

Development of Real-Time Crop Sorting System Based on Deep Learning and ESP32-CAM Integration

Zul Hasrizal Bohari^{1*}, Hazim Hafizullah Mohd Zaid¹, Mohamad Na'im Mohd Nasir¹,
Mohamad Fani Sulaima¹, Emy Zairah Ahmad¹, Ahmad Zaidi Abdullah², Muzamir Isa²

¹Faculty of Electrical Technology and Engineering, Universiti Teknikal Malaysia Melaka (UTeM), Malacca, Malaysia

²Faculty of Electrical Engineering & Technology, Universiti Malaysia Perlis (UniMAP), Perlis, Malaysia

Email: *zulhasrizal@utem.edu.my

How to cite this paper: Bohari, Z.H., Zaid, H.H.M., Nasir, M.N.M., Sulaima, M.F., Ahmad, E.Z., Abdullah, A.Z. and Isa, M. (2025) Development of Real-Time Crop Sorting System Based on Deep Learning and ESP32-CAM Integration. *Journal of Power and Energy Engineering*, **13**, 49-60.
<https://doi.org/10.4236/jpee.2025.138004>

Received: July 26, 2025

Accepted: August 18, 2025

Published: August 21, 2025

Copyright © 2025 by author(s) and Scientific Research Publishing Inc.
This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).
<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

This study presents the development of a real-time chili sorting system that integrates deep learning techniques with ESP32-CAM hardware to automate the grading of chili peppers based on ripeness and color. Manual chili grading, commonly used in agricultural practices, is often inefficient, inconsistent, and labour-intensive. To address these challenges, a YOLO-based object detection model was trained and deployed to classify chilies into red and green categories with high accuracy. The system incorporates an image acquisition setup using ESP32-CAM, a motorized conveyor system, and a microcontroller-based control mechanism for real-time operation. Experimental results demonstrate that the YOLO model can achieve detection accuracies exceeding 80%, effectively identifying and classifying chilies in dynamic environments. The proposed system enhances sorting precision, reduces processing time, and supports scalable implementation in agricultural supply chains. This research contributes to the advancement of smart agriculture by offering a low-cost, efficient, and scalable solution for post-harvest quality control.

Keywords

Image Processing, YOLO, Chili Grading, Smart Agriculture

1. Introduction

In recent years, the integration of artificial intelligence (AI) and machine vision into agricultural processes has gained significant traction, offering solutions to longstanding inefficiencies in post-harvest handling and quality assessment. One such application is the grading and classification of chili peppers, a crop that holds

substantial economic and cultural value across many regions. Traditional grading methods, which rely on manual inspection, are not only labour-intensive and time-consuming but also susceptible to inconsistencies and inaccuracies due to human fatigue and subjective judgment. These limitations can adversely affect product quality, market value, and food safety. To address these challenges, this study presents the development of an automated chili grading identification system utilizing image processing techniques and deep learning models, specifically the You Only Look Once (YOLO) object detection algorithm. This system is designed to accurately classify chilies based on critical attributes such as color, size, and ripeness, enabling consistent and high-speed sorting in real-time. Early tests of YOLO-based approaches have shown precision rates exceeding 80% in distinguishing red from green chilies, demonstrating its potential for reliable deployment in agricultural settings. Beyond technical efficiency, the impact of grading extends to broader social, economic, and environmental domains. Inconsistent grading practices can lead to food waste, financial losses, and public health risks due to the circulation of substandard produce. Furthermore, the inefficiencies of manual grading contribute to resource wastage and undermine efforts to meet the United Nations Sustainable Development Goals (UNSDGs) related to sustainable consumption and production. By automating the grading process, this project aims to enhance food safety, improve traceability, and support sustainable agricultural practices. The main objectives of this project are to build the application of a YOLO-based system for classifying and grading chilies by color and ripeness. Besides that, this project also develops a reliable and efficient grading system based on object detection techniques and this project evaluates the performance of a hardware prototype that integrates this system for real-time chili sorting. Ultimately, this research contributes to the advancement of precision agriculture by offering a scalable, cost-effective solution that supports both local farmers and larger agricultural stakeholders.

