Journal of Information Systems Engineering and Management
2025, 10(56s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Computer Vision-based Detection Models for Classifying Ripening Stages of Tomato Fruits

Rotimi-Williams Bello, Pius A. Owolawi, Etienne A. van Wyk, Chunling Tu
Department of Computer Systems Engineering, Faculty of Information and Communication Technology, Tshwane University of Technology, South Africa; email: bellorw@tut.ac.za

ARTICLE INFO

Received: 24 Dec 2024

Revised: 12 Feb 2025

Accepted: 26 Feb 2025

ABSTRACT

Introduction: Tomato fruits colour is a crucial aspect of tomato fruits cultivation as it guides farmers through the ripening stages. Accurate identification of the ripening stages of tomato fruits using their colours facilitates assessment of the level of their ripeness for timely harvesting, thereby improving quality control in horticulture. Four key ripening stages are involved in tomato fruits, namely green, turning, pink, and red ripe. These stages require constant monitoring to minimize waste and loss. However, little has been done to ensure constant monitoring of the ripening stages of this fruit, and most of the current methods employed for this task are uneconomical, labour intensive, time-consuming, and inaccurate.

Objectives: To address the abovementioned challenges and monitor the ripening progression of the tomato fruits.

Methods: YOLOv4 and ResNet50 were proposed for detecting and classifying their different colour stages. A dataset from Roboflow, which comprises 1050 images were collected and categorized into four classes of green, turning, pink, and red ripe, then augmented and employed in training the models on Google Colab, leveraging its cloud-based resources to efficiently manage the training process.

Results: The results obtained from the two models were compared to SSD MobileNet v2, and a proposed colour space model of HSV based on Histogram. In overall, YOLOv4 performed better in detection and worse in classification than ResNet50, however, the performance of the colour space model of HSV was better in classification than the models of YOLOv4 and ResNet50.

Conclusions: Comparing the performance of the models with each other in detection of tomato fruits, the confidence in the predictions of both models were close and similar, this simply means that at close and similar confidence thresholds of 57% and 59%, respectively, both models got their best F1-score and mAP; however, YOLOv4 model achieved a better F1-score of 89.50% and mAP of 83.20% than the F1-score of 77.01% and mAP of 71.01% achieved by the ResNet50 model. The colour space model of HSV for Green ripening tomato fruits obtained the highest percentage in all the evaluation metric for the classification with 97.45% precision, 98.11% recall, and 97.78% F1-score.

Keywords: Colour, HSV, Model, ResNet50, Tomato fruit, YOLOv4

1. INTRODUCTION

Farming of tomato fruits plays an important role in the agricultural sector of South Africa, thus contributing to both subsistent farming and economy [1]. The annual tomato fruits production in South Africa is around 600, 000 tonnes due to the important role they play; this is evident in the high rate of their consumption and the high rate at which they were produced worldwide annually in their million tonnes a few years back [2]. Three main types of tomato fruits are grown in South Africa, namely cherry tomato fruits, roma tomato fruits destined for processing, and fresh or round tomato fruits. Most of these tomato fruits are cultivated in open fields under the aid of irrigation. Several South Africa provinces grow tomato fruits, including Northwest, Eastern Cape, Western Cape, Limpopo, and

Journal of Information Systems Engineering and Management
2025, 10(56s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

Mpumalanga with Limpopo being one of the largest producers [3]. Tomato fruits ripening process is closely associated with colour change in the three pericarps; the exocarp (outermost layer), the mesocarp (middle layer), and the endocarp (innermost layer), which is one of the bases of tomato fruits life cycle [4]. Colour change is an essential quality attribute for monitoring and assessing tomato fruits' growth and ripening behaviour [5]-[6]. The colour of tomato fruits on the exocarp (outermost layer) has been used as a basis for determining their maturity (ripeness), harvest and consumption [4]. Although colour, quality and maturity (ripeness) of tomato fruits may not matter at the level of subsistent horticulture, there is great preference attached to their colour, quality and maturity (ripeness) at commercial level. The worldwide accepted methods for identifying and classifying ripening stages of tomato fruits are the methods approved by the United Fresh Fruit and Vegetable Association (UFFVA) and U.S. Department of Agriculture (USDA) colour chart [5].

This is also synonymous to the specified evaluation methods used in determining the progressive maturity (ripeness) of tomato fruits as promulgated by International Commission on Illumination (CIE) [7], among which are hue colour (HC - instrument based colorimetric measurement), ripening index (RI - based on visual colour), and red tomato (RT). The RI-based method, being a simple and rapid assessment method has been widely accepted by individual farmers for evaluating the degree of tomato fruits ripening. Furthermore, the amount in number of RT during harvest and storage period can also be used as an indicator of the maturity (ripeness) of the tomato fruits [8]. The duo of RI and RT as index-parameters for evaluating the degree of ripening of tomato fruits can be useful indicators for informing about the rate of ripening if carried out at regular time intervals.

Colour of tomato fruits has a lot to do with the manual method of harvesting; the manual method of harvesting is one of the factors that cause low labour productivity because of rate at which the method consumes both time and energy of the human harvester [9]-[10], thereby leading to loss of interest on the part of labour force. The economic and subsistent importance of tomato fruits in addition to the related costs of their production justify any painstaking research carried out to improve on their production and quality [10]. However, monitoring the degree of tomato fruits ripening for harvesting using their colours is a herculean task. The performance of any tomato fruits ripening detection models is evaluated based on their detection and manipulation strength in a heterogeneous environment especially when the target tomato fruits are of different reflectivity, postures, colours, and dimensions [11].

Tomato fruits are either green, turning, pink, or red in colour to ripe for harvest [6], and this colour plays an essential role in tomato fruits' features extraction and classification [12], which are important techniques for achieving a distinguished crop harvest. Varied lighting conditions and occlusions of some tomato fruits by the plant's body parts are major barriers that shield the tomato fruits from being visible and accessed by most of the tomato fruits ripening detection models. Less visible tomato fruits should not be an issue to a highly performing tomato fruits ripening detection model for a robust and damage-free harvest. Computer vision-based models are designed for different vision-based tasks and thus must possess all the phases to acquire, segment and classify image. Image segmentation refers to a computer vision technique that partitions a digital image into discrete groups of pixels—image segments—to inform object detection and related tasks [13].

YOLO (You Only Look Once) and ResNet50 are detection models developed by using the framework of CNNs (Convolutional Neural Networks) due to its performance in executing complex real-time tasks with strong learning capacity within an ample time [14]-[15]. Models such as YOLO and ResNet50 though efficient, their employment in detection and classification of tomato fruits ripening is still very low [16]. There is a need for more research and development of computer vision systems for tomato fruits horticulture and their management [17]. This is in line with what is proposed in this paper which focuses on the computer vision of the colour-based detection models for classifying colour stages of tomato fruits by training, testing and evaluating the computer vision models and the proposed colour space model of HSV on the dataset for classification comparison.

In this paper, the ripening process of four classes of tomato fruits (green, turning, pink, and red ripe) was monitored by detecting changes in their colours, thereby facilitating the assessment of the level of their ripeness for timely harvesting. This study aims to detect and classify tomato fruits by employing colour-based detection models (computer vision models and the proposed colour space model of HSV). The work in this paper is a step towards classifying colour stages of tomato fruits using colour-based detection model and thus improving quality control in horticulture. The contributions of the paper are as follows:

(1) Integration of YOLOv4 and ResNet50 with VEED for detection and classification of ripening process of four classes of tomato fruits.

(2) Utilization of HSV for ripening classification by separating colour (Hue), purity/intensity (Saturation), and brightness/intensity (Value), thereby facilitating the detection of subtle colour changes that occur as tomato fruits ripen.

(3) Augmentation of Roboflow tomato fruit datasets to bolster training of the proposed models.

