



LEEDS  
BECKETT  
UNIVERSITY

---

Citation:

Alaudeen, KM and Selvarajan, S and Manoharan, H and Jhaveri, RH (2024) Intelligent robotics harvesting system process for fruits grasping prediction. Scientific Reports, 14. pp. 1-15. ISSN 2045-2322 DOI: <https://doi.org/10.1038/s41598-024-52743-8>

Link to Leeds Beckett Repository record:

<https://eprints.leedsbeckett.ac.uk/id/eprint/10546/>

Document Version:

Article (Published Version)

---

Creative Commons: Attribution 4.0

© The Author(s) 2024

The aim of the Leeds Beckett Repository is to provide open access to our research, as required by funder policies and permitted by publishers and copyright law.

The Leeds Beckett repository holds a wide range of publications, each of which has been checked for copyright and the relevant embargo period has been applied by the Research Services team.

We operate on a standard take-down policy. If you are the author or publisher of an output and you would like it removed from the repository, please [contact us](#) and we will investigate on a case-by-case basis.

Each thesis in the repository has been cleared where necessary by the author for third party copyright. If you would like a thesis to be removed from the repository or believe there is an issue with copyright, please contact us on [openaccess@leedsbeckett.ac.uk](mailto:openaccess@leedsbeckett.ac.uk) and we will investigate on a case-by-case basis.



OPEN

# Intelligent robotics harvesting system process for fruits grasping prediction

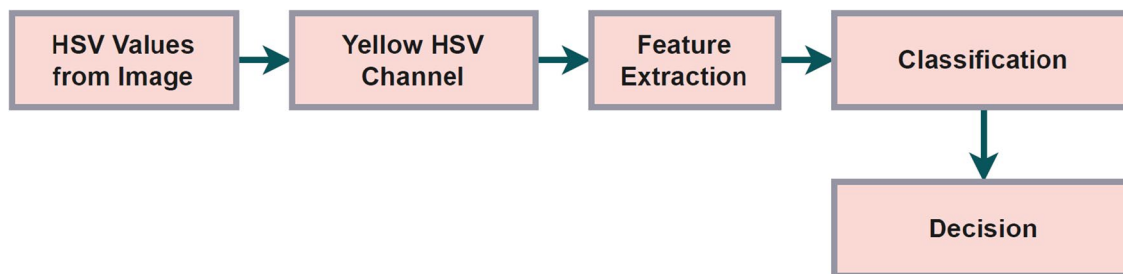
K. M. Alaaudeen<sup>1,6</sup>, Shitharth Selvarajan<sup>2,3,6</sup>✉, Hariprasath Manoharan<sup>4,6</sup> & Rutvij H. Jhaveri<sup>5,6</sup>

This paper proposes and executes an in-depth learning-based image processing approach for self-picking apples. The system includes a lightweight one-step detection network for fruit recognition. As well as computer vision to analyze the point class and anticipate a correct approach position for each fruit before grabbing. Using the raw inputs from a high-resolution camera, fruit recognition and instance segmentation are done on RGB photos. The computer vision classification and grasping systems are integrated and outcomes from tree-grown foods are provided as input information and output methodology poses for every apple and orange to robotic arm execution. Before RGB picture data is acquired from laboratory and plantation environments, the developed vision method will be evaluated. Robot harvest experiment is conducted in indoor as well as outdoor to evaluate the proposed harvesting system's performance. The research findings suggest that the proposed vision technique can control robotic harvesting effectively and precisely where the success rate of identification is increased above 95% in case of post prediction process with reattempts of less than 12%.

As the entire society is moving towards autonomous operations it is necessary to carry out different application operations such as military, civil aviation, medical applications, transportations etc. with the help of robotic systems. It is observed that in recent time most of the robotic systems are currently deployed in the process of agriculture as a replacement of humans as it is considered as time consuming process. In the process of Internet of Things (IoT) where multiple sensors are used in the detection process the agricultural lands will be surveyed during crop harvesting thereby appropriate crops will be planted and monitoring status is defined in an exact way. As a replacement of IoT the proposed method considers robotic system which is connected with multiple sensors for monitoring states. The major advantage of the replacement process is that the robotic systems can move in all different directions and it will analyze the entire area of agricultural land by using a computer vision procedure. In addition all the obstacles that are present in the mid-way is identified and changes are provided for processing autonomous operations by using a fixed camera at the front end of each robot. The above mentioned process is carried out at higher speed and the exact solution for plantation is provided by considering the depth of enclosures. Further the designed robot can able to identify the amount of water that is present in the agricultural land and the time period of yield will be determined in such cases and the outcomes of monitoring states is transmitted to corresponding individuals within short period of time.

Figure 1 portrays that images are captured at the front end of robotic systems where a High Standard Value (HSV) is grasped from image set that is present under same agricultural area. The HSV values are used for determining the possibility of initial state conditions where a yellow HSV channel is connected and entire feature of agricultural lands are extracted. Further every image is classified according to necessary contents and the robot provides an automated decision based on the current condition of crops or other types of plantations. Once the plantation is processed then the robotic systems can able to recognize different types of fruits and it will pick up the necessary fruit without any obstacles..

<sup>1</sup>Department of Computer Science and Engineering, Grace College of Engineering, Mullakkadu, Thoothukoodi, India. <sup>2</sup>School of Built Environment, Engineering and Computing, Leeds Beckett University, LS1 3HE Leeds, UK. <sup>3</sup>Department of Computer Science, Kebri Dehar University, Kebri Dehar, Ethiopia. <sup>4</sup>Department of Electronics and Communication Engineering, Panimalar Engineering College, Poonamallee, Chennai, Tamil Nadu, India. <sup>5</sup>Department of Computer Science and Engineering, School of Technology, Pandit Deendayal Energy University, Gandhinagar, India. <sup>6</sup>These authors contributed equally: K. M. Alaaudeen, Shitharth Selvarajan, Hariprasath Manoharan and Rutvij H. Jhaveri. ✉email: shitharths@kdu.edu.et



**Figure 1.** Robotic decision development with image processing units.

### Background and related works

In the mid-1980s, software engineering was utilized to enhance farming management practices and provide new advancements. Horticultural advancements are progressing in the recently introduced paradigm of Agriculture 4.0<sup>1</sup>. The executives in Agriculture 4.0 combine emerging innovations such as mechanical technology, the Internet of Things (IoT), machine vision, and artificial intelligence with a focus on economic yield. Agriculture 4.0<sup>2</sup> can be seen as an extension of the Industry 4.0 vision, specifically tailored for the agricultural sector. Agricultural cyber-physical systems (CPS)<sup>3,4</sup>, along with agricultural IoT, have recently garnered significant interest. Actuators and sensors that interact with the physical environment can be utilized to intelligently and automatically control crop growth settings. The issues confronting the development of CPS<sup>5,6</sup> include real-time processing, adaptability, hardware and software code verification, and scalability. Fruit recognition is a critical challenge in autonomous agricultural applications. Traditional approaches involve the utilisation of manually designed characteristics to represent the objects, followed by the application of machine learning techniques to identify or separate an object based on these extracted characteristics. Agricultural cyber-physical systems (CPS) have recently gained significant attention, surpassing the focus on Internet of Things (IoT) in farming<sup>7,8</sup>. Therefore, it is possible to control intelligent and automated crop development settings by interacting with real environmental variables using sensors and actuators. The development of CPS will encounter various problems, such as the need for continuous planning, adaptability, coordination of equipment programming, and diversity<sup>9–11</sup>. Exploration often centers around the development of intelligent agricultural robots that can efficiently operate on a large scale, especially those with practical material capabilities. Agrobots have been deployed to this location for precise horticultural tasks such as weed removal, harvesting, irrigation, and maintenance. An innovative concept for a robotic system designed specifically for harvesting cherry tomatoes is presented. The robot system comprises a railed vehicle, a stereo vision device, a fruit collector, a manipulator, and an end effector. An automated kiwifruit harvesting robot design has been documented.