2. Related Work

2.1. You Only Look Once (YOLO)

The You Only Look Once (YOLO) object detection algorithm has been successfully applied in several recent studies to detect, classify, and grade chili peppers efficiently and accurately [1]-[3]. The deep learning-based real-time object detection system YOLO splits the image into grids and concurrently predicts bounding boxes and class probabilities for objects within these grids, allowing YOLO to achieve high detection speed while maintaining strong accuracy. This end-to-end approach makes YOLO incredibly fast when compared to traditional methods [3] [4].

For automated chili grading identification, YOLO has several benefits. These include robust performance in complex natural environments where chilies may be occluded or clustered, real-time detection with high speed and accuracy, and the capacity to detect and classify multiple chili fruits at once [2] [3]. Additionally,

it may be coupled with robotic harvesting and sorting systems to increase automation and is expandable to other chili kinds and grading standards [4]. Nevertheless, YOLO relies on sizable, properly labelled datasets for maximum accuracy and demands a significant amount of processing power, especially during training. Model tuning and hyper parameter optimization can be difficult and time-consuming processes, and implementing YOLO in real-time field applications may need sophisticated hardware, which raises the cost and complexity of the system as a whole [1]. **Table 1** shows the modification of object detection system (YOLO) version by version [4] [5].

Table 1. Modification of object detection system (YOLO).

Release	Author	Tasks	Paper
YOLO	Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi	Object Detection, Basic Classification	You Only Look Once: Unified, Real-Time Object Detection
YOLOv2	Joseph Redmon, Ali Farhadi	Object Detection, Improved Classification	YOLO9000: Better, Faster, Stronger
YOLOv3	Joseph Redmon, Ali Farhadi	Object Detection, Multi-scale Detection	YOLOv3: An Incremental Improvement
YOLOv4	Alexey Bochkovskiy, Chien-Yao Wang, Hong-Yuan Mark Liao	Object Detection, Basic Object Tracking	YOLOv4: Optimal Speed and Accuracy of Object Detection
YOLOv5	Ultralytics	Object Detection, Basic Instance Segmentation (custom)	NO

2.2. Convolutional Neural Networks (CNNs)

Convolutional Neural Networks (CNNs) provide a strong and efficient method for classifying devaluating chilies based on images. CNNs, a deep learning technique, are well-suited for identifying patterns like colour, texture, and form in chili photos since they are made to automatically and adaptively learn spatial hierarchies of features from images [3] [6].

CNN was employed in another research to estimate the ripeness degree of green chili fruits. After being gathered and pre-processed, the images were run through fully connected, max-pooling, and convolutional layers. With an accuracy of about 85%, the CNN model was able to classify chilies into maturity categories such as unmaturred, moderately matured, and matured. The study demonstrated CNN's ability to identify minute visual changes associated with chili maturity by using hyper parameter tweaking and the Adam optimizer to enhance performance [6].

In conclusion, CNNs' capacity to learn intricate visual cues straight from photos, adapt to change in illumination and form, and provide excellent classification and grading accuracy make them ideal for automated chili grade recognition. An efficient and scalable solution for chili grading and maturity estimate may be obtained for your project by deploying a CNN with the proper pre-processing, feature extraction, and training on a dataset of well-labelled chili images [3] [6].

2.3. Artificial Neural Networks (ANN)

Similar studies have made extensive use of Artificial Neural Networks (ANN) to automate and enhance the chili grading process. As an illustration, consider a project that used Artificial Neural Networks (ANN) to create a chili grading system. The project involved taking pictures of chili grading system. The project involved taking pictures of chilies and processing them to extract characteristics like colour, size, and skin texture. These characteristics were used to train the ANN to categorize chilies based on their grade, taking into account variables such as size, ripeness, and colour intensity. A MATLAB graphical user interface was used to create the system, enabling users to input chili photos and obtain automatic grading scores that were saved for further examination. This method sought to solve the inconsistencies and inefficiencies of human grading, which is frequently time-consuming, labour-intensive, and prone to human error.

In a different research, an ANN architecture was created especially for digital image analysis-based red chili pepper sorting and grading. In this study, colour and texture elements were extracted from photographs and utilized as input variables for the Artificial Neural Network (ANN). Five output cells representing various grades, 22 hidden layer cells, and three input cells (for specific attributes) made up the network's structure. With an accuracy of about 84.5%, the ANN was able to successfully classify chili samples. Similarly, additional work has utilized ANN to identify dried chili peppers based on size and colour, further illustrating the adaptability of this technique for diverse chili varieties and grading criteria [3] [7].