2. RELATED WORK

2.1. Tomato fruits detection and classification by colour-based models

The state-of-the-art models for detecting and classifying tomato fruits in images using their colour feature analysis and computer vision approaches are the main focus of this section. Image segmentation accuracy and performance depend on the image colour as colour is what feature extraction technique uses in extracting features from an image, particularly when segmentation of foreground (ripe tomato fruits) from the background (heterogenous environment) matters. However, the different conditions (such as occlusions or variation in illuminations) under which colour segmentation is performed may limit its performance. To extract colour features of target object from image, there is a need to use colour spaces such as HSV (Hue, Saturation, Value), HIS (Hue, Intensity, Saturation), RGB (Red, Green, Blue), etc.

Multiple colour spaces are sometimes employed in developing tomato fruits harvest-aided model for a particular purpose, in which the techniques may involve identifying the tomato fruits in the image using colour space transformation from RGB to HIS [18]. In an effort to devise a better recognition method for ripe tomato fruits that could subdue occlusion and related mitigating factors, an algorithm was proposed by Rong et al. [19] for extraction and segmentation of colour information from an image using algorithms of different thresholds, morphological features and hybrid of YIQ (where Y represents luminance or brightness, I and Q represent chrominance or colour), RGB, and HSI colour spaces. Sharma et al. [5] proposed colour-based ripening index (RI %) and red tomato (RT %) parameters for assessing tomato fruits ripening progress and behaviour during storage.

Comparing the two parameters, they found RI (%) as a better assessment method for tomato fruits ripening than RT (%) due to its sensitivity (2.5×) to dynamic nature of the ripening, thus concluding its suitability for the progress assessment of tomato fruits ripening. A HSV colour space-based detection algorithm and watershed segmentation method were presented by Malik et al. [20] for detecting ripe tomato fruits and segmenting the clustered tomato fruits, respectively. This was like the method proposed by Indriani et al. [21] in which the colour space of HSV was combined with K-nearest neighbour and gray level co-occurrence matrix for the classification of tomato fruits into five different classes. Tomato fruits segmentation and classification algorithms based on Otsu technique was proposed by Sari et al. [22] for multiplication of Cb and V channels from the colour spaces of YUV (where Y represents luminance or brightness, and U and V represent chrominance or colour) and YCbCr (where Y represents luminance, and Cb and Cr represent the chrominance).

Nassiri et al. [23] applied fuzzy rule-based classification method to ripe tomato fruits classification by using the colour space of RGB. Fujinaga et al. [24] presented a method for recognizing tomato fruits by machine learning using infrared images and characteristics of tomato fruits. The work was carried out to address shortage of labour and aging in Japanese agriculture sector. The abovementioned works are some of the state-of-the-art colour-based models for detecting and classifying tomato fruits.

2.2. Tomato fruits detection and classification by computer vision models

Computer vision is a field of artificial intelligence developed for object detection, segmentation, and classification [25]. Computer vision is widely employed at the present time by researchers as solution to different agricultural problems such as horticulture. The agricultural evidence supporting computer vision application has led to a rapid paradigm shift in the cultivation of plants (tomato fruits). The acceptability of computer vision by the scientific community is due to its high levels of abstraction and the ability to derive complex information (features) from images, videos and other inputs automatically [26]. The architecture of computer vision solutions employed for object

Journal of Information Systems Engineering and Management
2025, 10(56s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

detection, segmentation and classification relies primarily upon CNNs, and Deep Learning (DL) technique within the family of Recurrent Neural Networks (RNNs) [27].

CNNs has the capacity to learn and interpret extremely complex and very extensive problems within a short time due to the interconnected weights and parallel operations by using advanced models [28]. YOLO is among the widely accepted and employed one-stage frameworks for object detection (even in real-time) [29], which is built on CNN-driven backbone for learning spatial hierarchy of object's features in image through backpropagation for possible timely object localization and detection in addition to fully convolutional layers (head) for object classification [30]. Likewise, ResNet50 is among the widely accepted and employed one-stage frameworks for object detection and classification (even in real-time) [31], which is built on CNN-driven backbone for learning spatial hierarchy of object's features in image through backpropagation for possible timely object localization and detection in addition to fully convolutional layers (head) for object classification [30].

The ResNet50 model became popular for several pitfall reasons that befell its predecessor networks such as difficulty in training deep neural networks; more expressive, less different; and vanishing/exploding gradient. As abovementioned, factors such as variations in illumination, overlapping, occlusion, etc, may reduce the performance of some feature extraction methods that are manual based when applied to agricultural problems [32]. Features extraction by computer vision models is by automatic manipulation which do not require manual manipulation, though, they may be applied to domain knowledge and human intuition for selection or designing features suitable for the problem [33]. Among the few papers that proposed the application of DL-driven computer vision models such as SSD, YOLO and ResNet50 to tomato fruits detection and classification are Moreira et al. [32], Liu et al. [34], Lawal [35] and Legaspi et al. [36].

By using SSD MobileNet v2 and YOLOv4 models, Moreira et al. [32] carried out the detection and classification of tomato fruits using two datasets of tomato fruits images at different stages of ripening to facilitate the detection capacity of the models and the systems were compared with their proposed model, histogram-based HSV colour space, for determining the ripening stages and the classification of individual tomato fruits. YOLOTomato was developed by Liu et al. [34] based on the existing model of YOLOv3 for tomato fruits detection in heterogenous environments. Lawal [35] furthered the detection and classification of ripening stages of tomato fruits by proposing a fusion of similar models with different activation functions. Legaspi et al. [36] employed image processing technique for tomato ripeness and size classification. More encouraging research results are being rolled out daily from the application of computer vision and DL models to complex agricultural tasks.

3. MATERIALS AND METHODS

3.1. Dataset collection and pre-processing

In this study, a dataset which comprises 1050 images of tomato fruits was collected from Roboflow. The dataset was categorized into four classes, namely green, turning, pink, and red ripe. The images were collected with the mindset of using them to train and test the computer vision models (YOLOv4 and ResNet50) and the proposed colour space model of HSV. Thus, necessitating performing on the images data augmentation such as geometric transformation, colour-based transformations, illumination transformation, noise injection, etc. [37], including trimming them to 520×520-pixel resolution to suit the input specifications for processing. This process increased the images to 2250 with 70% set aside for training the models, 10% for validating the models, and 20% for testing the models.

Moreover, in their downloaded form, the dataset contained some occlusion and overlapping; these were also normalized appropriately to ensure accurate training and prediction outcome by the models. Multiple detection and classification tasks are involved in this study, thus making the prediction problem both a detection problem and a classification problem [38]. Since it involves supervised learning, the models need to be provided with an annotated dataset (Figure 1). Thus, LabelMe [39], a web-based annotating tool for building image databases for computer vision research was employed for annotating the images before they were transferred in YOLO format [29] and ResNet format for the training of YOLO and ResNet frameworks and the proposed colour space model of HSV. Figure 2 shows the categories of tomato fruits classified according to their ripening colour as approved by UFFVA and USDA colour chart [5].

Journal of Information Systems Engineering and Management
2025, 10(56s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

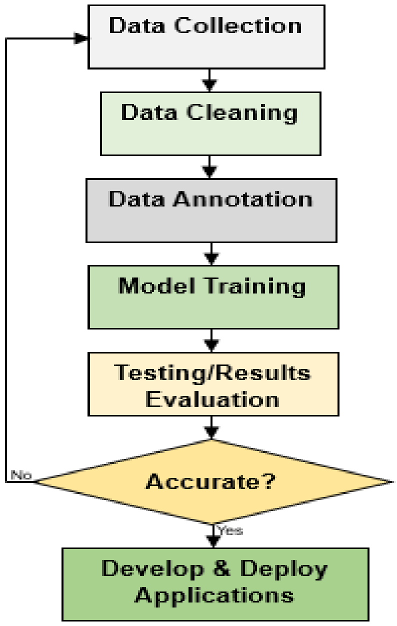


Figure 1. Process of image annotation towards training computer vision models.

Figure 2. Categories of tomato fruits classified according to their ripening colour: (a) Green, (b) Turning, (c) Pink, (d) Red ripe.