In<sup>12</sup>, the robot utilizes a Global-Positioning System (GPS) and machine vision to adhere to radio commands and autonomously explore; the structure consists of four picking arms that are controlled by a single centralized handling unit. A ground robot, equipped with a camera, was utilized alongside a small dosage framework to develop a compact and precise herbicide administration system<sup>13</sup>. Another artwork portrayed a self-governing robot specifically engineered for the purpose of harvesting apples. The system comprises three primary elements: a mobility apparatus, a visual system, and a mechanical arm outfitted with a gripper. The current architectural design includes a comprehensive system that consists of an RGB-D camera, a gripper, and a mechanical arm<sup>14</sup>. These components are smoothly incorporated into an autonomous wheeled robot. The mechanized weeding versatile stage provides three mechanical implements for weed control, namely the bolt cultivator, the prong, and the cutting tool. The choice of the suitable gadget relies on the particular weed that has been identified. The dominant pattern seen in the current literature is on the prevalence of mobile robots that are specifically developed and optimized for particular tasks. Multiple agricultural robots are currently available on the market, each designed for certain agricultural tasks. There exists a robot named Harvest that is specifically engineered for the purpose of harvesting strawberries<sup>15</sup>. Another robot, named Guss, is employed for the purpose of orchard spraying. In addition, there are agricultural robots named Oz, Ted, and Dino that are specifically utilized for the task of removing weeds<sup>16</sup>. The Digital Farmhand, Farm droid, and Clear Path14 are examples of commercially accessible versatile robotic platforms specifically engineered to carry out diverse agricultural functions. Collecting sensor data is a crucial component of computer hardware architecture. Indeed, greenhouses are not employed for cultivating all types of crops. The planting habitats of certain crops, such as dragon fruits, are undergoing expansion and modification. Cameras are strategically placed in designated regions inside agricultural systems to capture images of crop growth, enabling the monitoring of crop development. Nevertheless, this method of image capture restricts the comprehensiveness of recording crop growth images, as it can only follow a limited number of crop sides. The number of sensors used increases directly in relation to the size of the farm. Another method in recent years has been to employ robot movement to interact with sensors<sup>15</sup>. Data collection through the utilisation of sensors is highly versatile and efficient, thereby leading to potential cost reduction in the installation of sensor networks. Furthermore, certain researchers<sup>16</sup> devised the robots by employing integrated convolutional neural networks (CNNs) to identify certain crops for subsequent processing, hence enabling accurate documentation of crop growth images. Defoliation, lateral branch reduction, and green harvest are widely recognized viticulture procedures that significantly influence grape quality<sup>17</sup>. Activities such as canopy management, pre-harvest, post-harvest, and harvest are considered to have a beneficial effect on the quality of wine in valuable crops such as wine grapes. Regarding high-value crops such as wine grapes, it is believed that the management of shelter, pre-harvest, post-harvest, and storage practices have a significant impact on the

quality of wine. Agrobots are designed to enhance viticulture by reducing the need for human labor through the execution of viticulture tasks with the proficiency of a skilled professional<sup>18</sup>.

### Major contributions

To create a fully operational robotic harvesting system, there are still a lot of obstacles to overcome. Therefore this paper offers a yield scale multi-reason autonomous grape gatherer robot, or ARG for short, that is designed to do viticulture tasks such as collect, green reap, and defoliation instead of a skilled specialist. In practical implementation settings, traditional vision methods are susceptible to mistakes in precision, resilience, and efficiency. The development and evaluation of an automated apple harvesting vision approach using deep learning is the focus of this work. In this paper, the following contributions are highlighted:

- A computationally efficient one-stage instance segmentation procedure for conducting fruit identification and instance segmentation on sensory input is proposed using computer vision techniques.
- Using edge and contour points from an RGB camera, propose an enhanced computer vision-based network for ripe fruit detection and grasping estimate.
- System integrates and combines the above two functions while designing and implementing an accurate robotic system for autonomous fruit harvesting

### Paper organization

The remainder of this document is laid out as follows: The second section looks at studies on fruit recognition and grasping estimation that is comparable to the first. “Materials and methods” section introduces the suggested visual processing algorithm’s techniques. The experimental setup and results are presented in “Results” section. “Conclusions” section concludes the research work with future updates.

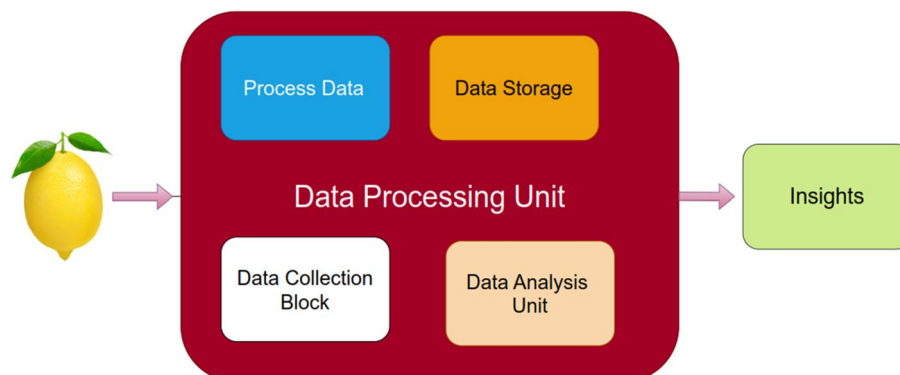
### Materials and methods

The mobile base is a special mobile vehicle mainly composed of a main control unit, four electric wheels, a 24 V power supply and a vehicle chassis. The entire robotic system, including the movable base, is programmed to find the correct position. Intelligent Robotics is a robotic arm with 6 degrees of freedom. Along with end effectors, manipulators support path development. Our end-effectors are built using soft robotic grippers that have been extensively studied for robotic grippers. The secure touch and the yielding mechanism are integrated in the suggested end effectors as the consequence of ribs and low modulus material. The Real Sense RGBD camera is a major component of the visual subsystem. This is used for fruit photography for processing the upcoming data. The collected fruit position and alignment data is used to control robotic reaping equipment. Figure 2 shows the overall operation of the robotic harvesting system.

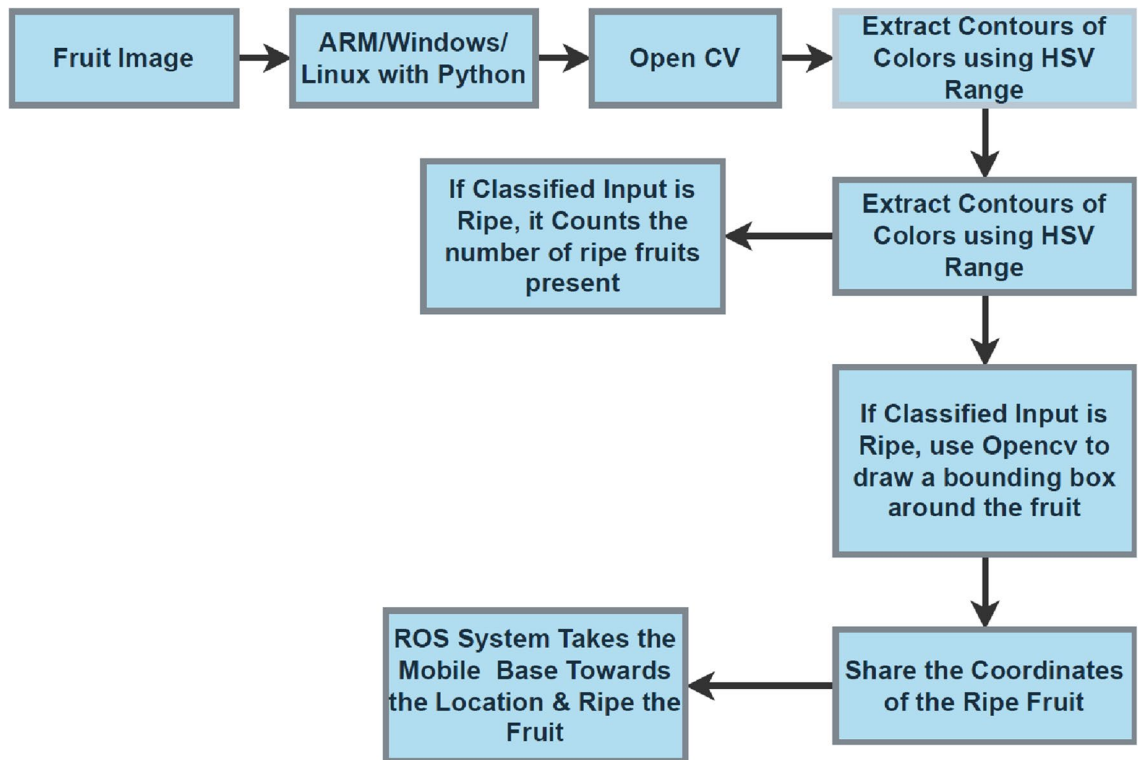
The two phases in the our vision technique are fruit identification and grabbing assessment. The RGB images are segmented and detected by the suggested vision algorithm in its first step. Contours are created by mixing the anticipated mask of every fruit with the depth image from Open CV’s (Open Computer Visions), input point clouds. Contours, using Open CV will anticipate the form, size, and approaching attitude of each fruit based on the output from the first phase in the second step. Figure 3 provides step-by-step processing for apple harvesting robotic systems.

