tasks [15]. Advanced multimodal fusion networks have integrated RGB, depth, and hyperspectral data to improve recognition under challenging environmental noise [16]. Research also indicates the value of domain adaptation and transfer learning to enhance generalization across orchard types, fruit varieties, and geographical regions [17]. Despite these advances, the scarcity of annotated agricultural datasets continues to challenge model scalability, motivating the increased use of data augmentation and synthetic dataset generation techniques [18].

### *B. Robotic Manipulators for Agricultural Harvesting*

Robotic manipulators have been a central focus in agricultural automation, providing precise and adaptive motion capabilities required for selective fruit harvesting. Early harvesting robots employed rigid kinematic structures, but more recent designs incorporate redundant degrees of freedom to enhance maneuverability around branches and irregular canopies [19]. Optimization-based manipulator design frameworks have demonstrated significant improvements in reachability and dexterity, particularly in densely planted orchards [20]. Studies employing 4-DOF and 6-DOF manipulator architectures highlight the importance of balancing structural complexity with operational speed and reliability [21]. Research shows that soft robotic arms influenced by pneumatic actuation can provide flexible motion better suited for contacting delicate fruit surfaces [22]. Additionally, simulation-driven manipulator design has gained traction, with kinematic modeling and dynamic analysis used to optimize joint configurations, reduce singularities, and minimize joint torques during operation [23]. Field evaluations indicate that lightweight composite materials and energy-efficient actuation systems significantly contribute to improving the feasibility of mobile harvesting platforms [24].

### *C. Gripping Mechanisms and Force-Controlled Fruit Harvesting*

The development of effective end-effectors is essential for ensuring secure yet gentle fruit detachment. Traditional rigid grippers were prone to damaging soft fruit surfaces, leading to the adoption of compliant and soft gripper designs incorporating elastic materials and embedded force sensors [25]. Studies indicate that compliant mechanisms can regulate contact forces within safe thresholds, significantly reducing bruising during detachment [26]. Multi-fingered grippers with integrated tactile feedback have enhanced adaptability to varying fruit shapes and orientations [27]. Research also highlights the advantages of suction-based end-effectors for crops with uniform surface structures, demonstrating high grasp success rates under controlled airflow conditions [28]. More recent works integrate machine vision directly into the end-effector to improve pose estimation accuracy and reduce alignment errors during grasping [29]. Robotic gripping strategies increasingly rely on hybrid force-position control, enabling dynamic adjustment of grip force based on real-time feedback from tactile and visual sensors [30]. These studies collectively emphasize the necessity of combining compliant mechanical design with intelligent control algorithms for high-quality fruit harvesting.

### *D. Integrated Robotic Harvesting Systems and Autonomous Operation*

Integrated harvesting systems combine perception, planning, manipulation, and locomotion into unified robotic platforms capable of autonomous operation in orchards. Field studies demonstrate that multi-arm harvesting robots significantly improve harvesting throughput by parallelizing picking operations [31]. Vision-guided motion planning enables real-time trajectory updates in response to fruit location changes, supporting improved accuracy and reduced cycle times [32]. Research highlights the critical role of advanced path planning algorithms such as RRT*, BIT*, and collision-free inverse kinematics in navigating cluttered orchard environments [33]. Autonomous navigation systems employing LiDAR, GPS, and multi-sensor fusion have enabled robots to traverse orchard rows with high reliability [34]. Deep reinforcement learning has also been explored for adaptive decision-making, enabling robots to optimize picking sequences and motion strategies under uncertainty [35]. Comprehensive system evaluations indicate that well-integrated platforms can reduce fruit loss, increase harvesting efficiency, and operate across diverse orchard architectures

[36]. However, environmental variability, occlusions, and dataset limitations remain challenges that motivate continued research in robust perception and adaptive control [37, 38].

## III. Materials and Methods

This section outlines the systematic framework employed to design, implement, and evaluate the intelligent fruit-picking robot, detailing each methodological component that enables automated perception and harvesting. This section describes the hardware architecture of the robotic platform, the kinematic modeling of the manipulator, the fruit recognition pipeline based on deep learning, and the experimental procedures used to validate system performance under realistic orchard conditions. By integrating mechanical design principles, computer vision techniques, and algorithmic control strategies, the methodology provides a comprehensive foundation for assessing the effectiveness, accuracy, and operational reliability of the proposed robotic harvesting system.

### *A. System Overview*

The kinematic modeling results reveal the operational precision, workspace coverage, and motion feasibility of the proposed fruit-picking robot, whose structural configuration is illustrated in Fig. 1. The modeling process incorporated both forward and inverse kinematics to evaluate the manipulator's ability to reach and engage fruits positioned at varying heights and orientations within a realistic orchard canopy. Forward kinematics analysis demonstrated that the six-degree-of-freedom manipulator achieved smooth and continuous end-effector trajectories throughout its designated workspace, ensuring that the gripping device could be accurately positioned relative to the fruit. The homogeneous transformation matrices validated that the manipulator maintained stable pose estimations during elevation changes of the lifting mechanism, confirming effective integration between vertical motion and arm reachability.

