International Journal of Autonomous Robotics and Intelligent Systems
Vol. 1, No. 1, June 2025 [107-115]

Real-Time Classification of Chilli Ripeness using Convolutional Neural Network (CNN)

M. Alif Zauhaili[1], N.S Khalid[1,2,3*], Ismail I. Ibrahim[1,2], MZ Hasan[1,2], H. Mansor[1,2], NDN Dalila[4], MH Muhammad Hafiz[4] and S.A. Abdul Shukor[1,2]
[1]Faculty of Electrical Engineering and Technology, Universiti Malaysia Perlis, Perlis, Malaysia
[2]Centre of Excellence for Intelligent Robotics & Autonomous Systems, Universiti Malaysia Perlis, Perlis, Malaysia
[3]Faculty of Engineering and Architecture, Ghent University, Belgium
[4]Horticultural Research Centre, Malaysian Agricultural Research and Development Institute (MARDI), Kedah, Malaysia

Received 7 July 2025, Revised 17 July 2025, Accepted 18 July 2025

ABSTRACT

Chilli harvesting plays an important role in Malaysia's economy as it is one of the crops with high demand in the country. Normally, farmers harvest and categorise the ripeness of chillies by using the naked eye which can lead to errors and human fatigue. To overcome the limitations of this manual harvesting, an automated real-time chilli vision system that can classify between ripe and unripe chillies were developed. This research involved with a diverse dataset of chilli images using various chilli varieties and growth stages. The YOLOv8 model was trained using Google Colab's GPU-accelerated environment to optimize the performance. The model's deployment for real-time inference and classification was facilitated through Visual Studio Code, with HSV colour analysis used to differentiate between ripe and unripe chillies. CNN was used to validate and analyse the accuracy of the proposed system. As a result, the system achieved an accuracy of 88% for chilli classification. These findings proved the potential of Artificial Intelligence (AI)-driven systems in supporting precision agriculture.

Keywords: Chillies, Convolutional Neural Network (CNN), Real-time, Ripeness, YOLOv8

1. INTRODUCTION

In Malaysia, chilli harvesting plays an important role because it is one of the crops with high demand [1]. According to the 2018 statistics from the Agriculture Department of Malaysia, the annual domestic demand for chillies stood at 55,420 tonnes, while the domestic production lagged behind at 24,428 tonnes per year. A portion of the produced chillies, amounting to 3,863 tonnes, was exported, leaving a shortfall in domestic supply and necessitating an additional 34,855 tonnes from external sources to meet the overall demand. This proved that the industry and demand of chilli in Malaysia plays an important role of agricultural needs. Unfortunately, chilli harvesting in Malaysia is done manually which can lead to human error and fatigue among the workers.

The current chillies ripeness classification method relies on hands-on techniques where normally farmers will be using their vision and touch, to determine the ripeness of chillies based on their experiences or predefined schedules. These methods involve with a lot of subjective judgments and may lead to inconsistencies in identifying ripe and unripe chillies. Figure 1 shows the traditional method for harvesting chillies.

* syahirahkhalid@unimap.edu.my

Figure 1. Traditional chillies harvesting [2]

Despite their widespread use, these manual methods are time-consuming and may lack of accuracy. This problem highlights the need for a more robust system that can reliably differentiate between chillies and the leaves, ensuring precise classification and reducing errors. The integration of Artificial Intelligence (AI) method and machine vision can increase the accuracy and efficiency of chilli classification during harvesting. To address these issues, an AI-based solution is required to enhance the accuracy of ripeness chilli sorting. The system should be capable of accurately detecting and classifying chillies based on their ripeness to improve overall system robustness.

1.1 Computer Vision in Precision Agriculture

Computer vision is mainly use to detect and classify the condition of the chilli in data collecting process. Computer vision usually works with a sensor such as a camera. From the data, suitable microcontroller will process the data taken in real-time and in split second will make the identification of the chilli and can do the classification as well. It is a critical step in the machine vision workflow because an inaccurate image can disrupt the entire process [3]. Figure 2 shows the example of image acquisition of chillies.