### Network architecture

To perform fruit identification, consider the lemon embed with this study uses computer vision and HSV channels to create an advanced, lightweight, one-step, Fully Convolutional detection network called “Single Shot Detection (SSD)”. Instead of using resnet50 as the backbone network and the previous network using a three-layer pyramidal network of features, the proposed classification based on SSD HSV uses simplified “computer vision” backbone and HSV values. On a 128-core GPU based on NVIDIA Maxwell™ architecture, the proposed SSD weighs 20.5 MB and averages 30 frames per second. Each classification level has an instance split branch and an edge coordinate detection block. To analyze data at multiple scales, computer vision, scaled convolutions with



**Figure 2.** Proposed architecture of intelligent robotics harvesting system process for fruits grasping prediction.



**Figure 3.** Proposed apple harvesting robot software designs.

different aspect ratios are often employed to detect objects of various shapes and sizes within an image. These convolutions, often utilized in object detection frameworks like Single Shot Multibox Detector (SSD) or You Only Look Once (YOLO), involve using convolutional kernels of different aspect ratios (typically rectangles with various width-to-height ratios) to capture objects with diverse shapes and uses scaled convolutions with different ratios (eg 1, 2, 4). Segmentation can be classified into two as follows: segmentation by mask and segmentation by detection. The confidence score, bounding box and class of the mesh elements are predicted by the detection segmentation method. Each classification level uses a bounding box which has a preset value of  $50 \times 50$  at C4 level and  $120 \times 120$  at C5 level respectively. The process of segmentation by binary branch is based on the SPR Net architecture, which allows you to predict binary masks for features in a feature map on a pixel-by-pixel basis. Although not used in this work, semantic branching in computer vision methods is also designed for semantic branching.

To increase the versatility of the classification test, we collected 1200 photos from different contexts. For example, farmlands in China, Melbourne and Australia have different settings and lighting, including artificial lighting, shadows, daylight, lighting in front, side and backlighting. Label images were used to confirm classification. HSV correction uses reflection (horizontal only), random saturation (0.8–1.2), scaling (0.8–1.2), rotation (10), and brightness (0.8–1.2). A high-definition camera collects images of the real world with a computer vision process. Computer vision supports conversion from RGB to HSV. With the help of HSV, the values are sent to a specific channel responsible for feature extraction of the classification system according to the selected threshold. When the threshold is exceeded, the program creates a bounding box and displays the result as shown in Fig. 4. This is the first part included in the data acquisition system to inject data into the ROS system for data collection.

### Grasping estimation

This graphic depicts an apple as a spherical shape. In natural circumstances, apples can be blocked from the RGB-D camera's view-angle. As a result, the visible area of the apple from the RGB-D camera's current view-angle shows the robotic arm approaching the target in the correct location. As an object pose estimation problem, we employ Contours with Open CV to provide grasping estimates. From the current view angle, as an approaching posture, a vector from the geometric center to the visible center of the apple is selected. Our system can predict the approaching posture using only a single-viewed point cloud as input, decreasing the operation time in half and the dynamics model of prosthetic can be established. The equation is shown as follow.

$$Mq + H(q, q) = \gamma^i - JF - cq - k_i q - T_k \quad (1)$$

where  $M$  represent the Matrix,  $H(q, q)$  a need to pick the object for grasping mechanism,  $\gamma^i$  denotes the torque used for the motion,  $F$  denoted the external force,  $k_i q$  need for restoring point and  $T_k$  initial condition of the robot.

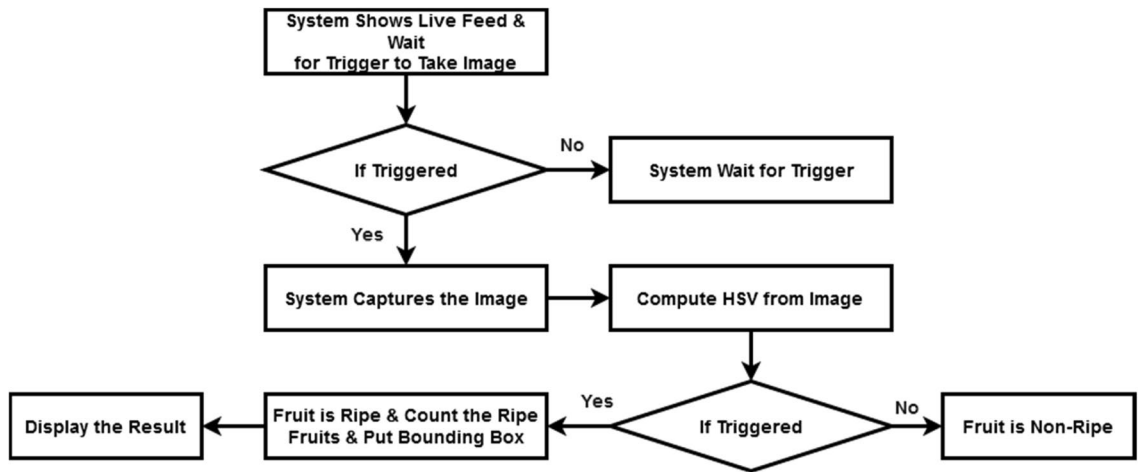


Figure 4. Fruit classification architecture.

### Contours using open CV architecture

In this work, maybe already familiar with the word contour. This term has been mentioned several times in previous posts. A contour line is a curved line that represents the threshold between two sets of values or intensities. Because the two terms are frequently used interchangeably, it might be perplexing. Simply expressed, the concept of edges exists in a local range, but the concept of contours exists at overall border as shown in Fig. 5. Edges are points where the values differ greatly from those of their neighbors. Contours, on the other hand, are closed curves formed by edges and representing a figure’s boundaries. The contour can be used for a wide range of purposes. With the help of the concept of image moment, we may locate the centroid of a picture or calculate the area of a boundary field. In popular usage, a ‘moment’ refers to a brief period. However, in physics, a moment is defined as the product of a distance and another physical quantity, which refers to how a physical quantity is dispersed or positioned. Image moment describes how image pixel intensities are dispersed according to their position in computer vision. We can derive the centroid or spatial information from the picture moment, which is a weighted average of image pixel intensities.

$$x = M_{10}/M_{00} \tag{2}$$

$$y = M_{01}/M_{00} \tag{3}$$

In the X-axis, the variable ‘M’ transitions from the positive to the negative axis initially, while in the Y-axis, it ranges from the negative to the positive axis in relation to the zero axis.

### Results

A high definition camera helps us to collect the real-time feed from the field from the image our computer vision techniques extracts the HSV(Hue Saturation Value) value classifies the fruit as ripe or non-ripe and share a contour, edge, and center co-ordinate to the moving vehicle which is basically a manipulator for industrial robotics a customized soft end-effector for an intelligent robot Logitech 1080UHD USB web camera, and an

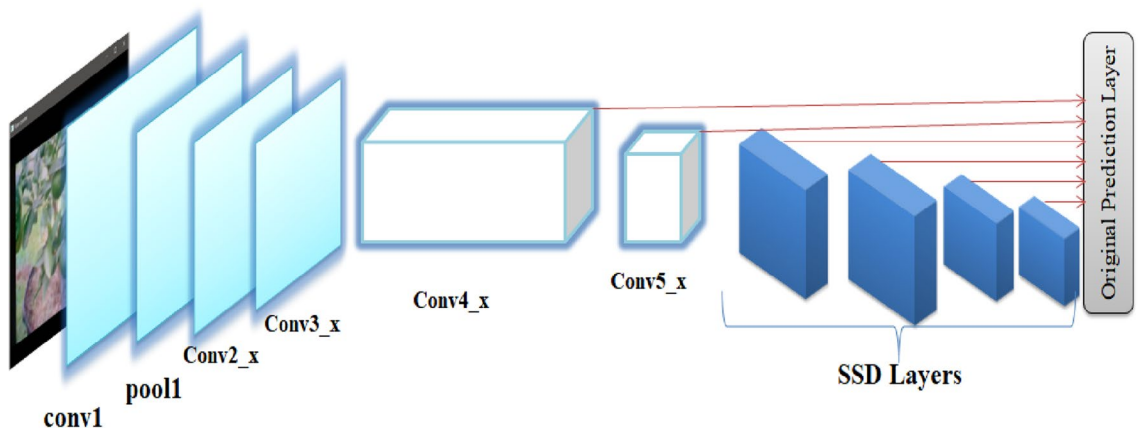


Figure 5. SSD architecture for fruits grasping prediction.