3.2. Overview of the computer vision models and proposed model

3.2.1. ResNet50

Among the popularly considered CNN architectures is the ResNet architecture, which was introduced by Microsoft Research in 2015 [40]. ResNet became popular due to a few limitations in state-of-the-art networks at the time. Some of the limitations are: (1) Complexity of deep neural networks and difficulty in training them, (2) More expressive and less different, and (3) Vanishing/exploding gradient. Among the common challenges facing the deep neural networks training is the vanishing/exploding gradient and this is caused by an oversight of numerical stability of the network's parameters. The moving from the deep layers to the shallow layers during back-propagation makes it

Journal of Information Systems Engineering and Management
2025, 10(56s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/ **Research Article**

necessary for the chain rule of differentiation to compel multiplication of the gradients which is often as small as the order of 10^ {-5}10−5 or more.

Mathematically, the multiplication of these small numbers with each other leads to them becoming infinitesimally smaller, thus resulting in making almost insignificant changes to the weights. Sometimes, the gradient is as large as the order 104 and more on the other end of the spectrum. The values keep approaching infinity as the multiplication occurs among the large gradients. There is difficulty in achieving convergence whenever such a large range of values is allowed to be in the numerical domain for weights. The abovementioned problem is known as vanishing/exploding gradient problem. The architecture of ResNet completely eliminates the occurrence of these problems by using the skip connections to act as super-highways to the gradient, thereby allowing its free flow with no barrier from a large magnitude. Figure 3 shows the building block of residual learning, mathematically, this is expressed as y = x + F(x) where y denotes the layer's final output.

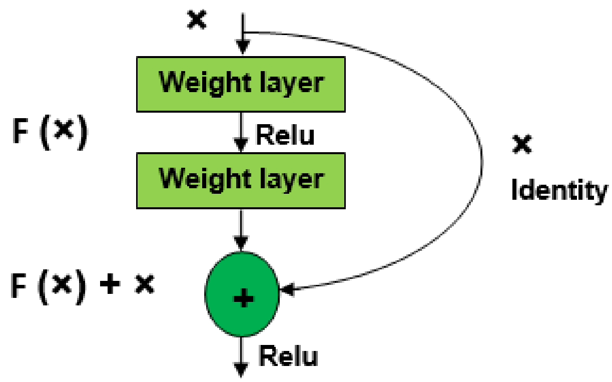


Figure 3. Residual learning as a building block [40].

The structure of the architecture enables the skip connections to skip any layers which has the tendency of damaging the model's performance in plain network. The architecture of ResNet-50 is divided into 6 different parts: (1) Input pre-processing, (2) Configuration (Cfg) [0] blocks, (3) Cfg [1] blocks, (4) Cfg [2] blocks, (5) Cfg [3] blocks, and (6) Fully connected layer. A varying number of Cfg blocks are used by ResNet architecture at different levels. For its deep residual learning capacity as abovementioned, we employed ResNet50 in this study for the detection and classification of ripening stages of tomato fruits.

We employed ResNet50 due to the following reasons:

(a) Residual learning: The use of residual connections makes ResNet50 ahead of other architectures. The residual connections significantly improve the deeper networks training and minimize vanishing gradients.

(b) Accuracy: Lighter models like Inception and MobileNet are outperformed by ReNet50 (due to its depth and design) in many standard image classification tasks like the tasks carried out in this study.

(c) Simpler architecture: The architecture of ResNet50 is straightforward compared to others like Inception and ResNeXt, thereby facilitating its implementation, tuning, and deployment.

(d) Training efficiency: ResNet50's residual architecture enables its high training efficiency and convergence, especially when deeper networks are involved.

(e) Real-time applications: While architectures like DETR is a high-performance model for object detection, ResNet50 has faster inference and capability for real-time applications with large-scale datasets where there is sufficiency of CNNs.

3.2.2. YOLOv4

YOLOv4 comprises CSPDarknet53 [41] as its backbone, SPP [42] and PAN [43] as its neck, and YOLOv3 [44] as its head. For the backbone, YOLOv4 uses: (1) Bag of Freebies (BoF), which includes augmentation of cut-mix and mosaic data, drop-block regularization, and smoothing of class label; (2) Bag of Specials (BoS), which includes mish

Journal of Information Systems Engineering and Management
2025, 10(56s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

activation, connections of cross-stage partial, and connections of multi-input weighted residual. For the detector, YOLOv4 uses: (1) Bag of Freebies (BoF), which includes complete intersection over union (ciou) loss, cross mini-batch normalization (cmbn), regularization of drop-block, augmentation of mosaic data, training of self-adversarial, grid sensitivity elimination, employing more than one anchor for one ground truth, cosine annealing scheduler [45], optimal hyperparameters, and random training shapes; (2) Bag of Specials (BoS), which includes mish activation, SPP-block, SAM-block, PAN path-aggregation block, and distance-iou-non-maximum suppression (diou-nms).

We employed YOLOv4 due to the following reasons: YOLOv4 has the endorsement of the original YOLO architecture. YOLOv4 often outperforms YOLOv5 and others in terms of accuracy and generalization due to the advanced architectural and training improvements made in YOLOv4. YOLOv4 generally offers better accuracy and more advanced techniques (like PANet, CIoU loss, and Mosaic augmentation) than its successors. Other reasons are higher mAP (mean Average Precision), better small object detection, real-time performance, optimized for both high-end and Edge devices, CSPDarknet53 backbone, feature aggregation with PANet, better handling of real-world challenges, improved training techniques, improved multi-class performance, mosaic augmentation, CutMix and self-adversarial training, complete intersection over union (CIoU), wide framework compatibility, community-driven and open-source.

A CNN commonly uses the following to improve the training of object detection; For activations, it uses Rectified Linear Unit (ReLU), ReLU6, parametric-ReLU, leaky-ReLU, Scaled Exponential Linear Unit (SeLu), and mish or swish. For regression loss of bounding box, it uses IoU, Generalized IoU (GIoU), CIoU, DIoU and Mean Square Error (MSE). For augmentation of data, it uses cut-out, mix-up, cut-mix. For method of regularization, it uses drop-out, drop-path [46], spatial drop-out [47], or drop-block. For normalizing the activations of the network by their mean and variance, it uses batch normalization [48], normalization of cross-gpu batch [49], normalization of filter response [50], or normalization of cross-iteration batch [51]. For connections of skip, it uses connections of residual, connections of weighted residual, connections of multi-input weighted residual, or partial connections of cross stage.

3.2.3. HSV Colour Space

In the colour model of RGB, Hue Saturation Value (HSV) sometimes refer to as HSB (B for brightness) and Hue Saturation Lightness (HSL) are two most popular and convenient cylindrical-coordinate representations of points developed for applications involving computer vision; others are hue, saturation, and intensity (HIS). For this study, we focus solely on HSV. The RGB geometry is rearranged by HSV representation to be more nonrational and remained germane than the cube (cartesian) representation. Usually, these definitions are consistent, but there is no standard in their usage, thereby making their abbreviations interchangeable and usable for any of the abovementioned colour models or any other related cylindrical models.

Figure 4 shows HSV representation in the cylinder. Because there is no difficulty in transforming device-dependent RGB models by HSV, the physical colours that it defines depend on the device’s primary colours of red, green and blue or of the particular space of RGB, and on the corrected gamma employed in representing the quantities of those primaries. Each unique RGB device is accompanied by a unique HSV space, and HSV value describes a different colour for each basis RGB space. In this study, HSV model is used for the detection and classification of ripening stages of tomato fruits using their feature detection and image segmentation, which is a computer vision and image analysis problem. In most cases, the algorithms of computer vision employed on colour images are direct extensions to algorithms developed for grayscale images.

Basically, the components of each colour are distinctly passed through the same algorithm. Based on this, it is essential, therefore, for features of interest to be easily distinguished in the colour dimensions employed. Object discrimination is made difficult by image descriptions in terms of RGB components due to the correlation of all the RGB components of the colour of a digital image’s object with the quantity of light striking the object.