When opposed to manual approach, the primary benefit of adopting Artificial Neural Networks (ANN) for chili grading is their capacity to learn intricate, non-linear correlations between picture data and quality scores, which leads to increased accuracy and consistency. Large datasets may be handled by ANNs, which can also adjust to new grading criteria when fresh data becomes available. Nevertheless, drawbacks include the demand for a significant quantity of labelled training data, the need for computing resources for both inference and training, and the possibility in comparison to simpler models. Furthermore, the effectiveness of the picture capture and feature extraction procedures has a significant impact on the system's performance.

2.4. Naive Bayes Classifier

The Naive Bayes Classifier is a practical and effective method for classification

tasks. For classification problems, the Naïve Bayes Classifier is a useful and efficient technique. One pertinent example is a research that distinguished four varieties of red chili peppers like red chilies, caplak red chilies, curly red chilies. To make feature extraction easier, digital photos of chilies were gathered for this study and pre-processed by first turning them into grayscale and then binary pictures. The HSV (Hue, Saturation, Value) colour model, which is more resilient to changes in illumination than the conventional RGB model, was used to extract the important properties. These retrieved characteristics were then used to classify the chili photos using the Naïve Bayes Classifier. The classifier demonstrated a 92.5% accuracy rate using 119 training samples and 123 testing samples [7] [8].

Based on the Bayes theorem, the Naïve Bayes Classifier streamlines computation while maintaining high classification performance by assuming feature independence. Following the extraction of data such HSV colour component, the classifier determines the likelihood that a chili sample falls into each grade or kind and places it in the class with highest probability. Because of its ease of implementation and computational efficiency, this method may be used with real-time or almost real-time grading systems [9]. Accuracy may be impacted by colour channel correlation, therefore the assumption that characteristics are independent may not always hold true. Furthermore, the effectiveness of the classifier is highly dependent on the calibre of feature extraction. For automated chili grading identification, the combination of HSV-based feature extraction and the Naïve Bayes Classifier provides a good balance of accuracy, speed, and ease of use, making it a strong contender of your project, particularly when you want to efficiently and reliably classify chilies based on colour features [8].

2.5. K-Nearest Neighbour (KNN)

A popular and easy-to-use supervised machine learning technique for classification problems is K-Nearest Neighbour (KNN). KNN can categorize a chili sample according to the grades of its closest neighbours in the feature space when it comes to automated chili grading identification. First, key characteristics including colour, texture, shape, and size are extracted from chili photos. Each chili sample is represented by a vector made up of these characteristics.

The approach uses distance metrics such as the Manhattan or Euclidean distance to determine the distance between the feature vector of a fresh chili and every sample in the training dataset during classification. The chili is then assigned to the grade or kind that occurs most frequently among the k closest samples, neighbour, that were chosen [8].

One study, for instance, used Principal Component Analysis (PCA) in conjunction with KNN to enhance computing performance by reducing the dimensionality of the feature set by up to 95%. With this method, the machine achieved a 90% classification accuracy for chili kinds. This demonstrates that KNN may be a useful and precise technique for chili grading when combined with efficient feature extraction and dimensionality reduction [10].

KNN's simplicity, ease of use, and efficacy in multi-class classification tasks without assuming anything about the underlying data distribution are among its benefits. However, because KNN must compute the distance to every training sample, it might be computationally demanding during prediction. Accuracy may be lowered by its sensitivity to noise and extraneous characteristics [9]. Furthermore, for optimal performance, the distance measure and the number of neighbour (k) must be chosen carefully. Normalizing features is necessary to prevent bias in distance computations.

In conclusion, KNN is a good algorithm for your automated chili grade detection project, particularly when paired with PCA and other reliable feature extraction methods. It is useful for differentiating chili grades based on visual traits because it strikes a nice mix between accuracy and simplicity [8] [9].