ARM Cortex-A53 1.4GHz central control computer make up the designed robotic harvesting system. With the help of HSV, it transfers the value to the particular channel that the channel takes care of the feature extraction for the classification system. The classification system classifies and decided to rip the fruit or not as shown in Fig. 1. If it is ripe fruit then the system sends the coordinates to ripe the fruit to the control system. The control system is built using Linux Ubuntu 18.04 and the Robot Operating System (ROS) in Melodic along with the mobile robot unit. ROS and Gazebo handle communication between the RGB camera and the computer which takes care of the computer vision process of classification. To evaluate the laboratory and real time outcomes for designed robotic systems the following case studies are considered.

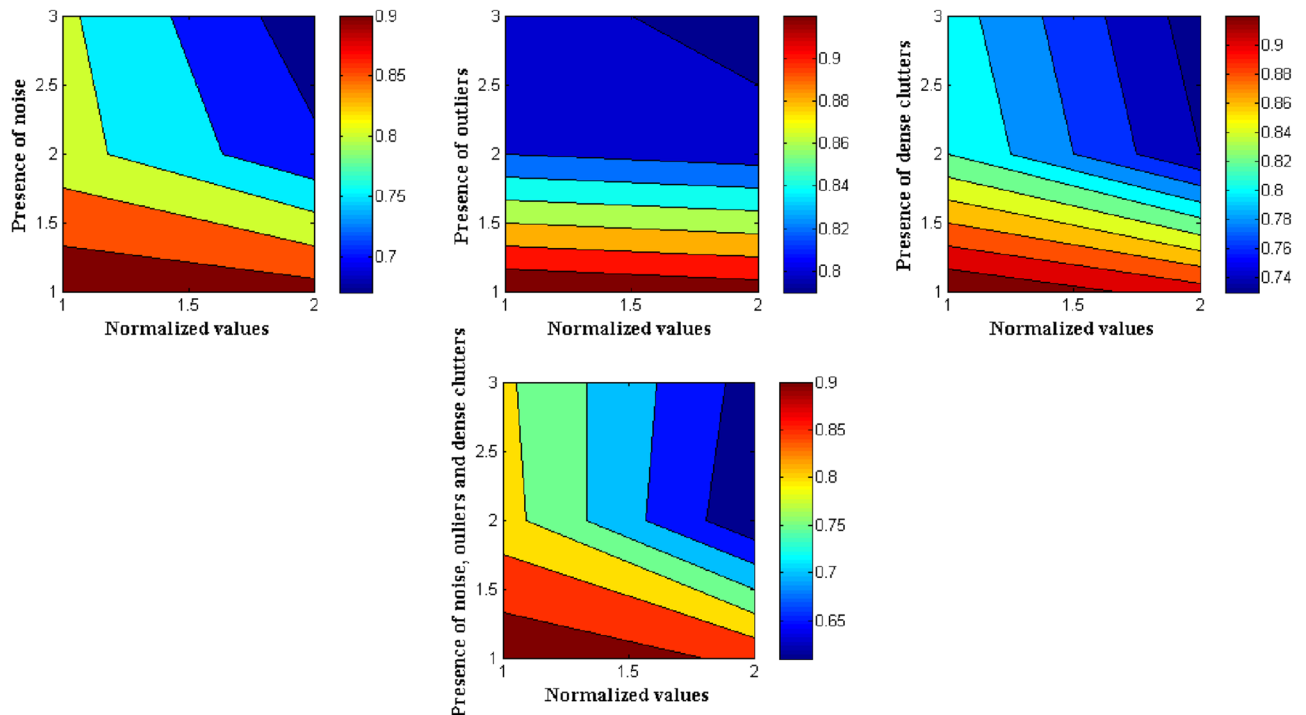
- Case study 1: Fruit form estimation.
- Case study 2: Accuracy of grasping.
- Case study 3: Robotic hold conditions.

All the case studies are performed with image classification technique using python coding and the sample code is provided as follows.

The unique deep learning-based technique was compared to two old methods, sphere Random Sample Consensus (sphere-RANSAC) and sphere Hough Transform (sphere-HT). The RANSAC and HT techniques both employed point clouds to estimate the form of the fruit. RGB-D pictures from both laboratory and orchard contexts were used to do this comparison. We included a thick clutter condition in the experiment to see how well the algorithm performed when the fruits were close together.

### Case study 1

Figure 6 shows the outcomes of a variety of tactics in various settings. Contours using an Open CV-based technique greatly improves the localization accuracy of the 2D bounding box (0.94 in typical conditions), outlier (1% to 5% of the aggregate of point clouds), and point clouds to assess the durability of various strategies in dealing with noisy and unexpected conditions. When dealing with outliers, three strategies showed comparable robustness. Both RANSAC and HT employed a voting system to estimate the shape's primitives, which proved to be resilient against outliers. When confronted with noisy data, Contours employing Open CV-based approaches displayed far stronger resilience, with just a 3% loss in outcomes when compared to the usual scenario, whereas exhibited substantial decreases in accuracy compared to the Contours using Open CV. Contours utilizing Open CV outperformed the other two strategies in the dense clutter case when compared to the other two techniques. When compared to existing approaches, the experimental findings show that the Contours using the Open CV-based method enhances grasping estimating accuracy and resilience.



**Figure 6.** Comparison of different activities with normalized values.

<b>//Sample Code</b>	
<pre> while True:     ret_val, img = cam.read()     if mirror:         img = cv2.flip(img, 1)         (h, w) = img.shape[:2]         blob = cv2.dnn.blobFromImage(cv2.resize(img, (300, 300)), 0.007843,(300, 300), 127.5)         #cv2.imshow('my webcam', img)         hsv = cv2.cvtColor(img, cv2.COLOR_BGR2HSV)         mask = cv2.inRange(hsv, ripe, raw)         cv2.imshow('Image',img)         cv2.imshow('Result',mask)          print("[INFO] computing object detections...")         net.setInput(blob)         detections = net.forward()         for i in np.arange(0, detections.shape[2]): # extract the confidence (i.e., probability)             associated with the prediction </pre>	<pre> IMAGE CLASSIFIER:  #!/usr/bin/env python import sys if sys.version_info[0] &gt;= 3:     import PySimpleGUI as sg import cv2 import os import numpy as np from sys import exit as exit from PIL import Image Width=540 Height=500 WidthB=50 HeightB=50 # Get the folder containing the images from the user folder = sg.PopupGetFolder('Image folder to open') if folder is None:     sg.PopupCancel('Cancelling')     exit(0)  # get list of PNG files in folder png_files = [folder + '\\\ + f for f in os.listdir(folder) if '.png' in f] filenames_only = [f for f in os.listdir(folder) if '.png' in f] #print(png_files, '\n', filenames_only) </pre>
<pre> #----- window = sg.Window('Image Dttset creator',     return_keyboard_events=True,     resizable=True,     #text_justification='c',     auto_size_text=True,     auto_size_buttons=True,     no_titlebar=False,     location=(0,0),     use_default_focus=False ).Layout(layout)  # loop reading the user input and displaying image, filename i=0 top_left=(0,0) bottom_right=(50,50) while True:      event, values = window.Read() </pre>	<pre> NumOFD=0 window.FindElement('NumOFD').Update(NumOFD) elif event in ('Prev', 'MouseWheel:Up', 'Up:38', 'Prior:33') and i &gt; 0:     i -= 1     NumOFD=0     window.FindElement('NumOFD').Update(NumOFD) elif event == 'graph':     mouse = values['graph']     if mouse == (None, None):         continue     print(mouse)     x = mouse[0]     y = mouse[1]     top_left=(x-WidthB,y-HeightB)     bottom_right=(x+WidthB,y+HeightB)     #letter_location = (box_x * 200 + 18, box_y * 200 + 17)     #graph.DrawRectangle((200, 200), (250, 300), line_color="red")     #DrawRectangle(top_left,bottom_right, fill_color=None, line_width=None)     #print(box_x, box_y)  elif event == 'Set Size':     #print(int(values['Width']),int(values['Height']))     Width =int(values['Width'])     Height=int(values['Height'])     print(Width,Height) elif event == 'Set Box Size': </pre>