Journal of Information Systems Engineering and Management
2025, 10(56s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

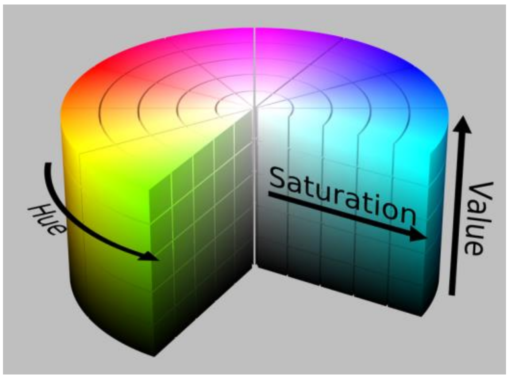


Figure 4. HSV cylinder [52].

The main reason of using HSV for ripening classification is for the separation of colour (Hue), purity/intensity (Saturation), and brightness/intensity (Value), thereby facilitating the detection of subtle colour changes that occur as tomato fruits ripen. There are several key steps involved in the flowchart for classifying the ripening stages of tomato fruits using the HSV (Hue, Saturation, Value) colour space. The detailed flowchart describing the process is shown in Figure 5. The description of the steps in Figure 5 is as follows.

(a) Image acquisition: Images of tomato fruits were collected at different stages of their ripening (Green, Turning, Pink, and Red ripe).

(b) Preprocessing: Images were cleaned and prepared by applying enhancement methods, such as resizing and cropping for noise removal and improvement of detection quality.

(c) Image convertion to HSV: Input image was converted from the RGB colour space to the HSV colour space, thereby isolating the colour information, which is essential for ripening stage classification.

(d) Hue, Saturation, and Value channels extraction: Individual channels (Hue, Saturation, Value) were extracted from the HSV image. Typically, Hue indicates the colour, which is most significant for determining ripeness.

(e) Thresholding application on HSV channels: Thresholding techniques were applied on HSV channels for isolation of specific ranges of Hue, Saturation, and Value that have correlation with different stages of ripening. For example, while Green tomato fruits indicate early ripening stage with a specific Hue range, Red tomatoes indicate fully ripe with a different Hue range.

(f) Feature (colour/texture) extraction: Relevant features were extracted from the HSV channels for classification of ripening stages. These include colour features such as mean or dominant Hue, Saturation, and Value values; texture features such as applying techniques like Gray Level Co-occurrence Matrix (GLCM) for texture patterns capturing related to ripening; and shape and size features analysis.

(g) Classification models: The ripening stage of the tomato fruits was classified using a machine learning and deep learning models.

(h) Stage prediction: The ripening stage was predicted by the models based on the extracted features.

(i) Output classification: The tomato fruits' ripening stage was classified and output, which is suitable for quality control applications or agricultural automation systems.

Journal of Information Systems Engineering and Management
2025, 10(56s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

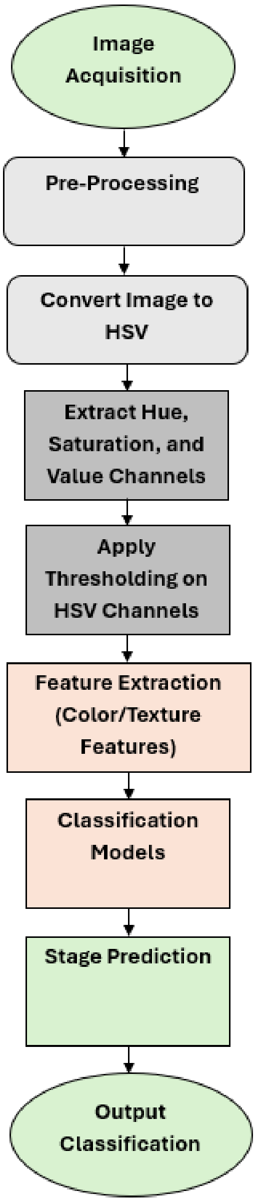


Figure 5. Flowchart illustrating the several key steps involved in the flowchart for classifying the ripening stages of tomato fruits using the HSV (Hue, Saturation, Value) colour space.

IV. EVALUATION METRIC

While the evaluation of the colour space model of HSV was restricted to only classification problem, the evaluation of the computer vision models was based on their capacity to detect and classify the ripening stages of tomato fruits in the image irrespective of the ripeness degree. To launch an accurate detection, an IoU (Intersection over Union) metric is often employed; IoU is employed to measure the area overlapped in an image between the bounding boxes of the predicted and the ground-truth divided by the area of union. Therefore, a detection is said to be accurate if and only if IoU is greater or equal to 50% (IoU ≥ 50%), and this expression is represented by Equation (1).

$$IoU = \frac{A\left(BB_p \cap BB_g\right)}{A(BB_p \cup BB_g)} \quad (1)$$

where, A denotes Area, BB denotes Bounding Box, p denotes predicted, g denotes ground-truth, ∩ denotes intersection, and U denotes union.

Journal of Information Systems Engineering and Management
2025, 10(56s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Other evaluation metric considered in this study are Precision (Equation (2)), Recall (Equation (3)), F1-score (Equation (4)) and mean Average Precision, mAP (Equation (5)).

$$Precision = \frac{TP}{TP + FP} \tag{2}$$

$$Recall = \frac{TP}{TP + FN} \tag{3}$$

$$F1 - score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{4}$$

where, T denotes True, P denotes Positive, F denotes False, and N denotes Negative.

$$AP = \frac{1}{n}\sum_{i=1}^{n} P(i)\,.\,\Delta R(i) \tag{5}$$

where, n denotes the total number of recall points, P(i) the precision at the i-th recall point and $\Delta R(i)$ is the change in recall between the i-th and (i+1)-th recall points.

$$mAP = \frac{1}{n}\sum_{n=1}^{n} AP_i \tag{6}$$

where, APi denotes the Average Precision for class i, and n denotes classes' total number.

V. MODELS IMPLEMENTATION

5.1. Training of the Models

For model training and inference scripts, this study employs TensorFlow framework, which resides in Google Colab, a hosted Jupyter Notebook service that doesn't require any setup for its usage, with free access to computing resources including GPUs (Graphics Processing Units) and TPUs (Tensor Processing Units) for machine learning and computer vision models development. A NVIDIA Tesla T4 with a 12 GB VRAM and high compute capability was allotted to all sessions so there won't be much effect of the difference in GPUs made available to each session. The two proposed models, YOLOv4 and ResNet50 were pre-trained with COCO dataset (Common Objects in Context) with 520×520-pixel resolution to suit the input specifications for processing, with 70% set aside for training the models, 10% for validating the models, and 20% for testing the models.

By fine-tuning, the pre-trained models were taking on source task (detection and segmentation) and trained further on the target task (the detection and classification of ripening stages of tomato fruits) by updating the models' weights using the target's task data (images of tomato fruits) while retaining some knowledge from the source task. To maintain equal training parameters between YOLOv4 and ResNet50 models, their batch size parameters were adjusted from the default to 32 for ResNet50 (with 100 epochs at learning rate of 0.1) and 64 for YOLOv4 (with 5000 epochs at learning rate of 0.01). These evaluation sessions are quite useful, since they allow monitoring the evolution of the training, meaning if the evaluation loss started to increase while the training loss decreased or remained constant, the computer vision model was over-fit to the training data.

A colour space of HSV was developed as an approach to augment the YOLOv4 and ResNet50 models for tomato fruits classification. To develop the model and give the tomato fruits a different perspective, seven of their images were selected from each class of their ripening. The Roboflow tomato fruits dataset being an open-source dataset, has a generalized perspective. Therefore, segmentation and extraction of region of ripening (interest) from the image was the first stage involved after inputting the image, and this was by using the annotation bounding-box's coordinates created by the LabelMe annotation tool that was used in labelling all the images. The conversion of the region of ripening on the images from RGB colour to colour space of HSV was the second stage, this is necessary to reduce the negative effects of noise that colour information of RGB would have caused by using only Hue of HSV. However, a colour histogram that focused only on hue channel was generated for each of the HSV images.