2.6. Model Selection and Justification

In conclusion, the literature review shows that YOLO is the best option for automated chilli grading applications, even though Convolutional Neural Networks (CNNs) and Naive Bayes Classifiers have shown great promise in recognising visual characteristics and categorising chilli grades [1] [6] [8]. Although CNN-based models usually work in a two-stage method that might lengthen inference time, they provide excellent classification accuracy by learning fine features and adjusting to fluctuations in illumination and shape [3] [6]. Using extracted features, Naive Bayes provides effective classification with a comparatively high accuracy; nevertheless, its performance is highly reliant on human feature engineering and is constrained when dealing with overlapping objects or complicated visual situations [8]. On the other hand, real-time simultaneous detection and categorisation of many items is made possible by YOLO's unified architecture, which provides speed and resilience in a variety of crowded and natural contexts where chillies may be obscured or appear in clusters [1]-[3]. Its usefulness is further supported by its scalability, flexibility to various grading standards, and demonstrated integration with automated systems [2] [3]. Thus, YOLO is chosen as the favoured method above conventional CNNs and statistical classifiers due to the requirements of real-time processing, automation, and high detection reliability in the chilli grading context.

3. Methodology

3.1. Design and Construction

The design of the automated chili grading system centers on a motorized conveyor belt that transports chili peppers through an inspection area for real-time grading. The conveyor is constructed to provide a stable and uniform movement of chillies, ensuring consistent image capture conditions. The speed of the conveyor is adjustable to optimize the balance between throughput and image processing time. To maintain consistent lighting and reduce the influence of external environmental variations, an adjustable lamp system is installed above the conveyor. This

lighting setup allows control over brightness and minimizes shadows and reflections on the chili surface, which is critical for accurate color detection and grading.

For image acquisition, ESP32-CAM camera is mounted above the conveyor to capture clear images of the moving chili peppers under controlled lighting conditions. The camera is connected to a processing unit, such as a computer or embedded system, where the YOLO object detection model is implemented. YOLO's capability to detect and classify objects in real-time makes it ideal for this application, as it can identify chili fruits and classify their grade based on color features despite challenges like occlusion or varying sizes. The system design also includes synchronization between the conveyor movement, image capture, and processing to ensure that each chili is accurately detected and graded without overlap or missed detections. The ESP32-CAM typically uses the OV2640 camera sensor with a maximum resolution of 2 megapixels, which corresponds to about 1600×1200 pixels. The default image output size can be smaller (e.g., 600×800 pixels) but can be optimized programmatically up to the sensor's max resolution. Regarding conveyor speed when using the ESP32-CAM, there is no fixed standard, but considering the camera's limited frame rate (e.g., roughly up to 10 - 15 fps at lower resolutions), the conveyor speed should be set low enough to keep the object (e.g., chili) fully within the camera frame while it's captured. Typical practical conveyor speeds compatible with ESP32-CAM might be around 0.1 to 0.2 meters per second to ensure clear images without motion blur, but this should be experimentally adjusted based on your exact setup, lighting, and required detection accuracy.

The construction phase involves integrating the mechanical components (conveyor, lamp, camera) with the electronic control system and software. The conveyor motor and lamp brightness are controlled via microcontrollers or programmable logic controllers (PLCs) to allow automated adjustments during operation. The camera feeds images continuously to the processing unit running the YOLO model, which outputs chili detection and grading results. These results can then be used to trigger actuators or sorting mechanisms that physically separate chilies into different grades. This integration of hardware and software components aims to create a scalable, efficient, and accurate chili grading system that reduces manual labor and enhances grading consistency.

3.2. Software Flow Chart

The process begins with receiving the input image captured by the ESP32 camera positioned above the conveyor belt. This image contains one or more chili peppers that need to be detected and classified. Before feeding the image into the YOLO model, it undergoes preprocessing steps such as resizing to a fixed dimension and normalization to ensure consistent input quality and improve model performance.

Next, the preprocessed image is passed through the YOLO object detection model. YOLO divides the image into a grid and simultaneously predicts bounding boxes and class probabilities for each grid cell. This enables the model to localize

each chili pepper by drawing bounding boxes around them and classify their grade based on color features detected within those boxes.

After detection, post-processing techniques are applied to eliminate redundant overlapping bounding boxes, ensuring that each chili is represented by a single, most confident detection. Finally, the model outputs the detected chili locations along with their classified grades, which are then sent to the sorting mechanism or further processing units. The overall flow for the classification process is portrayed in **Figure 1**.