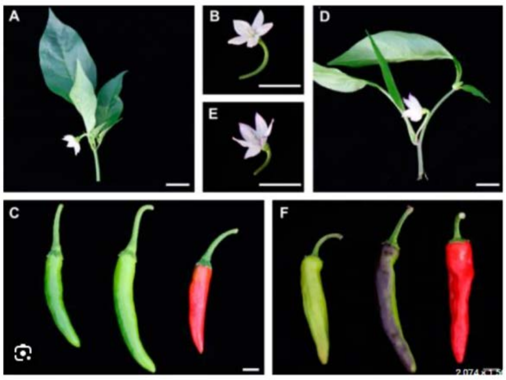

Figure 2. Example of image acquisition of chillies [4]

Precision agriculture helps farmers use their tools and materials more wisely by focusing on certain areas [5]. The technology enables the extraction of valuable insights from visual data, aiding farmers in making informed decisions about their crops. Investigating the applications of computer vision in crop monitoring provides a good foundation for understanding its potential in enhancing agricultural area. The use of digital image processing methods for simulating the visual capability of the human being has proven to be a dynamic feature in smart or precision agriculture [6]. This concept has provided with the automatic preventing and monitoring of plants, cultivation, disease management and water management, to increase the crop productivity and quality.

1.2 CNN for Data Processing

One of the Artificial Neural Network (ANN) technique that is particularly good at processing and recognising images is Convolutional Neural Network (CNN). The term "convolutional" refers to the fact that they use the mathematical operation of convolution to analyse input data. CNN consist of several "hidden" layers composed of interconnected layers of neurons. Before being sent to the output layer, the input layer receives the raw data and uses the hidden layers to process and transform it.

Convolutional and pooling layers are often combined to form CNN's hidden layers. Using a set of learnable filters, the convolutional layers take features out of the input data, like edges, corners and patterns. The pooling layers compress the data by providing an overview of each region's contents. CNN is widely employed in tasks related to object detection, image generation and image classification. The architecture of the CNN is shown in Figure 3.

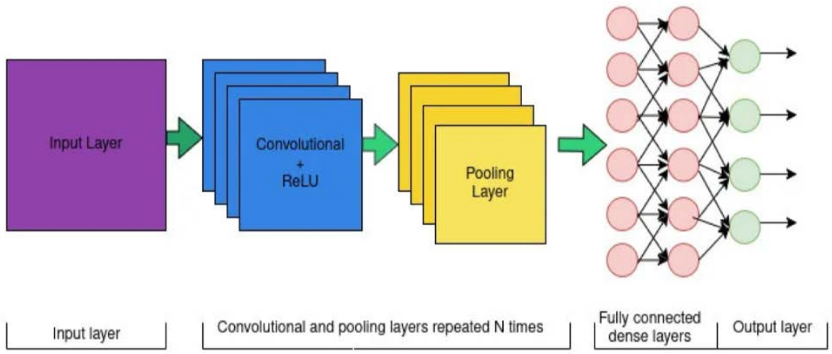

Figure 3. The general structure of CNN [7]

2. MATERIAL AND METHODS

This part will explain the details of the method used to develop an image recognition system to differentiate ripeness of chillies. By integrating deep learning method, the image processing algorithms aims to deliver a comprehensive solution for precise ripeness chillies classification.

2.1 Overall Methodology

The first step is to gather the related data from open-source repositories focusing on images of ripe (in red) and unripe (in green) chillies for training and testing the AI model. Visual Studio Code (VS Code) is then configured as the primary integrated development environment (IDE) for coding, offering robust debugging and version control features. Next, a pre-trained Convolutional Neural Network (CNN) model is trained using the collected dataset on Google Colab, which provides the computational power necessary for intensive training processes. Semantic segmentation techniques are applied to the trained model to enable accurate colour-based classification of chillies, distinguishing between ripe and unripe ones.

The classification code is developed and refined in VS Code to implement the trained model and integrate it into the system architecture. A web camera is then utilised to capture real-time images of chillies, for evaluating the system's performance in a dynamic environment. A flowchart is provided as in Figure 4.