VI. RESULTS AND DISCUSSION

Results based on the ripening tomato fruits detection and classification are presented in this section. The classifying colour space model of HSV was built by each sample's hue histogram mean, and the classification of the tomato fruits was by defining thresholds for each class. For the tomato fruits images input for classification, and given the coordinates of their bounding boxes, all in a single pass, the regions of ripening were segmented and converted to colour space of HSV by the colour space model of HSV via the colorimetric information. By this process, the class of the tomato fruits was predicted by the model. The purpose for applying colour space of HSV is for classification, the computer vision models of YOLOv4 and ResNet50 were applied for the detection of the tomato fruits.

The performance of the computer vision models employed in this study required that the best confidence threshold be defined before evaluating them. Table I presents the evaluation results of the computer vision models on ripening tomato fruits detection. Among the crucial machine learning metric is the F1-score, through which a balanced measure is provided of a model's precision and recall. The harmonic mean of precision and recall is used in deriving the formular for finding F1-score, thus making it indispensable component in the framework of precision recall F1-score. The confidence in the predictions of both models were close and similar, this simply means that at close and similar confidence thresholds of 57% and 59%, respectively, both models got their best F1-score and mAP. However, YOLOv4 model achieved a better F1-score of 89.50% and mAP of 83.20% than the F1-score of 77.01% and mAP of 71.01% achieved by the ResNet50 model.

Comparing the performance of the models with each other in detection of tomato fruits, the colour space model of HSV for Green ripening tomato fruits obtained the highest percentage in all the evaluation metric for the classification with 97.45% precision, 98.11% recall, and 97.78% F1-score. We focused more on precision, which focuses on the accuracy of positive predictions (reducing false positives) than the recall metric, which focuses on identifying all relevant cases (reducing false negatives). Therefore, Green and Red ripe are the most important determinants of the ripening process, which is the reason why we based our precision on these two and selected the best model based on them.

As presented in Table I, unlike the validation set benchmark, ability to understand the generalisation capacity of the computer vision models is by test set benchmark. The snapshot analysis in Figure 6, Figure 7 and Figure 8 reveal the computer vision models capacity through example images of the test set, respectively. In Figure 6, a snapshot shows the results of ripening tomato fruits classification by detection using YOLOv4. The green bar chart, turning bar chart, pink bar chart, and red bar chart represent Green, Turning, Pink, and Red ripe tomato fruits predictions, respectively. The bars are reflections of the bounding boxes generated in the VEED. In Figure 7, the snapshot shows the results of ripening tomato fruits classification by detection using ResNet50. The green bar chart, turning bar chart, pink bar chart, and red bar chart represent Green, Turning, Pink, and Red ripe tomato fruits predictions, respectively.

The bars are reflections of the bounding boxes generated in the VEED. In Figure 8, the snapshot shows the results of ripening tomato fruits classification by detection using colour space of HSV. The green bar chart, turning bar chart, pink bar chart, and red bar chart represent Green, Turning, Pink, and Red ripe tomato fruits predictions, respectively. The bars are reflections of the bounding boxes generated in the VEED. Due to the robustness of the two models, factors such as occlusion, variation in illumination, and heterogenous environment under which the tomato fruits images were acquired could not have negative effect on both models that were employed for the tomato fruits detection. The colour of each ripening tomato fruits was detected by the two models without mismatching tomato fruits objects for the background.

Table II presents the evaluation results of the proposed colour space model of HSV and the computer vision models on ripening tomato fruits classification. The colour space model of HSV for Green ripening tomato fruits obtained the highest percentage in all the evaluation metric for the classification with 97.45% precision, 98.11% recall, and 97.78% F1-score. In this study, we trained and tested YOLOv4 and ResNet50 on pre-processed tomato fruits image dataset for their detection, and the classification of their ripening were performed using the two computer vision models and a developed colour space model of HSV. All the models performed effectively in detecting and classifying the tomato fruits.

Journal of Information Systems Engineering and Management
2025, 10(56s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/ **Research Article**

Comparing the performance of the models with each other in detection of tomato fruits, the confidence in the predictions of both models were close and similar, this simply means that at close and similar confidence thresholds of 57% and 59%, respectively, both models got their best F1-score with YOLOv4 achieving a better mAP of 83.20% than the 71.01% mAP achieved by the ResNet50. The colour space model of HSV for Green ripening tomato fruits obtained the highest percentage in all the evaluation metric for the classification with 97.45% precision, 98.11% recall, and 97.78% F1-score. Moreover, to establish the effectiveness and performance of the models employed in this study, we compared the models with some existing related work. It was reported in Flores et al. [53] that lycopene increase depends on tomato fruits ripening or maturity progression. This was corroborated in Coyago-Cruz et al. [54] where it was more established, the role played by visual evaluation in examining the ripening status of tomato fruits since red colour and variation in the redness of tomato fruit is related to lycopene content. Coyago-Cruz et al. [54] highlighted that the state-of-the-art applications for assessing fruit colour and ripening are gaining popularity among horticulture community due to their non-destructive optical and analytical methods. The models proposed in this study are on a par with the models presented in Moreira et al. [32], Liu et al. [34], Lawal [35] and Legaspi et al. [36]. Table III shows the comparative evaluation results of YOLOv4, ResNet50, and SSD MobileNet v2 [32] on ripening tomato fruits detection. Table IV shows the comparative evaluation results of the colour space model of HSV and computer vision models on ripening tomato fruits classification. Figure 9 shows the results of ripening tomato fruits classification by detection in greater or lesser complexity situations using SSD MobileNet v2 [32]. The use of residual connections makes ResNet50 ahead of other architectures and significantly improve the deeper networks training and minimizes vanishing gradients [55]. Lighter models like Inception and MobileNet are outperformed by ReNet50 (due to its depth and design) in many standard image classification tasks like the tasks carried out in this study [56]. The architecture of ResNet50 is straightforward compared to others like Inception and ResNeXt, thereby facilitating its implementation, tuning, and deployment [57]. ResNet50's residual architecture enables its high training efficiency and convergence, especially when deeper networks are involved [57]. While architectures like DETR is a high-performance model for object detection, ResNet50 has faster inference and capability for real-time applications with large-scale datasets where there is sufficiency of CNNs [58]. The main reasons for using YOLOv4 in this study include (a) Compatibility with the available datasets, which are too small to train higher versions of the YOLO series, (b) High computational resources needed to train and implement higher versions of YOLO series, which are not readily available.

Table 1: Evaluation Results of YOLOV4 and RESNET50 on Ripening Tomato Fruits Detection

Computer vision models	Confidence threshold ≥	Precision	mAP	Recall	F1-Score
ResNet50	57%	79.01%	71.01%	75.01%	77.01%
YOLOv4	59%	90.46%	83.20%	88.54%	89.50%

Table 2: Evaluation Results of the Colour Space Model of HSV and Computer Vision Models on Ripening Tomato Fruits Classification

Computer vision models	Tomato ripening	Precision	Recall	F1-Score
ResNet50	Green	87.54%	85.01%	86.27%
	Turning	72.64%	68.20%	70.42%
	Pink	74.20%	59.03%	66.61%
	Red	89.43%	78.54%	83.98%
YOLOv4	Green	79.45%	71.01%	75.23%
	Turning	61.46%	58.20%	59.83%
	Pink	64.02%	51.03%	57.52%
	Red	82.34%	84.54%	83.44%
HSV Colour Space	Green	97.45%	98.11%	97.78%
	Turning	51.46%	68.20%	59.83%
	Pink	54.02%	50.03%	52.02%

Journal of Information Systems Engineering and Management
2025, 10(56s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