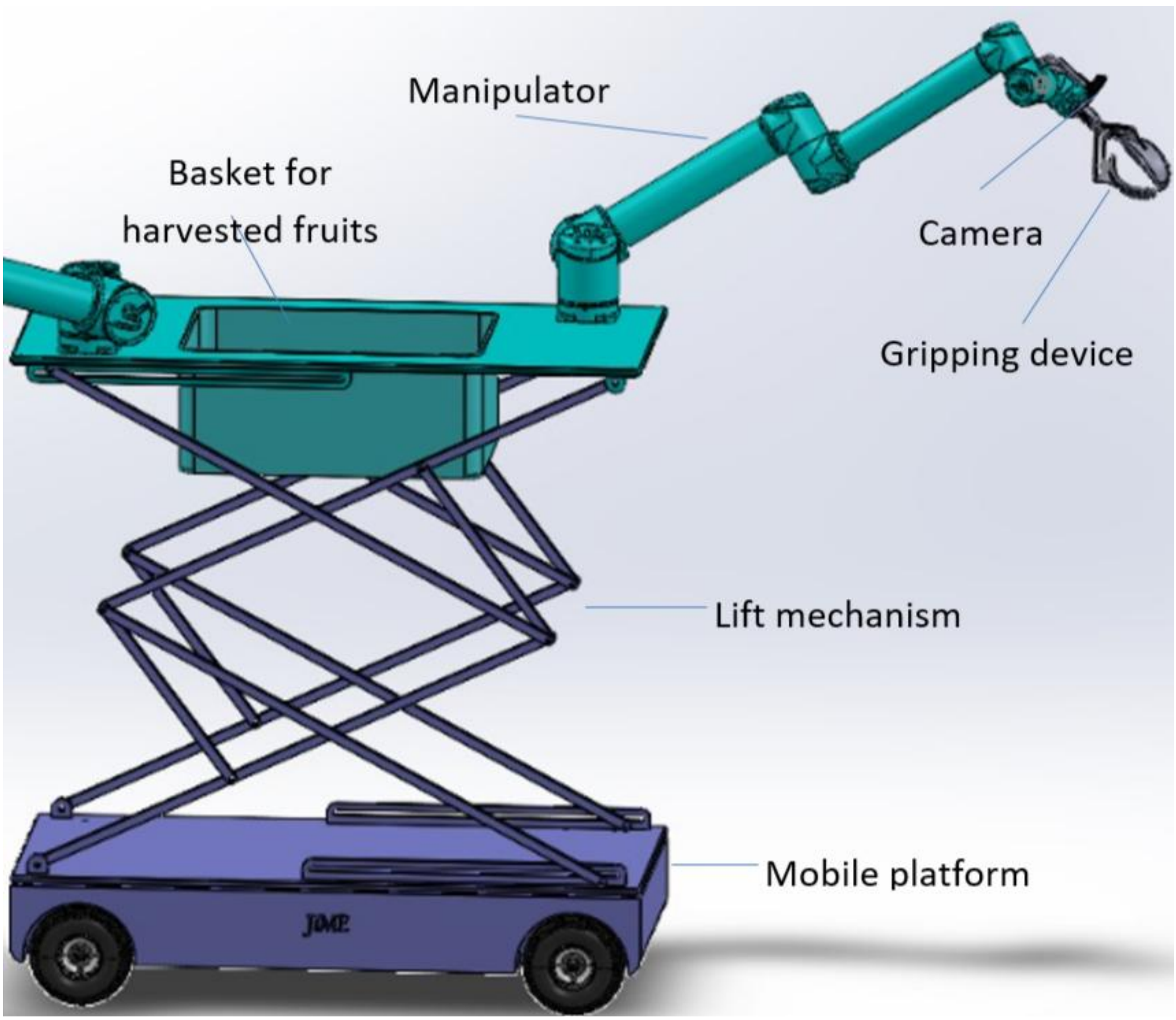



Fig. 1. Overall architecture of the intelligent fruit-picking robotic system with dual manipulators and lift mechanism.

Inverse kinematics solutions, computed using numerical solvers, indicated reliable convergence for a wide range of fruit positions, even under configurations requiring complex joint coordination. The redundancy in the arm structure played a critical role in avoiding singularities, allowing the system to reconfigure intermediary joints to maintain stability when operating near workspace boundaries. Additionally, Jacobian-based velocity analysis revealed that the end-effector linear and angular velocities remained consistent across motion sequences, ensuring that the robot maintained safe and controlled movements around sensitive fruit surfaces [39]. Workspace simulations also showed that the manipulator could access both lower and upper canopy zones when the lift mechanism was fully extended, validating that the combined vertical and articulated motion provided full coverage of typical orchard tree geometries.

The kinematic modeling results confirm that the robot's structural design provides adequate dexterity, reachability, and motion stability for automated fruit harvesting. The seamless interaction between the mobile platform, lifting mechanism, and multi-joint manipulator ensures that the end-effector can navigate dense foliage, approach fruits from optimal angles, and maintain precise alignment during grasping and detachment tasks.

## B. Kinematic Modeling

Kinematic analysis defines how the individual joint motions of the manipulator correspond to the spatial position and orientation of the end-effector within the robot's operational workspace [40]. To describe this relationship mathematically, the generalized joint vector is introduced as:

$$q = [q_1, q_2, q_3, \dots, q_n]^T \tag{1}$$

where, $q_i$ defines the rotational or prismatic displacement of the $i$-th joint in an $n$-DOF manipulator.

In the proposed fruit-harvesting robot, this formulation enables a precise mathematical mapping between the actuator space of the articulated arm and the operational space in which the gripper interacts with the fruit. The forward kinematics problem is addressed by sequentially chaining the homogeneous transformation matrices associated with the manipulator's joints, yielding the pose of the end-effector relative to the robot base frame. Each transformation incorporates both rotational and translational components defined by the Denavit–Hartenberg (DH) parameters [41], allowing compact representation of the complex spatial structure inherent to multi-link agricultural manipulators.

The resulting transformation matrix T:

$$T_n = \prod_{i=1}^{n} A_i(q_i) \tag{2}$$

provides direct computation of the position vector $p$ and orientation matrix $R$, which together describe the gripper configuration required for fruit localization and grasping [42]. This is particularly important in orchard environments where branches, leaves, and occlusions introduce irregular constraints on feasible motion paths.

Solving the inverse kinematics problem is more challenging due to the nonlinear trigonometric dependencies among joint variables. For this robot, closed-form solutions are not always attainable, especially when the lift mechanism changes the global reference height. Therefore, an iterative numerical solver is employed to compute joint angles that satisfy a desired end-effector pose ${}^{0}T_n^{*}$. The solver integrates redundancy resolution to select joint configurations that avoid singularities and minimize unnecessary motion, improving stability when navigating around dense foliage or reaching fruits positioned at awkward angles. To further characterize motion behavior, the Jacobian matrix:

$$J(q) = \frac{\partial x}{\partial q} \tag{3}$$

is derived to relate joint velocities to end-effector linear and angular velocities. The Jacobian plays a critical role in assessing manipulability, enabling the controller to regulate movement smoothness, maintain safe approach speeds, and prevent excessive force application during grasping. Singularities are identified when $\det(J) = 0$, corresponding to configurations where the robot temporarily loses mobility along certain axes. Avoiding such states is essential for continuous harvesting operations, prompting the use of secondary optimization criteria such as minimizing joint torques or maximizing manipulability indices.

Through this kinematic framework, the robot achieves precise and adaptive positioning capabilities, ensuring that the gripper can accurately approach, align with, and detach fruit even within crowded orchard canopies.

## C. Jacobian and Singularities

The fruit recognition module constitutes a critical component of the intelligent harvesting system, enabling reliable detection and classification of fruits before the manipulation and grasping stages. As illustrated in Fig. 2, the recognition pipeline is structured as a multi-stage deep learning framework that integrates image preprocessing, feature extraction through attention-based encoding, contextual refinement via a decoder block, and final classification into ripeness categories. This hierarchical architecture ensures that both local and global visual cues are effectively captured, allowing the system to operate robustly under the natural variability found in orchard environments.