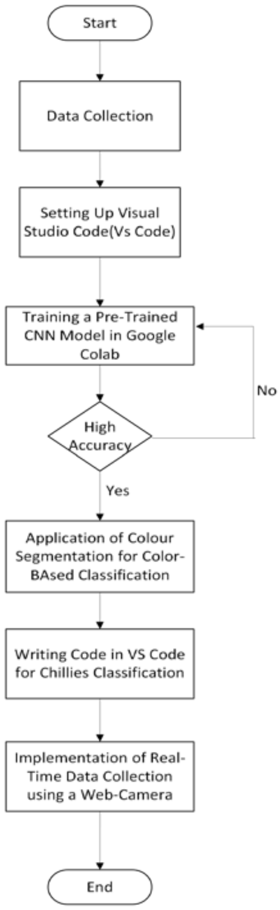

Figure 4. Flowchart of the methodology

2.2 Training a Pre-trained CNN Model in Google Colab

A pre-trained CNN model is fine-tuned in Colab to recognise and classify chillies based on images from the collected dataset. Colab is chosen for its access to powerful GPU resources and ease of collaboration, facilitating efficient model training and experimentation. The fine-tuning process involves preprocessing the images, including resizing and normalization, to prepare them for training. Transfer learning techniques are employed to leverage the knowledge learned by the pre-trained CNN model. The model architecture, such as ResNet or Inception, is adapted to the specific requirements of chillies classification. Hyperparameters are tuned iteratively to optimize model performance, guided by validation metrics such as accuracy and loss.

Once trained, the CNN model undergoes rigorous evaluation using a validation set to assess its ability to distinguish between chillies and leaves accurately. The model's performance is analysed, and adjustments are made as necessary to enhance its accuracy and generalization capabilities. The training process ensures that the CNN model is robust and capable of accurately identifying chillies in various real-world scenarios.

2.3 Implementation of Real-Time Data Collection using a Web Camera

The system captures video streams from the web camera and applies the trained model for real-time chilli classification. The implementation includes configuring the web camera settings, integrating it with the developed code in VS Code, and optimizing performance for real-time inference. The web camera setup ensures that high-quality images are captured and processed by the integrated model in real-time. The system continuously analyses incoming video frames, detects chillies using YOLOv8, and classifies them into ripe and unripe categories based on colour segmentation. Performance metrics such as frame rate, processing time per frame, and classification accuracy are monitored to evaluate the system's efficiency and reliability.

Real-world testing is conducted to validate the system's performance across different environmental conditions and operational scenarios. Feedback from testing is used to fine-tune parameters, optimize algorithms and enhance overall system performance. The iterative development approach ensures that the deployed system meets user requirements and achieves the desired level of accuracy and robustness. Figure 5 shows the implementation of real-time data collection using a web camera.

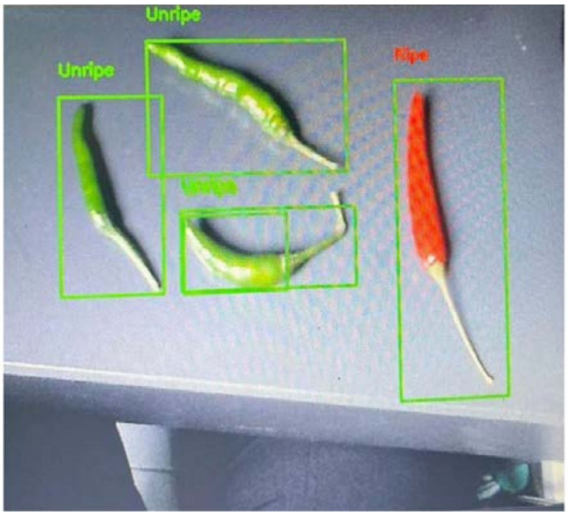

Figure 5. Sample of real-time data collection using a web camera

3. RESULTS AND DISCUSSION

3.1 CNN Evaluation

The evaluation of CNN model for chillies analysis was conducted over 50 epochs, with key performance metrics and loss values tracked throughout the training process. Figure 6 shows the results of chillies CNN train model.