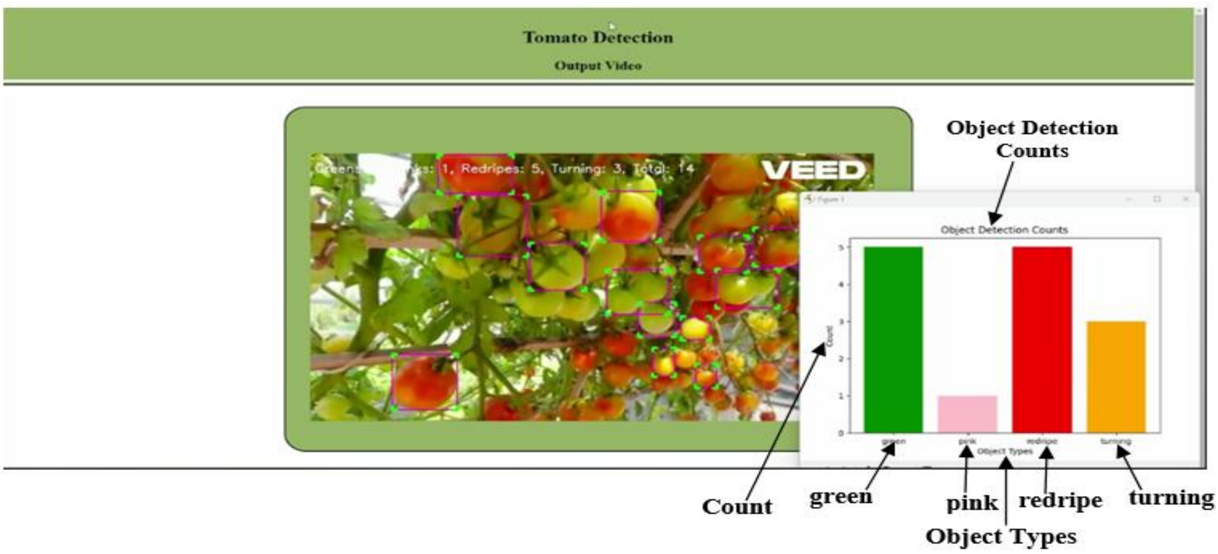
	Red	90.46%	74.54%	82.53%

Table 3: Comparative Evaluation Results of YOLOV4, RESNET50 and SSD MOBILENET V2 on Ripening Tomato Fruits Detection

Computer vision models	Confidence threshold ≥	Precision	mAP	Recall	F1-Score
ResNet50	57%	79.01%	71.01%	75.01%	77.01%
YOLOv4	59%	90.46%	83.20%	88.54%	89.50%
SSD MobileNet v2 [32]	56%	77.62%	65.38%	70.12%	73.68%

Table 4: Comparative Evaluation Results of the Colour Space Model of HSV and Computer Vision Models on Ripening Tomato Fruits Classification

Computer vision models	Tomato ripening	Precision	Recall	F1-Score
ResNet50	Green	87.54%	85.01%	86.27%
	Turning	72.64%	68.20%	70.42%
	Pink	74.20%	59.03%	66.61%
	Red	89.43%	78.54%	83.98%
YOLOv4	Green	79.45%	71.01%	75.23%
	Turning	61.46%	58.20%	59.83%
	Pink	64.02%	51.03%	57.52%
	Red	82.34%	84.54%	83.44%
HSV Colour Space	Green	97.45%	98.11%	97.78%
	Turning	51.46%	68.20%	59.83%
	Pink	54.02%	50.03%	52.02%
	Red	90.46%	74.54%	82.53%
SSD MobileNet v2 [32]	Green	77.27%	70.09%	73.68%
	Turning	59.38%	55.88%	57.63%
	Pink	60.61%	40.82%	50.72%
	Red	80.77%	84.00%	82.39%



Journal of Information Systems Engineering and Management
2025, 10(56s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/ **Research Article**

Figure 6. Snapshot showing results of ripening tomato fruits classification by detection using YOLOv4. The green bar chart, turning bar chart, pink bar chart, and red bar chart represent Green, Turning, Pink, and Red ripe tomato fruits predictions, respectively. The bars are reflections of the bounding boxes generated in the VEED.

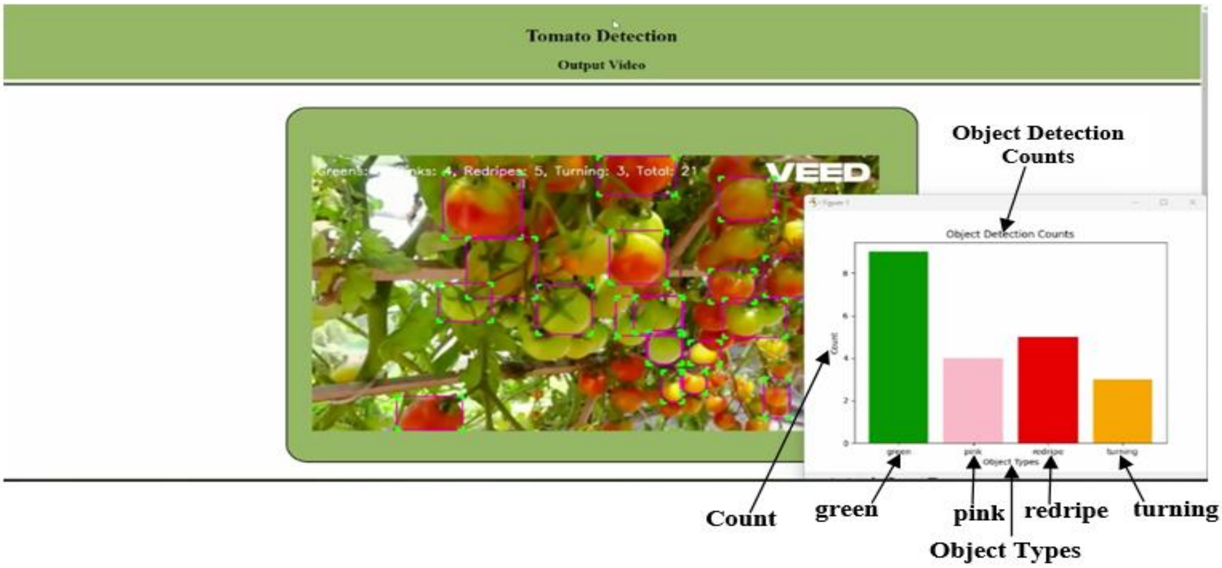


Figure 7. Snapshot showing results of ripening tomato fruits classification by detection using ResNet50. The green bar chart, turning bar chart, pink bar chart, and red bar chart represent Green, Turning, Pink, and Red ripe tomato fruits predictions, respectively. The bars are reflections of the bounding boxes generated in the VEED.

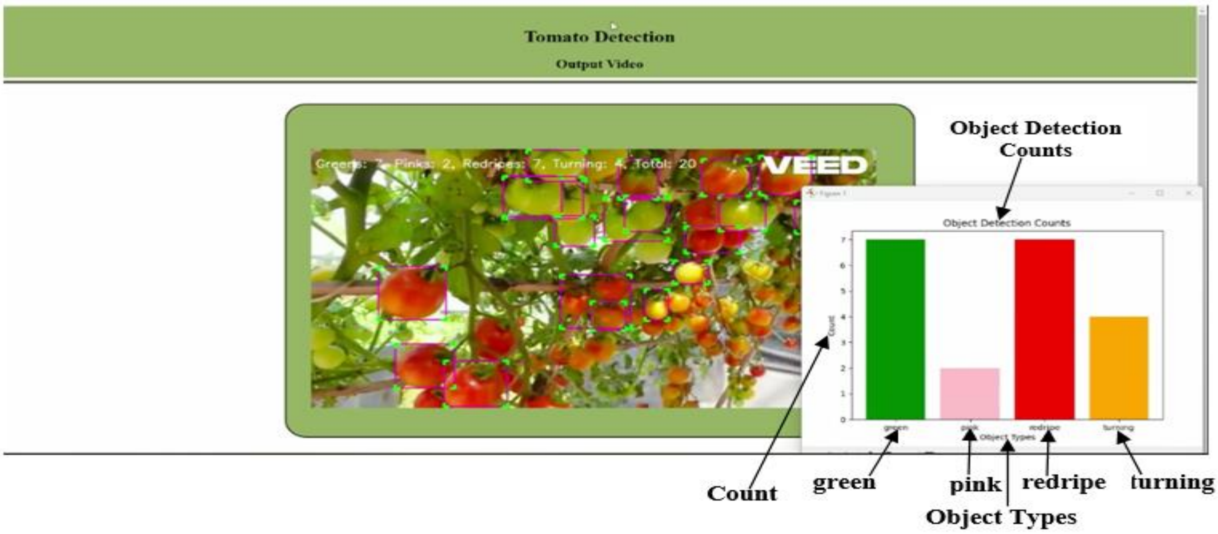


Figure 8. Snapshot showing results of ripening tomato fruits classification by detection using colour space of HSV. The green bar chart, turning bar chart, pink bar chart, and red bar chart represent Green, Turning, Pink, and Red ripe tomato fruits predictions, respectively. The bars are reflections of the bounding boxes generated in the VEED.