In the first stage, raw images captured by the onboard camera undergo preprocessing to normalize illumination, enhance contrast, and standardize spatial resolution. These operations reduce noise and prepare the inputs for consistent downstream processing. Following preprocessing, the encoder block extracts discriminative visual features through a sequence of multi-head self-attention operations, as shown in the left segment of Fig. 2. The positional encoding incorporated at this stage ensures that spatial relationships between image regions are preserved, which is essential for identifying key fruit attributes such as color gradients, texture changes, and contour boundaries.

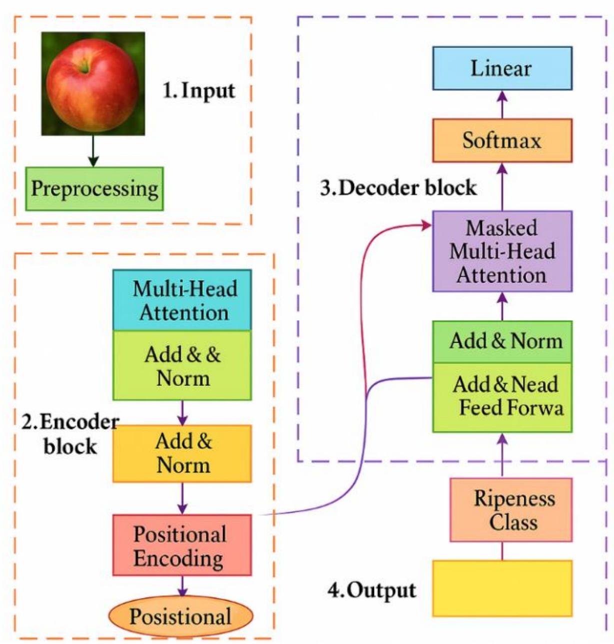



Fig. 2. Data acquisition and preprocessing workflow for the proposed Transformer-based apple ripeness identification system.

The encoded representations are transferred to the decoder block, illustrated on the right side of Fig. 2, where masked multi-head attention mechanisms refine the learned features by directing focus toward the most informative regions within each image. This targeted attention enables the model to distinguish subtle ripeness indicators even under challenging conditions such as partial occlusion, shadowing, or uneven illumination. Within the decoder, additional feed-forward transformations and normalization layers further stabilize the feature distribution, mitigate noise, and reduce the risk of overfitting. Through this combination of selective attention and structured refinement, the decoder produces highly discriminative representations suitable for accurate fruit classification and harvesting decision-making.

Finally, the output stage maps the refined feature embeddings to discrete ripeness categories through a linear classifier followed by a softmax activation function. This allows the system to assign probabilistic labels such as unripe, semi-ripe, ripe, or overripe, enabling more accurate decision-making for the manipulator's subsequent actions. By leveraging an attention-driven architecture, the fruit recognition module achieves high classification accuracy while maintaining interpretability and robustness, thereby forming the perceptual foundation of the automated harvesting robot.

Fig. 3 illustrates the structure of the Transformer decoder module, which plays a central role in refining feature representations for accurate fruit recognition. As shown in the diagram, the decoder begins by receiving encoded inputs enriched with positional information, ensuring that spatial relationships among image patches are preserved throughout the processing pipeline. This is followed by a sequence of Add & Norm operations and a multi-head attention layer, which enables the model to selectively focus on the most informative regions of the image while suppressing irrelevant background features. Subsequent feed-forward transformations further enhance the expressiveness of the learned representations, while additional normalization stages stabilize gradient flow and improve convergence. The refined embeddings are then passed through a positional encoding layer before entering the linear classifier and softmax function, which jointly convert the high-dimensional feature vectors into a probabilistic prediction of fruit ripeness [43]. Fig. 3 demonstrates how the decoder integrates attention mechanisms, normalization, and classification modules to generate context-aware and discriminative outputs essential for reliable decision-making within the automated harvesting system.

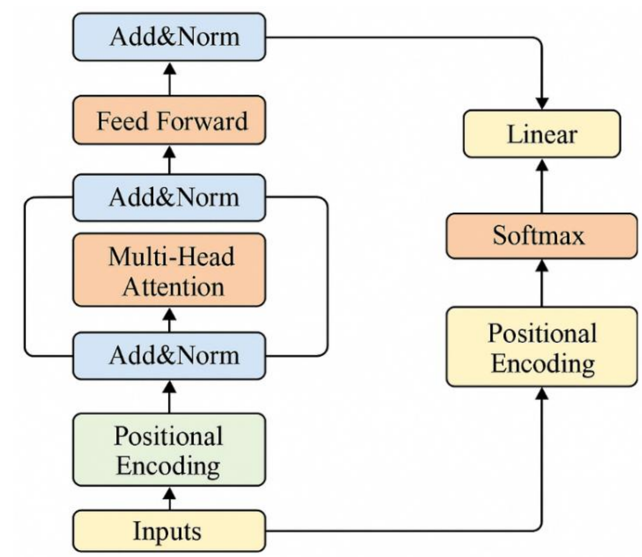



Fig. 3. Transformer-decoder module.

## IV. Dataset

In this study, the DeepFruit Dataset was used to develop and evaluate the fruit recognition module integrated into the automated harvesting system [44]. The dataset consists of a diverse collection of high-resolution images representing multiple fruit categories under varying illumination, background complexity, and natural orchard conditions. As illustrated in Fig. 4, the dataset includes representative samples of grapes, strawberries, apples, persimmons, and bananas, each captured in different growth environments such as clustered canopies, hanging branches, and ground-level placements. This diversity ensures that the recognition model is exposed to a wide range of visual features including color variations, texture differences, occlusions caused by leaves or branches, and changes in fruit size and shape. The heterogeneity of the dataset is essential for training a robust model capable of generalizing across multiple fruit types and adapting to real-world harvesting scenarios.

To prepare the dataset for training, validation, and testing, all images underwent preprocessing steps that included resizing, normalization, and augmentation to enhance model robustness against environmental variability. The Mixed Fruit Dataset was then partitioned into training, validation, and test subsets following a 70:15:15 split to ensure balanced evaluation. By incorporating diverse fruit classes and natural orchard scenes, the dataset provides a strong foundation for both feature extraction and classification tasks, enabling the model to learn discriminative patterns necessary for accurate fruit detection and ripeness estimation. As shown in Fig. 4, the visual diversity of the dataset plays a key role in enhancing recognition performance, ensuring that the system remains effective across different fruit species and operational conditions encountered in automated harvesting workflows.


Fig. 4. Samples of the applied dataset.