The experimental results for approaching posture prediction revealed that the Contours using the Open CV-based technique performed well. Experiments showed that Contours using Open CV grasping estimates can detect object grip orientation effectively and robustly in noisy, outlier-presented, and high-clutter settings. The key functions in explainable codes are provided as follows.



```

def load_food_name(food_type):
names = [line.rstrip('\n').lower() for line in open( food_type + '.dict')]
def recognize_food(img_path, list_foods):
start_time = datetime.now()
img = cv2.imread(img_path)
height, width = img.shape[:2]
img = cv2.resize(img, (800, int((height * 800) / width)))
image = vision.types.Image(content=content)
response = client.label_detection(image=image)
labels = response.label_annotations
for label in labels:
if len(text.description) == 10:
desc = label.description.lower()
score = round(label.score, 2)

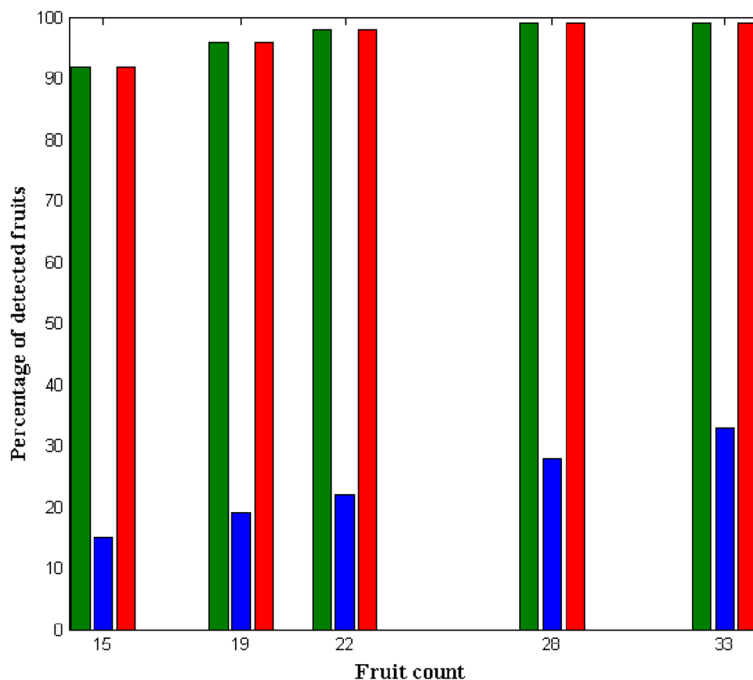
```

The experimental results for approaching posture prediction revealed that the Contours using the Open CV-based technique performed well. Experiments showed that Contours using Open CV grasping estimates can detect object grip orientation effectively and robustly in noisy, outlier-presented, and high-clutter settings.

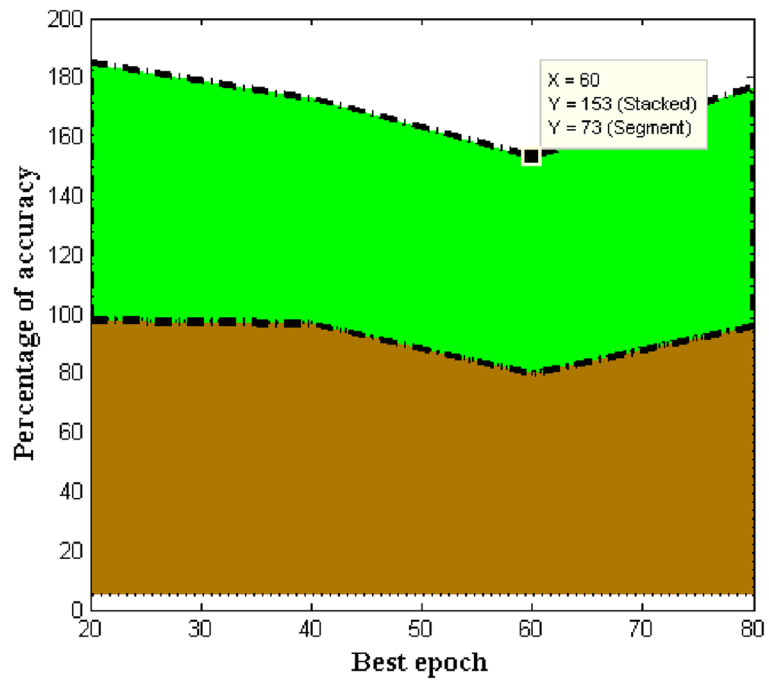
### Case study 2

In this work, we used Open CV grabbing estimations to conduct fruit recognition and contouring on RGB-D photographs obtained in apple orchards as shown in Fig. 7. Fruit identification and segmentation were assessed using the F1 score and the IoU. The performance of Contours using Open CV with HSV grasping estimate, Open CV with HSV with Contours, and Open CV with HSV with Contours were compared and deliberated in Fig. 8. Estimation was more difficult in the orchard situations than it was in the interior environments. In this setting, the Open CV with HSV with Contours and Open CV with HSV with Contours fared significantly worse than in the indoor testing, while the Contours using Open CV grasping estimate did significantly better. Contours' IoU3D in the scenario of an orchard was 0.88, 0.76, and 0.78, respectively, using Open CV grasping estimate, Open CV with HSV with Contours, and Open CV with HSV with Contours.

Even when sensory input was inadequate, contours employing OpenCV gripping estimates worked effectively in determining grip orientation. Contours estimated using OpenCV grasping had a mean orientation estimate error of 6.6, which is in an acceptable limit.



**Figure 7.** Fruit recognition and grasping estimation experiments: (a) only fruits in tree, (b) fruits with low ripe percent fruit and only low ripe percent fruit.



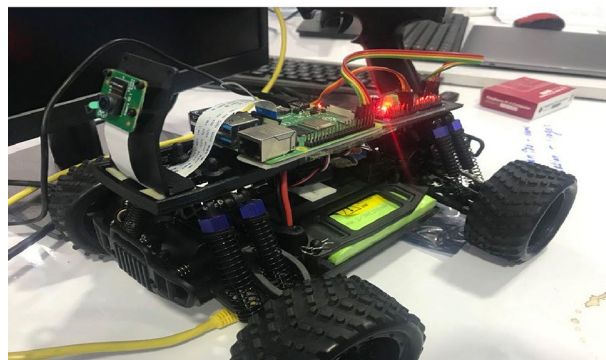
**Figure 8.** Comparison of accuracy using contours and open CV (existing vs proposed).

### Case study 3

To determine the success rate of robotic harvesting, we randomly assigned the number, distribution, and placement of oranges on the orange tree. The experimental setup in indoor environment is illustrated in Fig. 9 and the four phases of the robotic grasping procedure are represented in Fig. 10 where the fruit harvesting

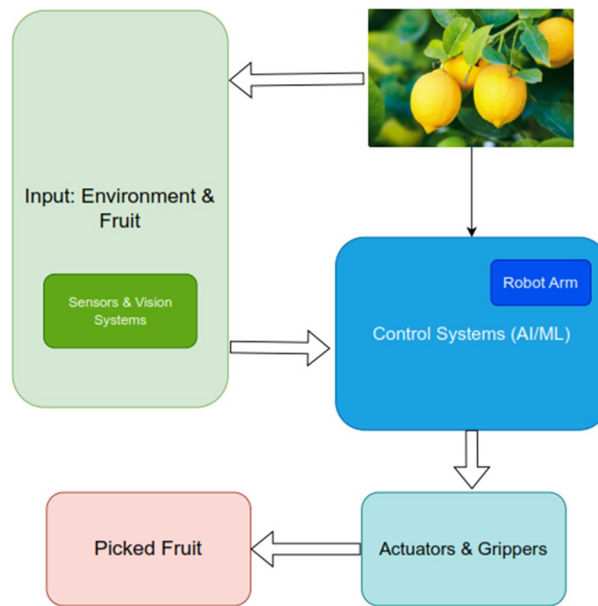


(a)



(b)

**Figure 9.** Experiment setup in indoor environment: (a) indoor connection segments, (b) designed robot.



**Figure 10.** Experiment setup in outdoor environment with four phase considerations.

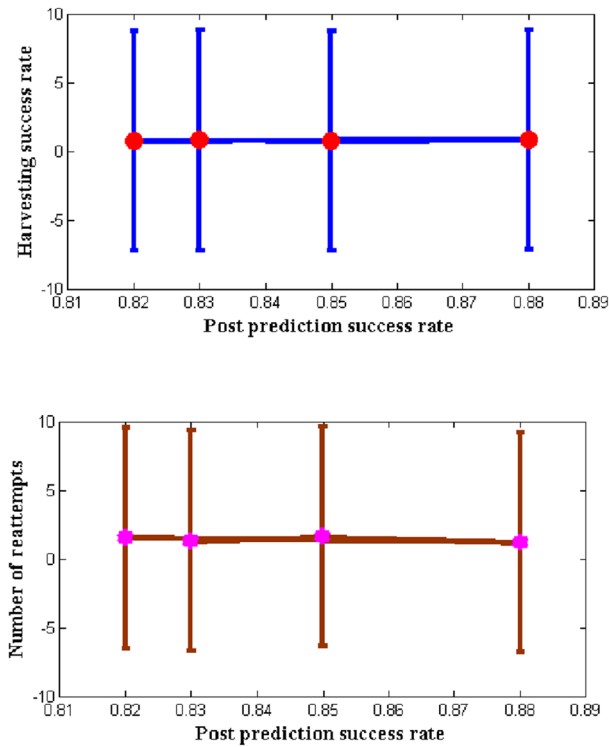
technique and the Pose prediction enabled harvesting method were tested and contrasted, as indicated in Table 1. The natural harvesting approach was just removing fruit without regard for the gripping position of each fruit.