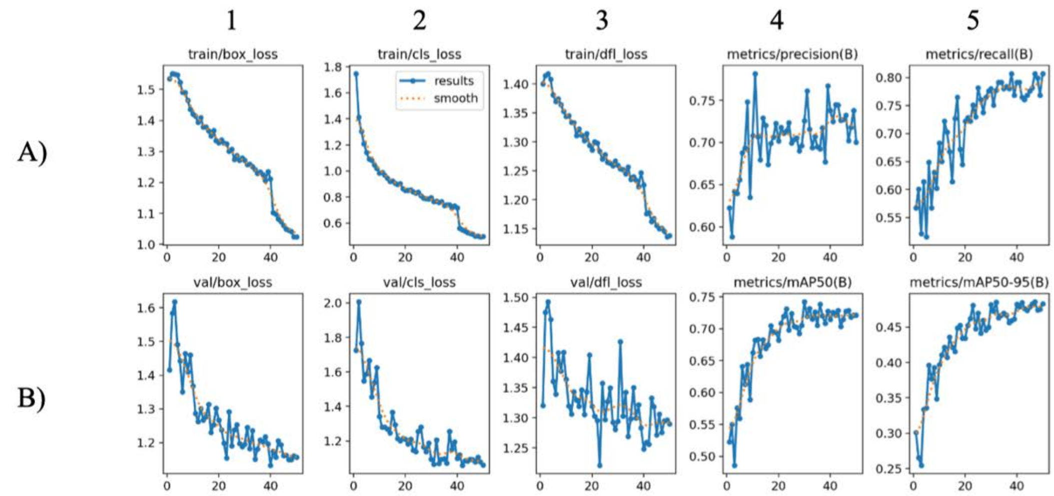

Figure 6. Results of chillies CNN train model

3.1.1 Training and Validation Losses of Chillies Classification

The first row of graphs in the Figure 6 depicts the training losses (1A-3A): Box Loss (1A), Class Loss (2A) and DFL Loss (3A). Each of these losses shows a consistent decrease over the epochs. This downward trend indicates that the model effectively learned and improved its predictions on the training data. The Box Loss (1A), started around 1.5 and steadily decreased to approximately nearly 1.0 by the 50th epoch. The Class Loss (2A), began near 1.6 and reduced to about 0.6, while the DFL Loss (3A), dropped from 1.4 to nearly 1.15.

The second row of graphs shows the corresponding validation losses (1B-3B). Similar to the training losses, the validation losses also demonstrated a downward trend, suggesting good generalization to unseen data. The Box Loss (1B), for validation started at around 1.6 and decreased to 1.2. The Class Loss (2B), dropped from approximately 2.0 to 1.2, and the DFL Loss (3B), fell from 1.5 to 1.25 by the end of the 50 epochs.

3.1.2 Performance Metrics of Chillies Classification

Performance metrics such as precision, recall, and mean Average Precision (mAP) were tracked throughout the training process. The precision metric (4A), as shown in the top row of Figure 6, fluctuated but shows an overall improvement, ending at around 0.75. The recall metric (5A), exhibited a similar pattern, starting from about 0.55 and increased to 0.80, indicating better identification of actual positives over time.

The mAP metrics, both at 50% IoU (mAP50) at (4B) and the average across IoU thresholds from 50% to 95% (mAP50-95) at (5B), shows significant improvements. The mAP50 (4B), metric improves from approximately 0.50 to 0.75, signifying increased accuracy in object detection at this threshold. The mAP50-95 (5B), metric rose from about 0.25 to 0.45, demonstrating the model's robustness across different IoU thresholds.

3.1.3 Learning Rate Adjustments of Chillies Classification

Throughout the training period, learning rates for all parameter groups were gradually reduced. This reduction played a crucial role in fine-tuning the model and achieving minimal loss, thereby enhancing overall performance. The gradual decrease in learning rates proved effective in optimizing the training process.