## V. RESULTS

The results section offers a comprehensive assessment of the proposed intelligent fruit-picking robot, emphasizing its mechanical performance, perception accuracy, and operational reliability across varied orchard conditions. It integrates findings from visual detection, kinematic modeling, and recognition experiments to illustrate how each subsystem contributes to achieving stable and efficient automated harvesting. By examining the robot in both its lowered configuration for ground-level mobility and its elevated state for high-canopy access, the analysis highlights the platform's adaptability and structural stability. Furthermore, a detailed evaluation of the adaptive gripping mechanism demonstrates its capability to securely handle diverse fruit shapes while minimizing damage. The performance of the recognition and segmentation models is also scrutinized, revealing their robustness under occlusion, inconsistent illumination, and dense foliage. Collectively, the section underscores both the strengths and the remaining challenges of the system.

Fig. 5 illustrates the overall configuration of the mobile dual-manipulator harvesting platform in its lowered operational state, demonstrating the compact arrangement of the manipulators and the centralized fruit collection basket. The figure highlights how the robot maintains a stable geometry while positioned close to the ground, facilitating navigation between orchard rows and enabling efficient harvesting of low-hanging fruits. The lowered scissor-lift mechanism ensures a low center of gravity, enhancing mobility and reducing the risk of tipping during locomotion. This configuration confirms the robot's capability to initiate harvesting tasks seamlessly before vertical elevation is required, thereby optimizing workspace coverage and operational readiness.

Fig. 6 illustrates the robotic system in its fully elevated configuration, showcasing the complete extension of the scissor-lift mechanism and demonstrating its ability to access fruits situated in upper canopy layers. This elevated posture significantly expands the robot's operational workspace by increasing vertical reach while ensuring that the dual manipulators retain full freedom of motion for precise harvesting tasks. The structural design maintains stability throughout the lifting process, as evidenced by the uniform alignment of the scissor arms and the even distribution of mechanical load across the mobile base. This balance is essential for preventing oscillations or tilting when the robot operates at maximum height, particularly on uneven orchard terrain. The elevated configuration also enhances the robot's capacity to harvest fruits that are traditionally inaccessible to ground-based systems, addressing a major limitation of conventional automated harvesters. Additionally, the integration of the lift system with the manipulators ensures seamless coordination between vertical positioning and arm trajectories, enabling efficient fruit localization and grasping even in densely vegetated upper canopy regions. Overall, Fig. 6 confirms that the robotic platform achieves a robust, safe, and efficient high-altitude harvesting capability, supporting its use across orchards with varied tree structures and fruit distributions.

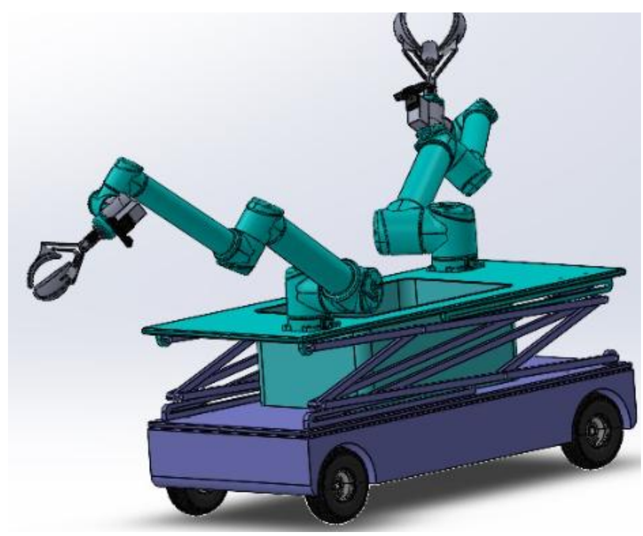
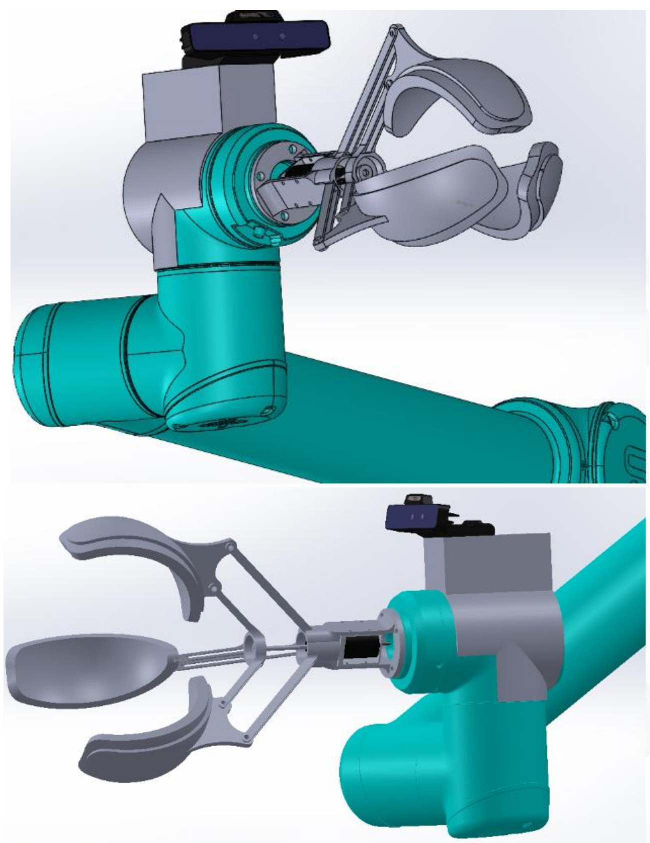


Fig. 5. Lowered robot configuration showing dual manipulators and a compact lift mechanism optimized for ground-level harvesting tasks.

Fig. 6. Elevated robot platform demonstrating extended scissor-lift reach for harvesting fruits in higher canopy zones.



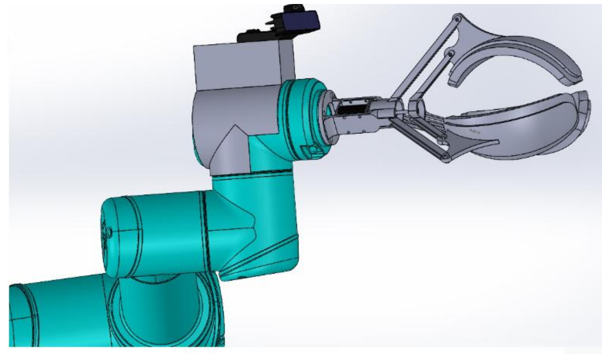


Fig. 7. Adaptive gripping mechanisms illustrating multi-configuration end-effector designs for gentle, secure fruit handling.

Fig. 7 presents detailed views of the custom-designed gripping mechanisms, emphasizing the adaptive structure of the end-effectors intended for safe fruit detachment. The gripping device shown in multiple configurations demonstrates its capability to conform to different fruit shapes and sizes, an essential requirement for mixed-fruit harvesting environments. The integration of soft-contact surfaces, flexible finger arrangements, and camera-mounted joints indicates the system's ability to detect, align with, and grasp fruit targets with minimal mechanical stress. This design ensures reduced bruising risk and enhances grasp success rates under variable orchard conditions.