The experimental setup for robotic system that is connected for identifying robotic harvesting focuses on indoor environments where fruits are identified by various colour units. Moreover to provide complete characterization the robotic design is separated in to four parts that are used for fruit verification, grasping, collection units and subsequent identification paths. Hence for the above mentioned functions it is necessary to use a detection system with high accurate camera that can able to identify all fruits without any discontinuity. Further it is possible to provide identification factor in a much easy way at outdoor environments due to the presence of bright features but in indoor environments additional identification units are added in such a way to provide grasping estimate of robots.

The accuracy of gripping position estimation was lower in both indoor and outdoor environments than on RGB-D picture data, as shown in Fig. 11. End-effector instability was found to be the cause of performance loss while the robotic arm was moving, which could lead to erroneous sensory input. From Fig. 11 it is observed that after harvesting the outcomes of robots are identified and plotted with respect to success rates and reattempts. In most of the grasping process that is completed by robotic actions it is realistic that success rate will be increased as compare to manual operation that is performed by individuals. The above mentioned situation is observed in the proposed method for fruit detections as the designed robot is having the capability to represent varying fruit type with proper identification procedures. Since the robot verifies the type of fruit before collecting it in carries the number of reattempts to pick the same fruit is avoided in case of proposed method. However the above mentioned identification process is carried out only for indoor environments therefore in real time the considered identifications will change for outdoor environmental conditions. To verify this case study post prediction success rate is observed as 82, 83, 85 and 88% after collected units are counted therefore for increase in predictions rates the harvesting success rate is increased to 72, 73, 80 and 85% respectively. Similarly for the post predicted success rate the number of reattempts are identified to be minimized at it falls beyond 12%.

References	Method	Pros and cons
19,20	Threshold and circle fitting	High false detection rate and lack of reliability
21,22	Support vector machine (SVM)	Simple to implement, over fitting, and reduced detection accuracy
23-25	Digital twin (DT) and texture classifier	Complex in detection, more time consumption for training and testing
26,27	Faster recurrent neural network (RCNN)	High computational complexity, and lack of scalability
28-30	Semantic segmentation and VGG	Over fitting, computational burden, and high training and testing time
31,32	Adaptive threshold and fusion	Low detection rate, inaccurate, and lack of reliability
33-35	Single shot detection	Ineffective, and high false detection

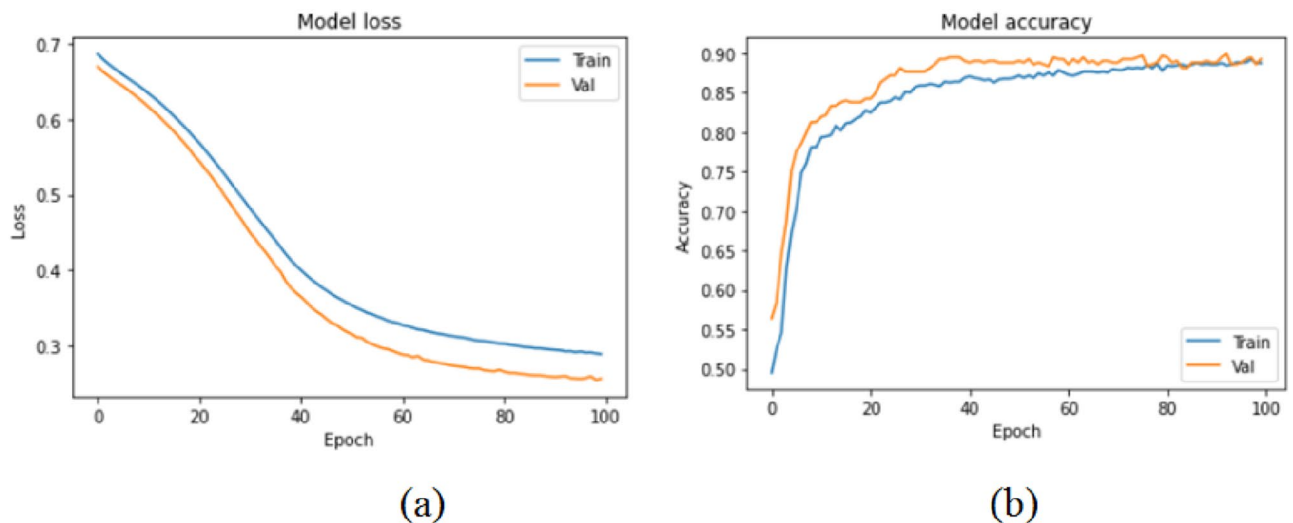
**Table 1.** Existing vs. proposed.



**Figure 11.** Success rate and number of reattempts for proposed and existing approach.

### Performance analysis

As a result, following each robotic arm movement, we introduced a 0.5 s delay to verify that the sensory data input was correct. Gripping failure can be caused by a variety of factors, including a weak grip and a crowded environment. In the first case, one or two fingers of our modified three-fingered end-effector may not have connection with the target fruit, causing the target to slip off the gripper and fail, whereas in the second case, the gripper may have contact with the nearby fruit and led to frostbite for these neighbors. When compared to the naïve harvesting strategy, harvesting in accordance with the prediction of poses dramatically improved the efficiency of harvesting using robots and minimized the miss rates in indoor as well as outdoor contexts. Each trial took 4 and 6.5 s for the naïve harvesting approach and Pose prediction assisted harvesting strategy, respectively. Overall, when tested in indoor as well as outdoor conditions, our suggested vision approach increased the accuracy and sturdiness of robotic harvesting systems as depicted in Fig. 12. The decrease in loss results in higher accuracy.



**Figure 12.** SSD architecture. (a) Model loss, (b) model accuracy.

The recommended deep learning-based solution, Contours using Open CV, outperforms conventional techniques such as Open CV with HSV along with pose estimation and contours alone in the image data trials in terms of robustness when processing noisy data. This is because noise has a significant impact on the accuracy of the voting framework. Among the three techniques, our method had the greatest accuracy in detecting the fruit form estimation in dense clutter. Furthermore, the findings of the experiment showed that Contours using Open CV effectively and robustly predicted the approaching position while grasping under complicated situations including noise, outliers, and thick clutter. The experiment indicated that computer vision with posture estimation, HSV with contours, and computer vision with HSV can all perform. When it comes to recognition of fruit and segmentation of the instances in orchard circumstances, it performs admirably. The proposed one-stage detector for fruit recognition has been proved to be accurate and computationally efficient. This light-weight computer vision SSD resulted in F1 score 0.94 as represented in Fig. 13. There call value is 0.826 and the value of accuracy is 0.9 on fruit detection. The Intersection over Union (IoU) which is a significant measure of accuracy. It indicates the overlap of the predicted and real bounding boxes. The SSD architecture is also efficient in accordance with the execution time its graphical representation as shown in Fig. 14.

Fruit recognition and segmentation activities are accomplished faster with this one-stage detector, resulting in a shorter average cycle time for fruit harvesting. In terms of future advances to the supplied approaches, the suggested Contours employing Open CV and contour functions might be merged into a single step. The robotic harvesting system's real-time performance is projected to increase with the achievement of fruit identification, segmentation, fruit modeling, and grabbing estimations in a single step. The inability to forecast the gripping position was mostly due to a defect in the test dataset sensory data. When the apple in the blue border box failed to generate a point cloud in the first place, the grasping approximation did not advance and was ruled a failure