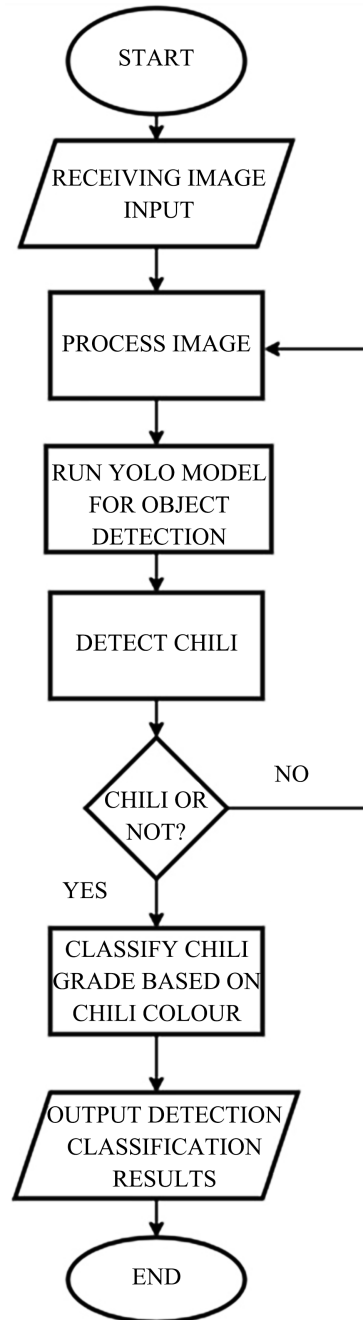


Figure 1. System block diagram for classification system.

3.3. Software Configuration

For the object detection experiment, a YOLO (You Only Look Once) deep learning model was implemented to enable real-time detection and grading of chili fruits. The specific YOLO version used was the Ultralytics YOLO framework, ensuring compatibility and support for state-of-the-art detection performance. All input images were resized to standard dimensions, such as 640×640 or 416×416 pixels, for both model training and inference, with the selection tailored to the hardware capabilities of the ESP32-CAM deployment platform. To maintain consistency and facilitate proper model learning, images were normalized by scaling pixel values to the 0 - 1 range (dividing each pixel value by 255), as required by the YOLO architecture. Bounding box annotations followed the YOLO-specific format of [center_x, center_y, width, height], with all coordinates normalized to the input image size, ensuring compatibility with the detection framework. For dataset preparation, training images were annotated using dedicated annotation tools such as LabelMe, and all images were acquired in the actual deployment environment to reflect operational conditions and enhance model robustness. This configuration ensures clear reproducibility and allows other researchers to replicate the experiment under similar conditions.

3.4. Dataset Configuration

A dataset created especially for the chilli grading assignment was assembled. To improve model robustness and operational relevance, all photos were taken in the deployment environment, accurately capturing the lighting and backdrop circumstances. The 30 photos in the collection cover a wide variety of chilli locations and orientations on the conveyor belt. In order to ensure a fair balance for graded classification, the dataset's class distribution comprises 30 photos of red chillies and 30 images of green chillies.

The LabelMe tool, which allowed for accurate bounding box labelling of each chilli instance within the photos, was used to carry out the annotation procedure. The YOLO-specific bounding box format, which is expressed as [center_x, center_y, width, height], was used in the annotations. All values were normalised in relation to the dimensions of the input picture. In order to represent real-world detection issues, the annotation standards mandated that every visible chilli in a picture be tagged, despite of location or partial occlusion. To guarantee label accuracy and consistency throughout the dataset, an experienced supervisor then went over and made corrections to the annotated photos. High detection accuracy and repeatability in ensuing tests are ensured by this meticulous approach to dataset preparation.

4. Result

4.1. Image That ESP32 CAM Captured

In a robust AI-powered chili detection and classification system, the crucial initial step involves data preparation and annotation, which is where a tool like LabelMe

becomes indispensable. While not explicitly depicted in the provided images, the process for training an object detection model, such as YOLO, necessitates meticulously labelling objects within images. This entails drawing precise bounding boxes or polygons around each chili pepper in the raw visual data (like that shown in figure).

In **Figure 2**, a green bounding box precisely encloses a green chili, accompanied by a clear “Green Chili” label in corresponding green text, visually confirming the model’s accurate classification of an unripe or mature green chili. Similarly, **Figure 3** illustrates a red chili, highlighted by a red bounding box and a “Red Chili” label, demonstrating the model’s capability to correctly identify a fully ripe chili.