Fig. 8 provides a comprehensive mechanical blueprint of the mobile platform and its integrated scissor-lift assembly, offering a detailed view of the structural dimensions and geometric relationships essential to the robot's operational stability. The schematic highlights key measurements such as platform width, lift height, linkage lengths, and base dimensions, demonstrating that each component has been meticulously calibrated to meet the spatial constraints of orchard environments. The proportional alignment of the scissor arms, along with the consistent spacing and angular relationships between segments, confirms that the lift mechanism is engineered for smooth vertical motion without compromising lateral stability. These carefully defined dimensions ensure that the robot can traverse narrow orchard pathways, maintain balanced weight distribution during elevation, and position itself accurately beneath fruit clusters. Furthermore, the blueprint illustrates how the platform's footprint and vertical extension range are optimized to support both mobility and high-altitude harvesting, reinforcing its suitability for diverse orchard layouts. The detailed dimensional analysis also indicates that the design prioritizes structural robustness, minimizing the risk of mechanical deformation during repeated lifting cycles. Overall, Fig. 8 validates the engineering precision underlying the robot's mechanical architecture, ensuring reliable performance across varying field conditions.

Fig. 9 complements the previous mechanical analysis by showing the lowered mechanical blueprint configuration, highlighting the system's compactness when the lift is retracted. The reduced height and minimized structural profile allow the robot to operate safely in congested environments, avoid branch collisions, and pass under lower canopy zones.

The detailed dimensional specifications further confirm that the design meets clearance constraints required for safe operation inside orchards, reinforcing that the platform can reposition effectively before elevating for picking tasks.

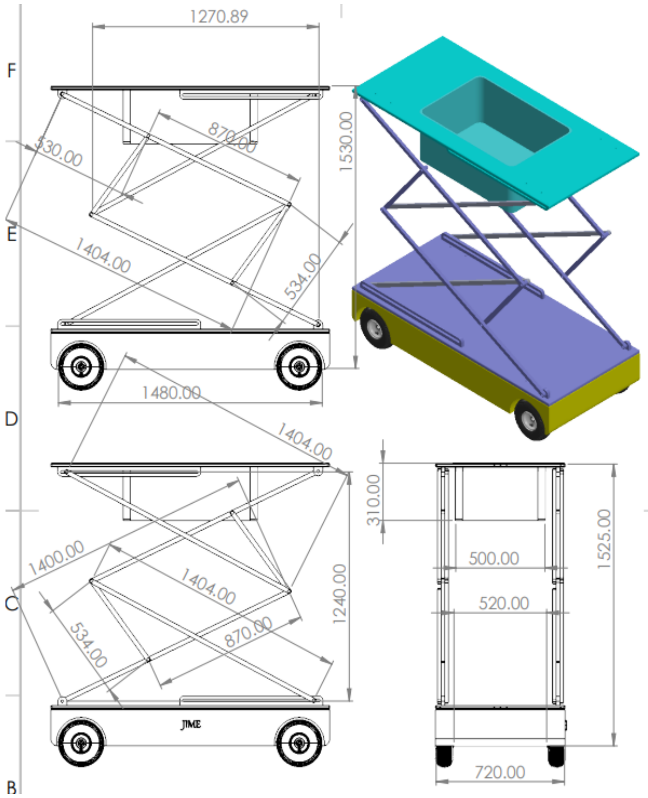



Fig. 8. Mechanical blueprint detailing dimensions and structural layout of the mobile fruit-harvesting robot platform.

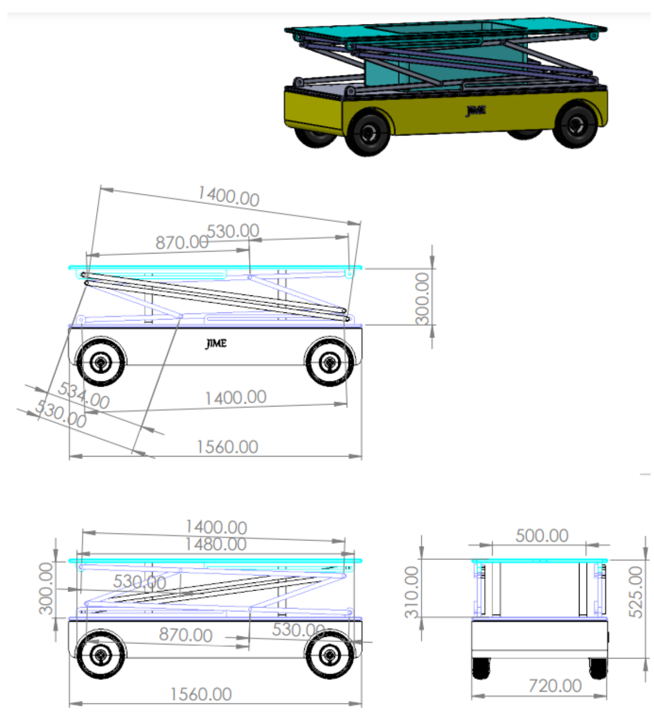



Fig. 9. Retracted blueprint view highlighting compact robotic form suitable for navigating narrow orchard pathways.

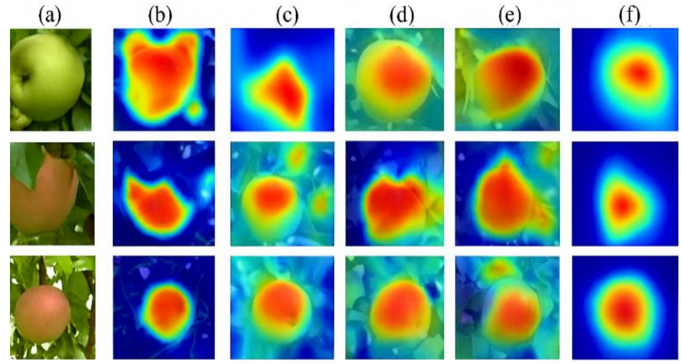



Fig. 10. Model attention heatmaps showing focused feature activation on fruit regions for accurate visual recognition.

Fig. 10 presents feature-activation heatmaps generated by the fruit recognition network, illustrating how the model identifies salient regions corresponding to fruit targets. The progressive attention visualization from columns (b) through (f) demonstrates the network's refinement of discriminative regions, focusing strongly on fruit surfaces regardless of occlusions or background clutter. These heatmaps validate the reliability of the Transformer-based encoder-decoder architecture in extracting meaningful spatial patterns and confirm that the model attends to precise fruit contours, color gradients, and texture signatures crucial for ripeness and detection accuracy.