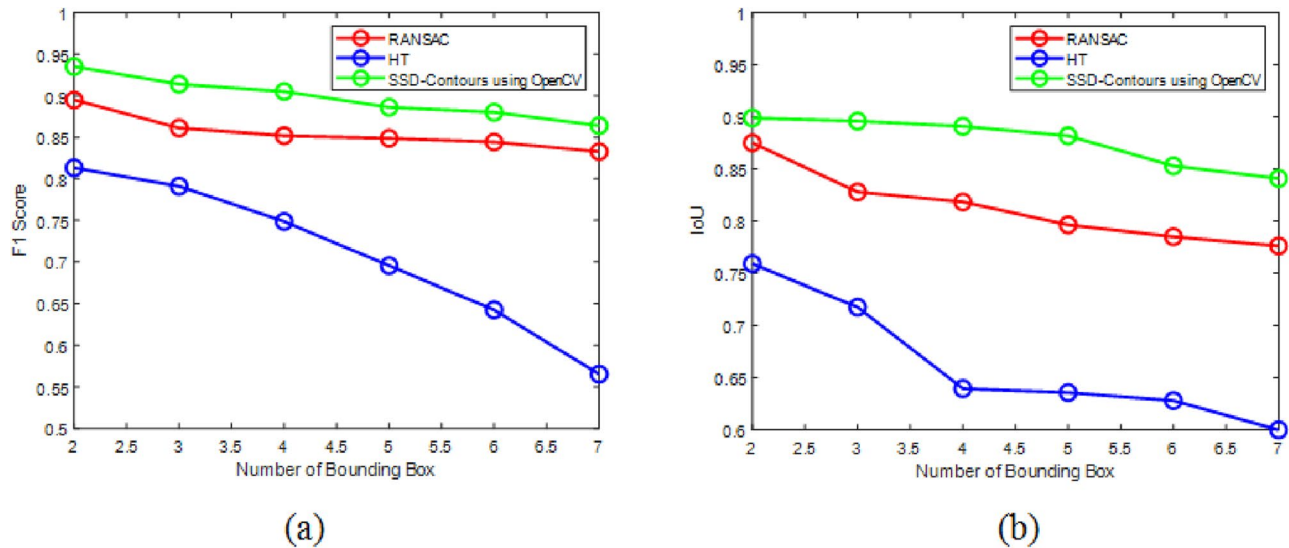


Figure 13. SSD contour (a) F1 score, (b) intersection over union (IoU).

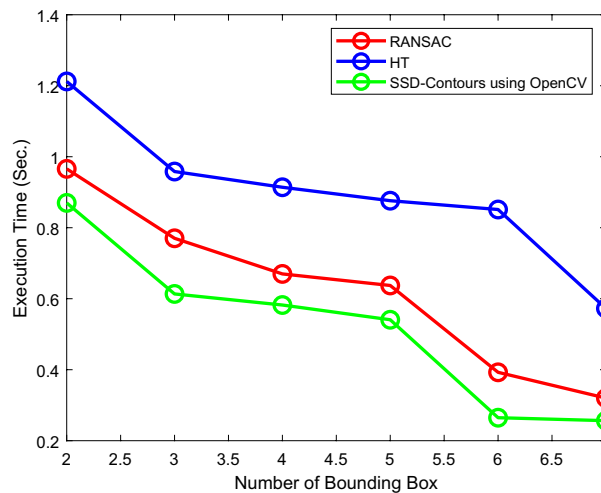


Figure 14. Execution time of SSD contours.

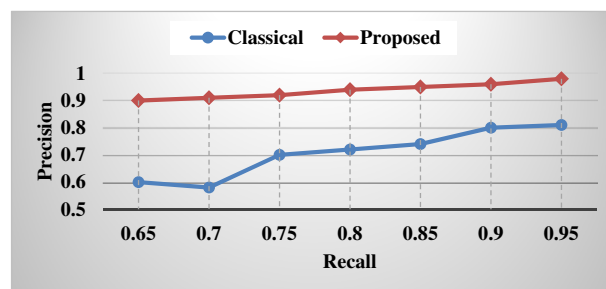
grasping estimation. Because there was no ground truth in this circumstance, there was no perfect value in the grabbing estimation. In this instance, contours using Open CV gripping estimates would always forecast a sphere with a short radius, which may be readily filtered out as outliers during execution.

Even though the former was allowed by posture assessment and the latter can only interpret to the identified fruit, our proposed technique beat the naive harvesting method in terms of higher harvesting success rates and shorter re-attempt durations in robotic harvesting. Gripping failure can result from several circumstances, including a weak grip and heavy debris. In the first case, one or two of our modified three-fingered end-fingers effector's may lose touch with the target fruit (contact with neighboring branches instead or an incorrect gripping posture), causing the target to fall from the gripper and fail. In a congested setting, the gripper can easily contact nearby fruit, causing it to fall. When compared to the naive harvesting strategy, harvesting based on predicting the pose dramatically improved the efficiency of robotic harvesting and minimized miss rate in indoor as well as outdoor situations.

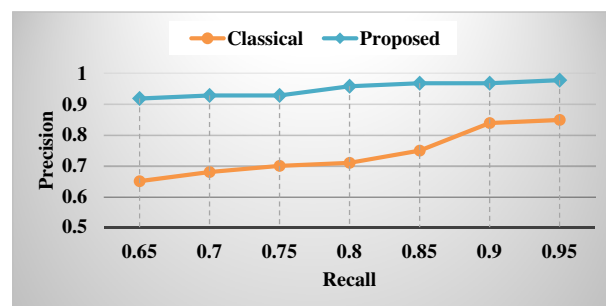
$$mAP = \frac{1}{NC} \sum_{t=i}^N P(t) \Delta R(t) \quad (4)$$

where,  $NC$  represents the number of categories,  $N$  is the count of thresholds,  $t$  indicates threshold,  $P(t)$  is the precision,  $R(t)$  is the recall, and  $mAP$  denotes the average of multiple categories of area under the precision and recall values. The increased value of  $mAP$  states the improved performance of the classifier. According to the analysis, it is estimated that the proposed classification model overwhelms the other deep learning classifiers with improved  $mAP$  for the different types of fruits such as apple, mango and orange. Figure 15 shows the detection results of the classical CNN and proposed models with respect to the precision and recall values. Based on the analysis, it is observed that the detection results for both ripe and unripe fruit detection of the proposed model are highly improved, when compared to the classical CNN model. Consequently, the accuracy of several fruit classification methodologies are assessed as shown in Fig. 16 where the different architecture models of CNN classifier is considered for analysis. Overall, the obtained results indicate that the proposed deep learning based classification model overwhelms the existing architectures with improved fruit detection results.

Where,  $NC$  represents the number of categories,  $N$  is the count of thresholds,  $t$  indicates threshold,  $P(t)$  is the precision,  $R(t)$  is the recall, and  $mAP$  denotes the average of multiple categories of area under the precision and recall values. The increased value of  $mAP$  states the improved performance of the classifier. According to the analysis, it is estimated that the proposed classification model overwhelms the other deep learning classifiers with improved  $mAP$  for the different types of fruits such as apple, mango and orange. Figure 15 shows the detection results of the classical CNN and proposed models with respect to the precision and recall values. Based on the analysis, it is observed that the detection results for both ripe and unripe fruit detection of the proposed model are highly improved, when compared to the classical CNN model. Consequently, the accuracy of several fruit classification methodologies are assessed as shown in Fig. 16 where the different architecture models of CNN

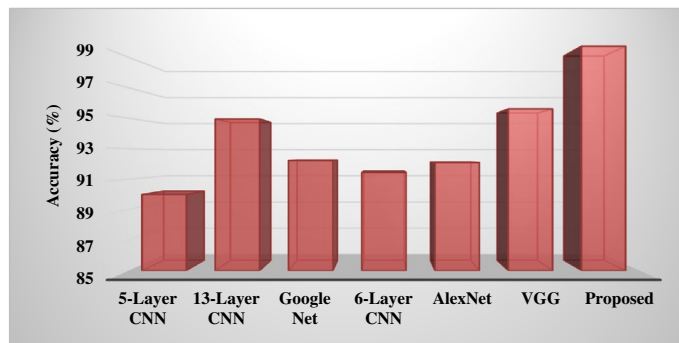


(a)



(b)

**Figure 15.** Detection rates. (a) Ripe fruit, (b) unripe fruit.



**Figure 16.** Accuracy of various fruit classification methodologies.

classifier is considered for analysis. Overall, the obtained results indicate that the proposed deep learning based classification model overwhelms the existing architectures with improved fruit detection results.

## Conclusions

This paper presents an experimental evaluation of a CNN-based robotic harvesting system designed to recognize and grasp fruits. The proposed methodology involves the utilization of a multi-functional network capable of simultaneously performing fruit identification and segmentation. Additionally, an SSD contours approach is employed, which utilizes a Convolutional Neural Network (CNN) based on Open CV, to analyze the point cloud data of the fruit. This analysis is used to estimate the grasping position for each fruit, thereby establishing the most optimal hold position. The adoption of a grasping stance is crucial when engaging in autonomous fruit picking activities. The fruit recognition network and SSD Contours, which have multiple functions, were trained and validated using RGB-D images obtained from both indoor and outdoor environments. The Open CV gripping estimate network was utilized in this process. The findings of the studies indicated that the proposed technique effectively facilitated accurate visual perception and grasping estimations. The computer vision Single Shot Multi Box Detector (SSD) attained F1 scores of 0.94, 0.826, and 0.9, along with recall and accuracy metrics for fruit identification. Additionally, the instance segmentation accuracy for fruit identification was measured at 0.82. Understanding estimates can be challenging. The Contours, employing Open CV grasping estimation, computer vision with posture estimation and HSV, and computer vision with the HSV algorithms, achieved IoU3D values of 0.88, 0.76, and 0.78, respectively, in the orchard scenario. The utilization of Open CV for contour detection outperforms the other two methods. The proposed robotic harvesting system underwent testing in both indoor and outdoor environments, demonstrating promising outcomes in terms of precision, resilience, and operational efficiency. The robotic harvesting system that has been designed demonstrates a harvest success rate of 0.75 and a cycle duration of 6.3 s.