Journal of Information Systems Engineering and Management
2025, 10(56s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/ **Research Article**

Figure 9. Snapshot showing results of ripening tomato fruits classification by detection in greater or lesser complexity situations using SSD MobileNet v2 [32]. Bounding boxes in blue represents annotations of the groundtruth; The bounding boxes in Green, Yellow, Orange and Red represents Green, Turning, Light Red (Purple) and Red tomato fruits predictions, respectively.

CONCLUSIONS

Computer vison-based detection models for classifying ripening stages of tomato fruits have been presented in this paper. YOLOv4 and ResNet50 were trained and tested on pre-processed tomato fruits image datasets for their detection, and the classification of their ripening were performed using the two computer vision models and a developed colour space model of HSV. All the models performed effectively in detecting and classifying the tomato fruits.

Comparing the performance of the models with each other in detection of tomato fruits, the confidence in the predictions of both models were close and similar, this simply means that at close and similar confidence thresholds of 57% and 59%, respectively, both models got their best F1-score and mAP; however, YOLOv4 model achieved a better F1-score of 89.50% and mAP of 83.20% than the F1-score of 77.01% and mAP of 71.01% achieved by the ResNet50 model. The colour space model of HSV for Green ripening tomato fruits obtained the highest percentage in all the evaluation metric for the classification with 97.45% precision, 98.11% recall, and 97.78% F1-score.

Despite the significant results obtained by the proposed models in this study, several limitations still affected their performance. Some of these limitations include variability in tomato fruits appearance, lighting conditions, scarcity of quality data, limited generalization, and dependence on specific features. To address these limitations, we propose several future improvement directions, which include embracing multimodal inputs, transfer learning, improved data augmentation, multi-scale and multi-view approaches, improved generalization, and integration with agricultural automation. With continuous improvements in the abovementioned areas, computer vision systems can become even more powerful tools in automating agricultural processes.

REFERENCES

[1] S. O. Salami, O. D. Adegbaju, O. A. Idris, M. O. Jimoh, T. L. Olatunji, S. Omonona, ..., and C. P. Laubscher, "South African wild fruits and vegetables under a changing climate: The implications on health and economy," South African Journal of Botany, vol. 145, 13-27, 2022.

[2] J. Dube, G. Ddamulira, and M. Maphosa, "Tomato breeding in sub-Saharan Africa-challenges and opportunities: A review," African Crop Science Journal, vol. 28, no. 1, 131-140, 2020.
[3] I. V. Diphoko, "Agricultural commercialization through market participation by smallholder tomato farmers in Bojanala District, North West Province," Doctoral dissertation, North-West University, South Africa, 2023.
[4] M. Quinet, T. Angosto, F. J. Yuste-Lisbona, R. Blanchard-Gros, S. Bigot, J. P. Martinez, and S. Lutts, "Tomato fruit development and metabolism," Frontiers in Plant Science, vol. 10, no. 1554, 2019.
[5] L. Sharma, K. V. Ramesh, V. Paul, and R. Pandey, "Ripening index: a better parameter for colour based assessment of ripening behaviour of tomato fruits," Plant Physiology Reports, vol. 25, 171-177, 2020.
[6] R. Kumar, V. Paul, R. Pandey, R. N. Sahoo, and V. K. Gupta, "Reflectance based non-destructive determination of colour and ripeness of tomato fruits," Physiology and Molecular Biology of Plants, vol. 28, no. 1, 275-288, 2022.
[7] T. N. Tadesse, A. M. Ibrahim, and W. G. Abtew, "Degradation and formation of fruit color in tomato (Solanum lycopersicum L.) in response to storage temperature," American Journal of Food Technology, vol. 10, no. 4, 147-157, 2015.
[8] M. Al-Dairi, P. B. Pathare, and R. Al-Yahyai, "Effect of postharvest transport and storage on color and firmness quality of tomato," Horticulturae, vol. 7, no. 7, 163, 2021.
[9] G. Abera, A. M. Ibrahim, S. F. Forsido, and C. G. Kuyu, "Assessment on post-harvest losses of tomato (Lycopersicon esculentem Mill.) in selected districts of East Shewa Zone of Ethiopia using a commodity system analysis methodology," Heliyon, vol. 6, no. 4, 2020.
[10] R. W. Bello, P. A. Owolawi, E. A. van Wyk, and C. Tu, "SAM-IE: SAM-enabled image enhancement for segmentation of infected cucumber leaves," International Journal of Innovative Research and Scientific Studies, vol. 8, no. 2, 824-832, 2025.
[11] W. Mandil, V. Rajendran, K. Nazari, and A. Ghalamzan-Esfahani, "Tactile-sensing technologies: Trends, challenges and outlook in agri-food manipulation," Sensors, vol. 23, no. 17, 7362, 2023.
[12] J. Wu, B. Zhang, J. Zhou, Y. Xiong, B. Gu, and X. Yang, "Automatic recognition of ripening tomatoes by combining multi-feature fusion with a bi-layer classification strategy for harvesting robots," Sensors, vol. 19, no. 3, 612, 2019.
[13] K. N. E. Alam Siddiquee, M. S. Islam, M. Y. U. Dowla, K. M. Rezaul, and V. Grout, "Detection, quantification and classification of ripened tomatoes: a comparative analysis of image processing and machine learning," IET Image Processing, vol. 14, no. 11, 2442-2456, 2020.
[14] P. Jiang, D. Ergu, F. Liu, Y. Cai, and B. Ma, "A review of Yolo algorithm developments," Procedia Computer Science, vol. 199, 1066-1073, 2022.
[15] S. Mascarenhas, and M. Agarwal, "A comparison between VGG16, VGG19 and ResNet50 architecture frameworks for image classification," In IEEE International Conference on Disruptive Technologies for Multi-disciplinary Research and Applications (CENTCON), 2021, vol. 1, pp. 96-99.
[16] T. C. Padilha, G. Moreira, S. A. Magalhães, F. N. dos Santos, M. Cunha, and M. Oliveira, "Tomato detection using deep learning for robotics application," In Progress in Artificial Intelligence: 20th EPIA Conference on Artificial Intelligence, EPIA 2021, Virtual Event, Springer International Publishing, September 7–9, 2021, Proceedings 20, 2021, pp. 27-38.
[17] M. K. Tripathi, and D. D. Maktedar, "A role of computer vision in fruits and vegetables among various horticulture products of agriculture fields: A survey," Information Processing in Agriculture, vol. 7, no. 2, 183-203, 2020.
[18] T. Fujinaga, S. Yasukawa, and K. Ishii, "Development and evaluation of a tomato fruit suction cutting device," In 2021 IEEE/SICE International Symposium on System Integration (SII), 2021, pp. 628-633.
[19] Q. Rong, C. Hu, X. Hu, and M. Xu, "Picking point recognition for ripe tomatoes using semantic segmentation and morphological processing," Computers and Electronics in Agriculture, vol. 210, no. 107923, 2023.
[20] M. H. Malik, T. Zhang, H. Li, M. Zhang, S. Shabbir, and A. Saeed, "Mature tomato fruit detection algorithm based on improved HSV and watershed algorithm," IFAC-PapersOnLine, vol. 51, no. 17, 431-436, 2018.
[21] O. R. Indriani, E. J. Kusuma, C. A. Sari, and E. H. Rachmawanto, "Tomatoes classification using K-NN based on GLCM and HSV color space," In IEEE International Conference on Innovative and Creative Information Technology (ICITech), 2017, pp. 1-6.