In conclusion, the CNN model exhibited consistent improvements in both training and validation performance metrics over 50 epochs. The observed reduction in losses and increase in precision, recall and mean Average Precision metrics indicate that the model is effectively learning and generalizing from the dataset.

3.2 Chillies Detection

The confusion matrix analysis reveals important performance metrics of the model's ability to classify chillies versus background in the dataset. From the matrix, it is observed that the model correctly identifies chillies (True Positive Rate) with an accuracy of 88%, indicating a high sensitivity in detecting chillies when they are present. However, there is a 12% false positive rate, meaning that in some instances, the model incorrectly identifies background as chillies. On the other hand, the model shows a 0% true negative rate, suggesting that it does not correctly identify background when it is present. Additionally, the false negative rate also stands at 0%, indicating that the model never misses a chilli when it is actually present.

This evaluation underscores the model's strength in detecting chillies but highlights a significant limitation in its ability to distinguish background accurately. Addressing the false positive and true negative rates could potentially improve the model's overall performance and reliability in real-world applications. These insights from the confusion matrix analysis provide valuable guidance for further refining the model's training strategy and optimizing its capabilities for enhanced precision in agricultural monitoring and crop classification tasks. Figure 7 shows the normalize confusion matrix of the chillies and Figure 8 shows the results of the trained data in Colab, in which the detected objects are being labelled as 0, which means that the detected objects are being classified as the class with index 0 as the first class listed in dataset's class file.