## Data availability

The datasets used and/or analysed during the current study are available from the corresponding author on request.

Received: 20 September 2023; Accepted: 23 January 2024

Published online: 03 February 2024

## References

- Rose, D. C., Wheeler, R., Winter, M., Lobley, M. & Chivers, C.-A. Agriculture 4.0: Making it work for people, production, and the planet. *Land Use Policy* **100**, 104933 (2021).
- Duckett, T. *et al.* Agricultural robotics: the future of robotic agriculture. arXiv preprint [arXiv:1806.06762](https://arxiv.org/abs/1806.06762) (2018).
- Rad, C.-R., Hancu, O., Takacs, I.-A. & Olteanu, G. Smart monitoring of potato crop: A cyber-physical system architecture model in the field of precision agriculture. *Agric. Agric. Sci. Procedia* **6**, 73–79 (2015).
- Grimm, T., Janßen, B., Navarro, O. & Hübner, M. The value of FPGAs as reconfigurable hardware enabling Cyber-Physical Systems. In *2015 IEEE 20th Conference on Emerging Technologies & Factory Automation (ETFA)*, 1–8 (2015).
- Rijswijk, K. *et al.* Digital transformation of agriculture and rural areas: A socio-cyber-physical system framework to support responsabilisation. *J. Rural Stud.* **85**, 79–90 (2021).
- Bechar, A. & Vigneault, C. Agricultural robots for field operations. Part 2: Operations and systems. *Biosyst. Eng.* **153**, 110–128 (2017).
- Bechar, A. & Vigneault, C. Agricultural robots for field operations: Concepts and components. *Biosyst. Eng.* **149**, 94–111 (2016).
- Feng, Q., Zou, W., Fan, P., Zhang, C. & Wang, X. Design and test of robotic harvesting system for cherry tomato. *Int. J. Agric. Biol. Eng.* **11**, 96–100 (2018).
- Scarfe, A. J., Flemmer, R. C., Bakker, H. & Flemmer, C. L. Development of an autonomous kiwifruit picking robot. In *2009 4th International Conference on Autonomous Robots and Agents*, 380–384 (2009).
- Søgaard, H. & Lund, I. Application accuracy of a machine vision-controlled robotic micro-dosing system. *Biosyst. Eng.* **96**, 315–322 (2007).
- Mavridou, E., Vrochidou, E., Papakostas, G. A., Pachidis, T. & Kaburlasos, V. G. Machine vision systems in precision agriculture for crop farming. *J. Imaging* **5**, 89 (2019).
- Xiong, Y., Ge, Y., Grimstad, L. & From, P. J. An autonomous strawberry-harvesting robot: Design, development, integration, and field evaluation. *J. Field Robot.* **37**, 202–224 (2020).

13. McCool, C. *et al.* Efficacy of mechanical weeding tools: A study into alternative weed management strategies enabled by robotics. *IEEE Robot. Autom. Lett.* **3**, 1184–1190 (2018).
14. Dunbabin, M. & Marques, L. Robots for environmental monitoring: Significant advancements and applications. *IEEE Robot. Autom. Mag.* **19**, 24–39 (2012).
15. Val, P. B., Valls, M. G. & Cuñado, M. B. A simple data-muling protocol. *IEEE Trans. Ind. Inform.* **10**, 895–902 (2013).
16. Atefi, A., Ge, Y., Pitla, S. & Schnable, J. In vivo human-like robotic phenotyping of leaf traits in maize and sorghum in greenhouse. *Comput. Electron. Agric.* **163**, 104854 (2019).
17. Drenjančević, M., Jukić, V., Zmaić, K., Kujundžić, T. & Rastija, V. Effects of early leaf removal on grape yield, chemical characteristics, and antioxidant activity of grape variety Cabernet Sauvignon and wine from eastern Croatia. *Acta Agric. Scand. Sect. B Soil Plant Sci.* **67**, 705–711 (2017).
18. Martin, D. *et al.* Grape cluster microclimate influences the aroma composition of Sauvignon blanc wine. *Food Chem.* **210**, 640–647 (2016).
19. Suresh Kumar, M. & Mohan, S. Selective fruit harvesting: Research, trends and developments towards fruit detection and localization—A review. *Proc. Inst. Mech. Eng. Part C J. Mech. Eng. Sci.* **237**(6), 09544062221128443 (2022).
20. Bai, Y., Mao, S., Zhou, J. & Zhang, B. Clustered tomato detection and picking point location using machine learning-aided image analysis for automatic robotic harvesting. *Precis. Agric.* **24**, 1–17 (2022).
21. Sepúlveda, D., Fernández, R., Navas, E., Armada, M. & González-De-Santos, P. Robotic aubergine harvesting using dual-arm manipulation. *IEEE Access* **8**, 121889–121904 (2020).
22. He, L., Zahid, A. & Mahmud, M. S. Robotic tree fruit harvesting: Status, challenges, and prosperities. In *Sensing, Data Managing, and Control Technologies for Agricultural Systems*, 299–332 (Springer, 2022).
23. Ghazal, S. *et al.* Analysis of visual features and classifiers for fruit classification problem. *Comput. Electron. Agric.* **187**, 106267 (2021).
24. Karkee, M., Zhang, Q. & Silwal, A. Agricultural robots for precision agricultural tasks in tree fruit orchards. In *Innovation in Agricultural Robotics for Precision Agriculture: A Roadmap for Integrating Robots in Precision Agriculture*, 63–89 (2021).
25. Altaheri, H. *et al.* Date fruit dataset for intelligent harvesting. *Data Brief* **26**, 104514 (2019).
26. Wan, S. & Goudos, S. Faster R-CNN for multi-class fruit detection using a robotic vision system. *Comput. Netw.* **168**, 107036 (2020).
27. Hu, C., Liu, X., Pan, Z. & Li, P. Automatic detection of single ripe tomato on plant combining faster R-CNN and intuitionistic fuzzy set. *IEEE Access* **7**, 154683–154696 (2019).
28. He, Y., Pan, F., Wang, B., Teng, Z. & Wu, J. Transfer learning based fruits image segmentation for fruit-picking robots. In *2020 IEEE 3rd International Conference on Computer and Communication Engineering Technology (CCET)*, 71–75 (2020).
29. Chen, C., Li, B., Liu, J., Bao, T. & Ren, N. Monocular positioning of sweet peppers: An instance segmentation approach for harvest robots. *Biosyst. Eng.* **196**, 15–28 (2020).
30. Barbole, D. K., Jadhav, P. M. & Patil, S. A review on fruit detection and segmentation techniques in agricultural field. In *Second International Conference on Image Processing and Capsule Networks: ICIPCN 2021*, vol. 2, 269–288 (2022).
31. Zhang, C., Zou, K. & Pan, Y. A method of apple image segmentation based on color-texture fusion feature and machine learning. *Agronomy* **10**, 972 (2020).
32. Zhou, Y. *et al.* Adaptive active positioning of camellia oleifera fruit picking points: Classical image processing and YOLOv7 fusion algorithm. *Appl. Sci.* **12**, 12959 (2022).
33. Magalhães, S. A. *et al.* Evaluating the single-shot multibox detector and YOLO deep learning models for the detection of tomatoes in a greenhouse. *Sensors* **21**, 3569 (2021).
34. Kuznetsova, A., Maleva, T. & Soloviev, V. Using YOLOv3 algorithm with pre-and post-processing for apple detection in fruit-harvesting robot. *Agronomy* **10**, 1016 (2020).
35. Ge, Y., Xiong, Y., Tenorio, G. L. & From, P. J. Fruit localization and environment perception for strawberry harvesting robots. *IEEE Access* **7**, 147642–147652 (2019).

### Author contributions

Data curation: K.M.A., S.S., H.M. Writing original draft, K.M.A., S.S., H.M. Supervision: K.M.A., S.S., H.M. Methodology: K.M.A., S.S., H.M. Visualization: R.H.J. Validation: R.H.J. Resources: K.M.A., S.S., H.M. Overall review and editing: S.S., R.H.J. All authors reviewed the final manuscript.

### Competing interests

The authors declare no competing interests.

### Additional information

**Correspondence** and requests for materials should be addressed to S.S.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024