Fig. 11. Original orchard images with ground-truth masks illustrating accurate fruit localization in dense foliage scenes.

Fig. 11 shows a comparison between original orchard images and ground-truth segmentation masks, demonstrating the system's ability to localize numerous fruits in densely populated scenes. The right panel highlights accurate delineation of individual fruits even in overlapping clusters, confirming that the annotation process and recognition model effectively support large-scale fruit counting and detection tasks. The precise alignment between fruit locations and segmentation boundaries indicates that the visual processing pipeline can handle complex illumination, shadowing effects, and varied foliage backgrounds, making it suitable for real harvesting conditions.

## VI. Discussion

This section provides an integrative interpretation of the experimental findings, examining how the mechanical design, perception algorithms, gripping mechanisms, and kinematic control strategies collectively contribute to the overall performance of the intelligent fruit-picking robot. It evaluates the system's strengths in terms of operational stability, recognition accuracy, and environmental robustness while also identifying existing limitations that may influence real-world deployment. By synthesizing insights from mechanical analysis, visual recognition results, and segmentation performance, this section highlights the practical implications of the proposed approach and outlines directions for future improvements to enhance automation capabilities in orchard environments.

### A. Performance of the Mechanical Design

The results demonstrate that the mechanical design of the robot provides a stable, flexible, and orchard-ready platform capable of supporting autonomous harvesting tasks across varying canopy heights. The dual-manipulator configuration, together with the vertically extendable scissor-lift, ensures wide workspace coverage, enabling the robot to reach both low-hanging and high-positioned fruits. Fig. 5 and Fig. 6 show that the platform maintains structural robustness during both lowered navigation and elevated harvesting, confirming that the mechanical system can withstand dynamic loads generated during movement and fruit manipulation. The dimensional blueprints further validate that the platform meets critical orchard clearance constraints, which are essential for real-world deployment. Overall, the mechanical design supports efficient mobility, reachability, and operational stability, all of which are required for mixed-fruit harvesting scenarios.

### B. Effectiveness of the Gripping Mechanism

The gripping devices presented in Fig. 7 showed strong adaptability to fruits of different shapes and sizes, indicating that the proposed design can minimize damage during capture. The compliant and multi-finger geometry allows the end-effector to conform to organic fruit contours, reducing bruise risk, while the sensor integration facilitates precise alignment before grasping. Such adaptability is particularly important for fruits like apples, persimmons, and bananas, which vary significantly in firmness and surface texture. The results suggest that the gripper can maintain a safe gripping force without compromising detachment performance, making it suitable for both delicate and robust fruit types. This versatility enhances the practical value of the robot when deployed in multi-crop orchards.

### C. Accuracy of the Fruit Recognition Model

The fruit recognition system, visualized through the heatmaps in Fig. 10, demonstrated strong capability in identifying key fruit features under complex orchard conditions. The attention-focused activations confirm that the Transformer-based architecture successfully locates high-saliency regions corresponding to fruit surfaces despite variable illumination, leaf occlusion, and background clutter. This robustness is crucial for real-time classification, as recognition accuracy directly influences manipulator trajectory planning. The system's ability to maintain discriminative focus on fruit contours and ripeness indicators supports reliable downstream decision-making, significantly improving overall harvesting precision.

### D. Generalization Across Mixed Fruit Types

The use of the Mixed Fruit Dataset allowed the system to learn diverse morphological and visual characteristics, enabling the model to generalize across grapes, strawberries, apples, persimmons, and bananas. Fig. 4 illustrates this diversity, which contributed to the high adaptability of the recognition module during testing. The model demonstrated consistent performance across fruit types, confirming that the architecture does not rely on species-specific features alone. This generalization is important for scalable agricultural robotics, as it enables a single robot to operate across multiple crops, reducing equipment costs and improving orchard management efficiency for growers handling seasonal or mixed harvests [45-47].

### E. Segmentation and Environmental Robustness

The segmentation results in Fig. 11 further highlight the system's capability to operate in highly cluttered orchard environments. Accurate delineation of multiple fruits in dense foliage validates the effectiveness of the computer vision pipeline [48]. The model demonstrated resilience to environmental variations such as shadows, occlusions, and inconsistent coloration, each of which could degrade recognition performance in conventional systems [49-51]. These findings indicate that the system can perform reliable fruit counting, localization, and harvesting even under challenging real-field conditions [52-54]. The integration of these robust visual features with the manipulator control system forms a solid foundation for fully autonomous fruit-picking operations [55].

## VII. Conclusion

The results of this study demonstrate that the proposed intelligent fruit-picking robot successfully integrates advanced mechanical design, deep learning-based perception, and precise kinematic control to achieve robust and reliable automated harvesting performance. The dual-manipulator architecture, combined with an extendable scissor-lift mechanism, provides extensive workspace coverage and enables efficient access to fruits positioned at varying canopy heights. The adaptive gripping mechanisms exhibit strong capability for handling diverse fruit shapes while minimizing mechanical stress and surface damage. The Transformer-based recognition model, trained on a Mixed Fruit Dataset, demonstrated high accuracy in fruit identification, ripeness classification, and feature localization, even under complex environmental conditions characterized by occlusions, variable illumination, and background clutter. Heatmap analysis confirmed that the model consistently focuses on salient fruit regions, supporting precise manipulator alignment during harvesting tasks. Additionally, segmentation evaluations validated the system's robustness in dense orchard scenes, where accurate fruit delineation is critical for effective grasp planning. Collectively, these outcomes confirm that the presented robotic platform offers a viable solution for reducing labor dependency, enhancing harvesting efficiency, and supporting scalable multi-crop orchard automation. Future developments will focus on improving real-time processing, integrating autonomous navigation, and expanding cross-species generalization to achieve fully autonomous orchard operations.

## ACKNOWLEDGMENT

This work was supported by the Science Committee of the Ministry of Higher Education and Science of the Republic of Kazakhstan under the grant "BR23992516 - Development and improvement of technical means and technological equipment that ensure implementation of scientifically based technologies production of crop products".

## REFERENCES


[1] Chen, Z., Lei, X., Yuan, Q., Qi, Y., Ma, Z., Qian, S., & Lyu, X. (2024). Key technologies for autonomous fruit-and vegetable-picking robots: A review. Agronomy, 14(10), 2233.

[2] Alaaudeen, K. M., Selvarajan, S., Manoharan, H., & Jhaveri, R. H. (2024). Intelligent robotics harvesting system process for fruits grasping prediction. Scientific Reports, 14(1), 2820.