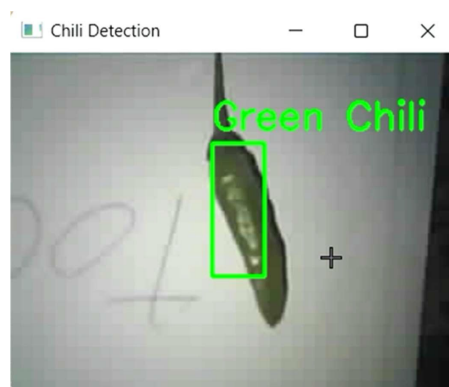


Figure 2. Green chili classification.



Figure 3. Red chili classification.

4.2. Data from Detection

The provided image displays the output from a Python program running within an IDLE Shell, specifically demonstrating the real-time performance of an object detection model, likely a YOLO (You Only Look Once) variant. The repeated “No chili detected” messages suggest that the model is continuously processing frames or images, and for many of these, it doesn’t identify any objects that it’s trained to recognize as chilies, or the chilies present do not meet a certain confidence threshold for detection. This is a common occurrence in real-time detection scenarios where the object of interest may only appear intermittently or be obscured.

Malaysia Melaka (UTeM) for the financial support and provision of research facilities that made this publication possible.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

- [1] Wang, Y., Ouyang, C., Peng, H., Deng, J., Yang, L., Chen, H., *et al.* (2025) YOLO-ALW: An Enhanced High-Precision Model for Chili Maturity Detection. *Sensors*, **25**, Article 1405. <https://doi.org/10.3390/s25051405>
- [2] Abubeker, K.M., Abhijit, Akhil, S., Akshat Kumar, V.K. and Jose, B.K. (2023) Computer Vision Assisted Real-Time Bird Eye Chili Classification Using YOLO V5 Framework. *Journal of Artificial Intelligence and Technology*, **4**, 265-271. <https://doi.org/10.37965/jait.2023.0251>
- [3] Salim, R. and Fajar, A.N. (2024) Object Detection of Chili Using Convolutional Neural Network YOLOV7. *Journal of Theoretical and Applied Information Technology*, **102**, 2419-2427. <https://www.jatit.org>
- [4] Warni, E., Indrabayu, Achmad, A. and Syahsir, A.R.R. (2025) Harnessing YOLO for Loose Fruits Detection: Boosting Productivity in Palm Oil Plantations. 2025 *International Conference on Advancement in Data Science, E-Learning and Information System (ICADEIS)*, Bandung, 3-4 February 2025, 1-6. <https://doi.org/10.1109/icadeis65852.2025.10933352>
- [5] Brucal, S.G.E., de Jesus, L.C.M. and Samaniego, L.A. (2024) Development of a Localized Tomato Leaf Disease Detection Using YoloV9 Model via RoboFlow 3.0. 2024 *IEEE 13th Global Conference on Consumer Electronics (GCCE)*, Kitakyushu, 29 October-1 November 2024, 601-603. <https://doi.org/10.1109/gcce62371.2024.10760343>
- [6] Zainudin, M.N.S. (2021) A Framework for Chili Fruits Maturity Estimation Using Deep Convolutional Neural Network. *Przegląd Elektrotechniczny*, **1**, 79-83. <https://doi.org/10.15199/48.2021.12.13>
- [7] Moya, V., Quito, A., Pilco, A., Vásquez, J.P. and Vargas, C. (2024) Crop Detection and Maturity Classification Using a YOLOv5-Based Image Analysis. *Emerging Science Journal*, **8**, 496-512. <https://doi.org/10.28991/esj-2024-08-02-08>
- [8] Krisna, D.A.N. and Salamah, U. (2022) Perbandingan Algoritma Naïve Bayes Dan K Nearest Neighbor Untuk Klasifikasi Berita Hoax Kesehatan Di Media Sosial Twitter. *Jurnal Teknik Informatika Kaputama (JTIK)*, **6**, 1-115.
- [9] Roujip, R.S. (2022) Chili Grading System Using Ann Approach. Universiti Malaysia Sabah. <https://eprints.ums.edu.my/id/eprint/33199/>
- [10] Julianda, R., Tundo, and Sugeng, (2025) Chili Type Detection System Using Principal Component Analysis Method. *International Journal Software Engineering and Computer Science (IJSECS)*, **5**, 102-112. <https://doi.org/10.35870/ijsecs.v5i1.3735>