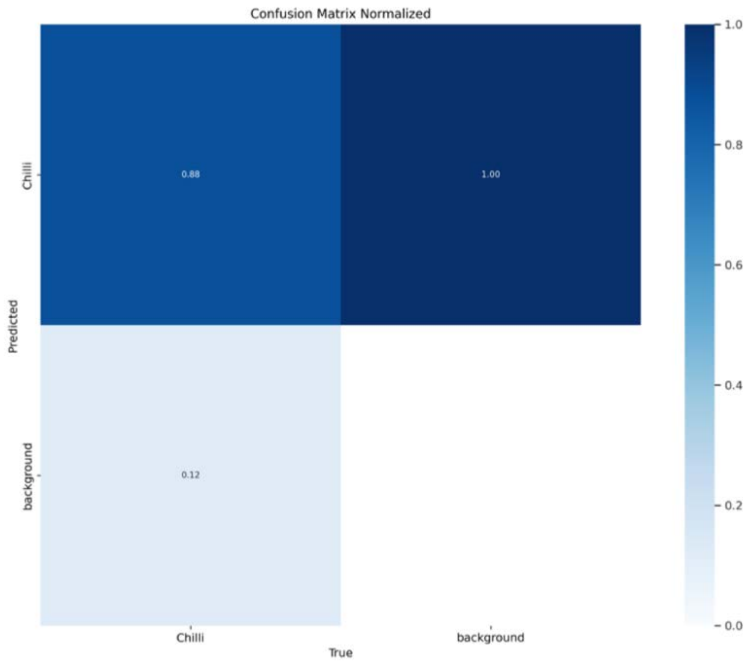

Figure 7. The normalize confusion matrix

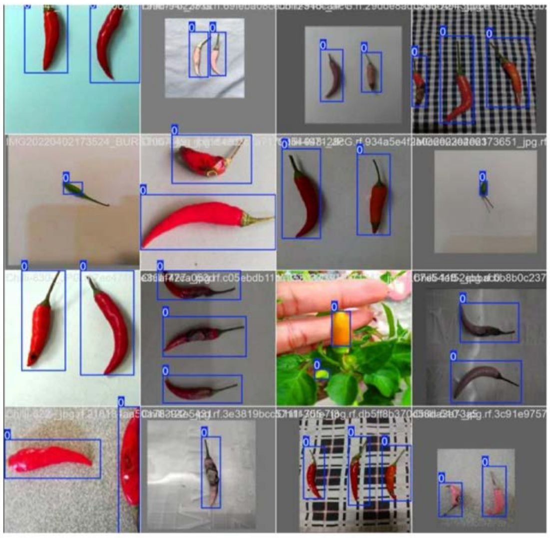

Figure 8. Results of the train data in Google Colab

3.3 Real-Time Results of Ripe and Unripe Chillies

The system successfully identifies the chillies and distinguishes between ripe and unripe ones. The implementation utilizes the trained YOLOv8 models to detect chillies in real-time. After identifying the chillies, the system analyses the colour of each detected chillies using a predefined HSV colour range to determine its ripeness. This allows the system to accurately classify chillies as either ripe or unripe, providing reliable and actionable results in real-time. The combination of advanced object detection and colour analysis ensures that the system performs with high precision and efficiency, even in complex environments. Figure 9 shows the real-time result of the system in distinguishing ripe and unripe chillies.

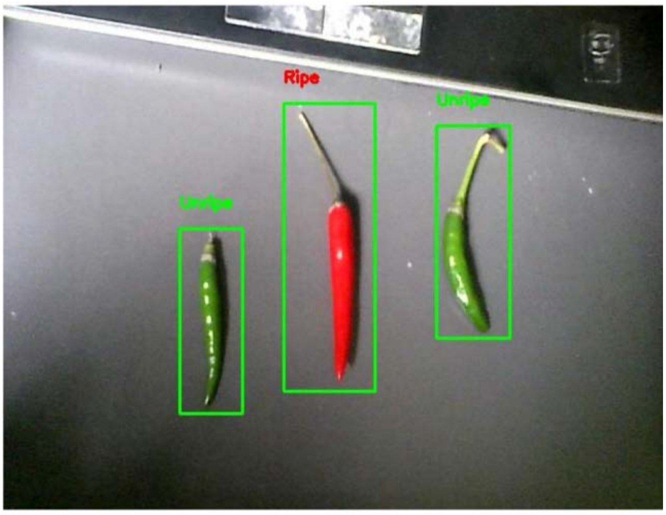

Figure 9. Real-time results of the system in distinguishing ripe and unripe chilli

4. CONCLUSION

This research demonstrates the feasibility and effectiveness of an AI method specifically for real-time chilli ripeness classification system utilizing the YOLOv8 object detection model. The integration of CNN for validation further strengthened the system's accuracy and reliability. This system offers a promising solution for enhancing productivity and decision-making in chilli farming, particularly within the Malaysian agricultural context. By automating the chilli sorting process, this system could reduce human error and fatigue among of workers.

REFERENCES

[1] "Young farmers can reduce Malaysia's dependency on imported chillies, says deputy minister," [Online]. Available: https://www.malaymail.com/news/malaysia/2020/06/03/young-farmers-can reduce-malaysias-dependency-on-imported-chillies-says-dep/1872186 [Accessed Jan. 16, 2024].

[2] "How to harvest chillies & storage tips," [Online]. Available: https://plantura.garden/uk/vegetables/chillies/how-to-harvest-chillies [Accessed Jan. 16, 2024].

[3] V. K. Mishra, S. Kumar, N. Shukla, 2017. Image acquisition and techniques to perform image acquisition. SAMRIDDHI: A Journal of Physical Sciences, Engineering and Technology, 9(01), 21-24.

[4] S.H. Lim, D.H. Kim, M.C. Cho, J.Y. Lee, 2022. Chili Pepper AN2 (CaAN2): A Visible Selection Marker for Nondestructive Monitoring of Transgenic Plants. Plants, vol. 11, no. 6, p. 820, doi: https://doi.org/10.3390/plants11060820.

[5] J.I. Burns, 2014. Agricultural Crop Monitoring with Computer Vision. Doctoral dissertation, Virginia Tech.

[6] S. S. Chouhan, U. P. Singh, S. Jain, 2020. Applications of Computer Vision in Plant Pathology: A Survey. Archives of Computational Methods in Engineering, 27, 611-632.

[7] L. Alzubaidi et al., 2021. Review of Deep Learning: Concepts, CNN Architectures, Challenges, Applications, Future Directions. Journal of Big Data, vol. 8, no. 1, doi: 10.1186/s40537-021-00444-8.