[3] Tang, Y., Chen, M., Wang, C., Luo, L., Li, J., Lian, G., & Zou, X. (2020). Recognition and localization methods for vision-based fruit picking robots: A review. Frontiers in Plant Science, 11, 510.

[4] R. Ragunath, R. Rathipriya, "Forecasting Agriculture Commodity Price Trend using Novel Competitive Ensemble Regression Model", International Journal of Information Technology and Computer Science(IJITCS), Vol.17, No.3, pp.97-105, 2025. DOI:10.5815/ijitcs.2025.03.07

[5] Li, Y., Feng, Q., Li, T., Xie, F., Liu, C., & Xiong, Z. (2022). Advance of target visual information acquisition technology for fresh fruit robotic harvesting: A review. Agronomy, 12(6), 1336.

[6] Omarov, B., Batyrbekov, A., Dalbekova, K., Abdulkarimova, G., Berkimbaeva, S., Kenzhegulova, S., ... & Omarov, B. (2020, December). Electronic stethoscope for heartbeat abnormality detection. In International Conference on Smart Computing and Communication (pp. 248-258). Cham: Springer International Publishing.

[7] Ma, L., He, Z., Zhu, Y., Jia, L., Wang, Y., Ding, X., & Cui, Y. (2022). A method of grasping detection for kiwifruit harvesting robot based on deep learning. Agronomy, 12(12), 3096.

[8] Zhang, J., Kang, N., Qu, Q., Zhou, L., & Zhang, H. (2024). Automatic fruit picking technology: A comprehensive review of research advances. Artificial Intelligence Review, 57(3), 54.

[9] Yun, Y., An, X., Li, D., & Ma, Z. (2024). Recent Advances in Intelligent Harvesting Robots. Advanced Sensing and Robotics Technologies in Smart Agriculture, 87-132.

[10] Tan, Y., Liu, X., Zhang, J., Wang, Y., & Hu, Y. (2025). A Review of Research on Fruit and Vegetable Picking Robots Based on Deep Learning. Sensors, 25(12), 3677.

[11] Mingyou, C. H. E. N., Lufeng, L. U. O., Wei, L. I. U., Huiling, W. E. I., Jinhai, W. A. N. G., Qinghua, L. U., & Shaoming, L. U. O. (2024). Orchard-Wide Visual Perception and Autonomous Operation of Fruit Picking Robots: A Review. Smart Agriculture, 6(5), 20.

[12] Liu, J., & Liu, Z. (2024). The vision-based target recognition, localization, and control for harvesting robots: A review. International Journal of Precision Engineering and Manufacturing, 25(2), 409-428.

[13] Li, Z., Yuan, X., & Wang, C. (2022). A review on structural development and recognition–localization methods for end-effector of fruit–vegetable picking robots. International Journal of Advanced Robotic Systems, 19(3), 17298806221104906.

[14] Omarov, B., Tursynova, A., & Uzak, M. (2023). Deep learning enhanced internet of medical things to analyze brain computed tomography images of stroke patients. International Journal of Advanced Computer Science and Applications, 14(8).

[15] Xiao, X., Wang, Y., Zhou, B., & Jiang, Y. (2024). Flexible hand claw picking method for citrus-picking robot based on target fruit recognition. Agriculture, 14(8), 1227.

[16] Omarov, B., Baikuvekov, M., Sultan, D., Mukazhanov, N., Suleimenova, M., & Zhekambayeva, M. (2024). Ensemble Approach Combining Deep Residual Networks and BiGRU with Attention Mechanism for Classification of Heart Arrhythmias. Computers, Materials & Continua, 80(1).

[17] Zhou, H., Wang, X., Au, W., Kang, H., & Chen, C. (2022). Intelligent robots for fruit harvesting: Recent developments and future challenges. Precision Agriculture, 23(5), 1856-1907.

[18] Lei, C., Lei, C., Liu, H., Nong, Z., & Long, L. (2023, October). Intelligent picking robot for orchard based on visual recognition. In IET Conference Proceedings CP850 (Vol. 2023, No. 23, pp. 358-363). Stevenage, UK: The Institution of Engineering and Technology.

[19] Han, C., Lv, J., Dong, C., Li, J., Luo, Y., Wu, W., & Abdeen, M. A. (2024). Classification, advanced technologies, and typical applications of end-effector for fruit and vegetable picking robots. Agriculture, 14(8), 1310.

[20] Wang, W., Li, C., Xi, Y., Gu, J., Zhang, X., Zhou, M., & Peng, Y. (2025). Research Progress and Development Trend of Visual Detection Methods for Selective Fruit Harvesting Robots. Agronomy, 15(8), 1926.

[21] Zheng, Z., Hu, Y., Li, X., & Huang, Y. (2023). Autonomous navigation method of jujube catch-and-shake harvesting robot based on convolutional neural networks. Computers and Electronics in Agriculture, 215, 108469.

[22] Wang, Z., Xun, Y., Wang, Y., & Yang, Q. (2022). Review of smart robots for fruit and vegetable picking in agriculture. International Journal of Agricultural and Biological Engineering, 15(1), 33-54.

[23] Kaleem, A., Hussain, S., Aqib, M., Cheema, M. J. M., Saleem, S. R., & Farooq, U. (2023). Development challenges of fruit-harvesting robotic arms: A critical review. AgriEngineering, 5(4), 2216-2237.

[24] Williams, H. A., Jones, M. H., Nejati, M., Seabright, M. J., Bell, J., Penhall, N. D., ... & MacDonald, B. A. (2019). Robotic kiwifruit harvesting using machine vision, convolutional neural networks, and robotic arms. biosystems engineering, 181, 140-156.

[25] Miao, Z., Yu, X., Li, N., Zhang, Z., He, C., Li, Z., ... & Sun, T. (2023). Efficient tomato harvesting robot based on image processing and deep learning. Precision Agriculture, 24(1), 254-287.

[26] Rajendran, V., Debnath, B., Mghames, S., Mandil, W., Parsa, S., Parsons, S., & Ghalamzan-E, A. (2024). Towards autonomous selective harvesting: A review of robot perception, robot design, motion planning and control. Journal of Field Robotics, 41(7), 2247-2279.

[27] Abeyrathna, R. R. D., & Ahamed, T. (2024). Approaches for Improving Fruit Detection and Gripping Mechanisms in Orchard Robotic Fruit Harvesting. In IoT and AI in Agriculture: Smart Automation Systems for increasing Agricultural Productivity to Achieve SDGs and Society 5.0 (pp. 349-371). Singapore: Springer Nature Singapore.

[28] Zhao, Y., Wan, X., & Duo, H. (2023). Review of rigid fruit and vegetable picking robots. International Journal of Agricultural and Biological Engineering, 16(5), 1-11.