Journal of Information Systems Engineering and Management
2025, 10(56s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

[22] Y. A. Sari, S. Adinugroho, P. P. Adikara, and A. Izzah, “Multiplication of V and Cb color channel using Otsu thresholding for tomato maturity clustering,” In IEEE International Conference on Sustainable Information Engineering and Technology (SIET), 2017, pp. 209-214.
[23] S. M. Nassiri, A. Tahavoor, and A. Jafari, “Fuzzy logic classification of mature tomatoes based on physical properties fusion,” Information Processing in Agriculture, vol. 9, no. 4, 547-555, 2022.
[24] T. Fujinaga, S. Yasukawa, B. Li, and K. Ishii, “Recognition of tomato fruit regardless of maturity by machine learning using infrared image and specular reflection,” In The 2018 International Conference on Artificial Life and Robotics (ICAROB 2018), 2018, pp. 761-766.
[25] A. Z. Da Costa, H. E. Figueroa, and J. A. Fracarolli, “Computer vision based detection of external defects on tomatoes using deep learning,” Biosystems Engineering, vol. 190, 131-144, 2020.
[26] J. Naranjo-Torres, M. Mora, R. Hernández-García, R. J. Barrientos, C. Fredes, and A. Valenzuela, “A review of convolutional neural network applied to fruit image processing,” Applied Sciences, vol. 10, no. 10, 3443, 2020.
[27] C. B. Murthy, M. F. Hashmi, N. D. Bokde, and Z. W. Geem, “Investigations of object detection in images/videos using various deep learning techniques and embedded platforms—A comprehensive review,” Applied Sciences, vol. 10, no. 9, 3280, 2020.
[28] M. M. Taye, “Theoretical understanding of convolutional neural network: Concepts, architectures, applications, future directions,” Computation, vol. 11, no. 3, 52, 2023.
[29] A. Vijayakumar, and S. Vairavasundaram, “Yolo-based object detection models: A review and its applications,” Multimedia Tools and Applications, 1-40, 2024.
[30] U. Sirisha, S. P. Praveen, P. N. Srinivasu, P. Barsocchi, and A. K. Bhoi, “Statistical analysis of design aspects of various YOLO-based deep learning models for object detection,” International Journal of Computational Intelligence Systems, vol. 16, no. 1, 126, 2023.
[31] P. Goswami, L. Aggarwal, A. Kumar, R. Kanwar, and U. Vasisht, “Real-time evaluation of object detection models across open world scenarios,” Applied Soft Computing, vol. 163, no. 111921, 2024.
[32] G. Moreira, S. A. Magalhães, T. Pinho, F. N. dos Santos, and M. Cunha, “Benchmark of deep learning and a proposed HSV colour space models for the detection and classification of greenhouse tomato,” Agronomy, vol. 12, no. 2, 356, 2022.
[33] A. Koirala, K. B. Walsh, Z. Wang, and C. McCarthy, “Deep learning–method overview and review of use for fruit detection and yield estimation,” Computers and Electronics in Agriculture, vol. 162, 219-234, 2019.
[34] G. Liu, J. C. Nouaze, P. L. Touko Mbouembe, and J. H. Kim, “YOLO-tomato: A robust algorithm for tomato detection based on YOLOv3,” Sensors, vol. 20, no. 7, 2145, 2020.
[35] O. Lawal, “Tomato detection based on modified YOLOv3 framework,” Sci. Rep., vol. 2021, no. 11, 1447, 2020.
[36] J. Legaspi, J. R. Pangilinan, and N. Linsangan, “Tomato Ripeness and Size Classification Using Image Processing,” In 2022 5th International Seminar on Research of Information Technology and Intelligent Systems (ISRITI), 2022, pp. 613-618.
[37] A. Mumuni, and F. Mumuni, “Data augmentation: A comprehensive survey of modern approaches,” Array, vol. 16, no. 100258, 2022.
[38] M. Grandini, E. Bagli, and G. Visani, “Metrics for multi-class classification: An overview,” arXiv 2020, arXiv:stat.ML/2008.05756, 2020.
[39] B. C. Russell, A. Torralba, K. P. Murphy, and W. T. Freeman, “LabelMe: a database and web-based tool for image annotation,” International Journal of Computer Vision, vol. 77, 157-173, 2008.
[40] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 770-778.
[41] C. Y. Wang, H. Y. M. Liao, Y. H. Wu, P. Y. Chen, J. W. Hsieh, and I. H. Yeh, “CSPNet: A new backbone that can enhance learning capability of CNN,” In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, 2020, pp. 390-391.
[42] K. He, X. Zhang, S. Ren, and J. Sun, “Spatial pyramid pooling in deep convolutional networks for visual recognition,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 37, no. 9, 1904-1916, 2015.
[43] S. Liu, L. Qi, H. Qin, J. Shi, and J. Jia, “Path aggregation network for instance segmentation,” In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 8759-8768.
[44] J. Redmon, “Yolov3: An incremental improvement,” arXiv preprint arXiv:1804.02767, 2018.

Journal of Information Systems Engineering and Management
2025, 10(56s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/ **Research Article**

[45] I. Loshchilov, and F. Hutter, "Sgdr: Stochastic gradient descent with warm restarts," arXiv preprint arXiv:1608.03983, 2016.
[46] G. Larsson, M. Maire, and G. Shakhnarovich, "Fractalnet: Ultra-deep neural networks without residuals," arXiv preprint arXiv:1605.07648, 2016.
[47] J. Tompson, R. Goroshin, A. Jain, Y. LeCun, and C. Bregler, "Efficient object localization using convolutional networks," In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 648-656.
[48] S. Ioffe, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," arXiv preprint arXiv:1502.03167, 2015.
[49] H. Zhang, K. Dana, J. Shi, Z. Zhang, X. Wang, A. Tyagi, and A. Agrawal, "Context encoding for semantic segmentation," In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 7151-7160.
[50] S. Singh, and S. Krishnan, "Filter response normalization layer: Eliminating batch dependence in the training of deep neural networks," In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 11237-11246.
[51] Z. Yao, Y. Cao, S. Zheng, G. Huang, and S. Lin, "Cross-iteration batch normalization," In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021, pp. 12331-12340.
[52] HSV_color_solid_cylinder.png: SharkD derivative work: SharkD Talk HSV_color_solid_cylinder.png, CC BY-SA 3.0, https://commons.wikimedia.org/w/index.php?curid=9801673 (accessed on 10/10/2024).
[53] P. Flores, E. Sánchez, J. Fenoll, and P. Hellín, "Genotypic variability of carotenoids in traditional tomato cultivars," Food Research International, vol. 100, 510-516, 2017.
[54] E. Coyago-Cruz, M. Corell, A. Moriana, P. Mapelli-Brahm, D. Hernanz, C. M. Stinco, ..., and A. J. Meléndez-Martínez, "Study of commercial quality parameters, sugars, phenolics, carotenoids and plastids in different tomato varieties," Food Chemistry, vol. 277, 480-489,2019.
[55] L. Borawar, and R. Kaur, "ResNet: Solving vanishing gradient in deep networks," In Proceedings of International Conference on Recent Trends in Computing: ICRTC 2022, Singapore: Springer Nature Singapore, 2023, pp. 235-247.
[56] M. R. M. Ariefwan, I. G. S. M. Diyasa, and K. M. Hindrayani, "InceptionV3, ResNet50, ResNet18 and MobileNetV2 performance comparison on face recognition classification," Literasi Nusantara, vol. 4, no. 1, 1-10, 2023.
[57] S. N. Kumar, I. B. Kalyani, A. Govina, R. R. Chandra, and P. Choubey, "Corn leaf disease detection using resnext50, resnext101, and inception v3 deep neural networks," In Machine Vision and Augmented Intelligence: Select Proceedings of MAI 2022, Singapore: Springer Nature Singapore, 2023, pp. 303-313.
[58] H. G. Suriyage, and H. M. S. C. Rathnayake, "Weed Detection in Agriculture using End-to-End Object Detection with Transformers in ResNet-50," In 2024 International Research Conference on Smart Computing and Systems Engineering (SCSE), IEEE, 2024, pp. 1-6.