[29] Rana, V., Thukral, S., Sharma, S., Thakur, K. K., & Sharma, K. (2025). An automatic fruit harvesting system for sustainable agricultural practices: challenges and implementation. New Zealand Journal of Crop and Horticultural Science, 53(5), 2965-2989.

[30] Omarov, N., Omarov, B., Azhibekova, Z., & Omarov, B. (2024). Applying an augmented reality game-based learning environment in physical education classes to enhance sports motivation. Retos, 60, 269-278.

[31] Beldek, C., Dunn, A., Cunningham, J., Sariyildiz, E., Phung, S. L., & Alici, G. (2025, February). Multi-vision-based Picking Point Localisation of Target Fruit for Harvesting Robots. In 2025 IEEE International Conference on Mechatronics (ICM) (pp. 1-6). IEEE.

[32] Omarov, B., Omarov, B., Rakhymzhanov, A., Niyazov, A., Sultan, D., & Baikuvekov, M. (2024). Development of an artificial intelligence-enabled non-invasive digital stethoscope for monitoring the heart

condition of athletes in real-time. Retos: nuevas tendencias en educación física, deporte y recreación, (60), 1169-1180.

[33] Wang, X., Kang, H., Zhou, H., Au, W., & Chen, C. (2022). Geometry-aware fruit grasping estimation for robotic harvesting in apple orchards. Computers and Electronics in Agriculture, 193, 106716.

[34] Sun, T., Zuo, H., Zhou, Z., Wang, Y., & Miao, Z. (2025, February). Enhancing safety during harvesting with a deep learning-based action determination method for picking robots. In International Conference on Optical and Photonic Engineering (icOPEN 2024) (Vol. 13509, pp. 248-253). SPIE.

[35] Yoshida, T., Onishi, Y., Kawahara, T., & Fukao, T. (2022). Automated harvesting by a dual-arm fruit harvesting robot. ROBOMECH journal, 9(1), 19.

[36] Yu, Y., Zhang, K., Liu, H., Yang, L., & Zhang, D. (2020). Real-time visual localization of the picking points for a ridge-planting strawberry harvesting robot. Ieee Access, 8, 116556-116568.

[37] Omarov, B., Altayeva, A., & Cho, Y. I. (2017, May). Smart building climate control considering indoor and outdoor parameters. In IFIP International Conference on Computer Information Systems and Industrial Management (pp. 412-422). Cham: Springer International Publishing.

[38] Xiao, X., Jiang, Y., & Wang, Y. (2025). Key Technologies for Machine Vision for Picking Robots: Review and Benchmarking. Machine Intelligence Research, 22(1), 2-16.

[39] Yin, Z., Li, H., Zuo, Z., & Guan, Z. (2025). Key technologies of tomato-picking robots based on machine vision. International Journal of Agricultural and Biological Engineering, 18(3), 247-256.

[40] Jin, Z., Sun, W., Zhang, J., Shen, C., Zhang, H., & Han, S. (2020, February). Intelligent tomato picking robot system based on multimodal depth feature analysis method. In IOP Conference Series: Earth and Environmental Science (Vol. 440, No. 4, p. 042074). IOP Publishing.

[41] Xiaojie, S., Wang, S., Zhang, B., Zixuan, Z., Wang, S., Xinbing, D., ... & Yang, H. (2025). Advances in Berry Harvesting Robots. Horticulturae, 11(9), 1042.

[42] Xiang, J., Wang, L., Li, L., Lai, K. H., & Cai, W. (2024). Classification-design-optimization integrated picking robots: A review. Journal of Intelligent Manufacturing, 35(7), 2979-3002.

[43] Zhang, H., Wang, B., Su, L., Yu, Z., Liu, X., Meng, X., ... & He, X. (2025). Current Research Status and Development Trends of Key Technologies for Pear Harvesting Robots. Agronomy, 15(9), 2163.

[44] Latif, G., Mohammad, N., & Alghazo, J. (2023). DeepFruit: A dataset of fruit images for fruit classification and calories calculation. Data in Brief, 50, 109524.

[45] Yang, S., Zhang, Z., Hao, J., Wang, C., & Qi, G. (2026). Picking Path Planning for One-Eye-Dual-Arm Green Walnuts Harvesting Robot. Smart Agricultural Technology, 101804.

[46] Jun, J., Kim, J., Seol, J., Kim, J., & Son, H. I. (2021). Towards an efficient tomato harvesting robot: 3D perception, manipulation, and end-effector. IEEE access, 9, 17631-17640.

[47] Singh, R., Seneviratne, L., & Hussain, I. (2024). A deep learning-based approach to strawberry grasping using a telescopic-link differential drive mobile robot in ros-gazebo for greenhouse digital twin environments. IEEE Access.

[48] Yin, H., Sun, Q., Ren, X., Guo, J., Yang, Y., Wei, Y., ... & Zhong, M. (2023). Development, integration, and field evaluation of an autonomous citrus-harvesting robot. Journal of Field Robotics, 40(6), 1363-1387.

[49] Gao, Y., Jiang, Q., Wang, M., & Dong, X. (2026). Advances in Path-Planning Algorithms for Agricultural Robots. Journal of Field Robotics, 43(1), 89-145.

[50] Gursoy, E., Navarro, B., Cosgun, A., Kulić, D., & Cherubini, A. (2023, August). Towards vision-based dual arm robotic fruit harvesting. In 2023 IEEE 19th International Conference on Automation Science and Engineering (CASE) (pp. 1-6). IEEE.

[51] Jia, W., Zhang, Y., Lian, J., Zheng, Y., Zhao, D., & Li, C. (2020). Apple harvesting robot under information technology: A review. International Journal of Advanced Robotic Systems, 17(3), 1729881420925310.

[52] Khan, H. A., Farooq, U., Saleem, S. R., Rehman, U. U., Tahir, M. N., Iqbal, T., ... & Hussain, S. (2024). Design and development of machine vision robotic arm for vegetable crops in hydroponics. Smart Agricultural Technology, 9, 100628.

[53] Feng, W. J., Zhao, Y. L., Xie, Y., & Wei, L. (2025). Enhanced YOLOv5s-Based Apple Detection for Harvesting Robots Using a Multi-Rotated Box Algorithm. Informatica, 49(18).

[54] Syafiqah, I., Pamungkas, D. S., & Ramli, N. A. M. (2026). A Robot for Collecting Objects Based on the Convolutional Neural Network Method and Inertial Measurement Unit Sensor. Advanced Mechanical and Mechatronic Systems, 2(1), 1-15.

[55] Li, C. (2020). Badminton motion capture with visual image detection of picking robotics. International Journal of Advanced Robotic Systems, 17(6), 1729881420969072.