2026-04

Machine vision-based pepper breed classification Using yolov10 and transfer learning models

Tanvir, K.M.

IUB

MACHINE VISION-BASED PEPPER BREED CLASSIFICATION USING YOLOv10 AND TRANSFER LEARNING MODELS

April 2026

Prepared by

K.M. Tanvir

ID: 2220778

Athina Sarkar Chaitee

ID: 2220940

Mahmuda Akter Moni

ID:2220183

Department of Computer Science and Engineering

Independent University, Bangladesh

Supervised by

Dr. Md. Tarek Habib

Associate Professor

Department of Computer Science and Engineering

Independent University, Bangladesh

Attestation

We are aware of the fact that plagiarism is strictly prohibited, and it conflicts with our university's rules and regulations. We guarantee the authenticity of our work. We have used some copyrighted material and models of others in our project which has been properly cited following the international standards proper guidelines.

Author Name:

K.M. Tanvir

Signature:

. . . .

Author Name:

Athina Sarkar Chaitee

Signature:

. . . .

Author Name:

Mahmuda Akter Moni

Signature:

. . . .

Evaluation Committee

Supervisor	
Name: Dr. Md. Tarek Habib	Signature: ___________________

Internal Examiner 1	
Name: _______________________	Signature: ___________________

Internal Examiner 2	
Name: _______________________	Signature: ___________________

External Examiner	
Name: _______________________	Signature: ___________________

Acknowledgement

First, we are grateful to God as He gives us the strength and patience to be able to complete this research. It is our great pleasure to thank our esteemed Associate Professor, Supervisor, Dr. Md. Tarek Habib, Independent University, Bangladesh. His zealous advice, prudent ideas, big forbearance, and that liberality of his precious time was a decisive part in the efficient achievement of our project. His knowledge and guidance to a great deal added to the quality and richness of our work, and we are more grateful to his perpetual encouragement in the course of our studies.

We also want to like to thank Professor Shipra Banik, PhD, Independent University, Bangladesh. She inspired and helped us by her encouraging advice remain committed and concentrated in this project. We wish to offer our thanks to Md. Ataur Rahman for technical assistance and practical observations did much bring to advance our project. His support was able to assist us in overcoming difficulties and reaching higher results.

Also, we want to thank Fab Lab for help and support. The cooperative and innovative culture offered by Fab Lab played a major role in the project delivery. Finally, we acknowledge and appreciate the help by everyone who supported us directly or indirectly on this endeavor.

Abstract

One of the most popular crops grown and eaten in Bangladesh is peppers, which include different types of different peppers; capsicum (red, yellow and green), local chili (red and green), Shimla, and Bombay chili. These types are not only the focus of culinary activities but also have a great nutritional and financial importance. Peppers contain necessary vitamins, such as C, A, B6, K, E, and antioxidants, such as carotenoids and flavonoids and minerals, such as potassium. Their nutritional quality highlights their significance in the human health, and their economic value highlights the importance of proper classification of breeds in agriculture, quality, and organization of the market. Though it has been proved that deep learning has been effectively used in the classification of pepper in previous research, the majority of the current research has been centered on a limited number of varieties and relatively small datasets. This paper fills these gaps by providing a novel balanced dataset of 1,000 images of ten pepper types. It was made sure that the data set represented each class equally, thus removing bias and allowing a fair analysis of the performance of the model. Three models were used to evaluate the effectiveness of deep learning to classify pepper breed: VGG16, ResNet50 and YOLOv10. VGG16 and ResNet50 are typical CNN models commonly used to classify images, whereas YOLOv10 is an object detector model that can be used generically to perform classification. The experimental findings showed that ResNet50 and VGG16 had classification accuracy of 96% and YOLOv10 had a classification accuracy of 94.25% with a higher real time inference speed.

List of Publication

M. A. Moni, A. S. Chaitee, K. M. Tanvir, M. A. Rahman, S. Banik and M. T. Habib, "Machine Vision-based Pepper Breed Classification using YOLOv10 and Transfer Learning Models," 2026 6th International Conference on Image Processing and Capsule Networks (ICIPCN), Dhulikhel, Nepal, 2026, pp. 544-549, doi: 10.1109/ICIPCN67432.2026.11438307. (Published)

Letter of Transmittal

25th April 2026

To,

Dr. Md. Tarek Habib

Associate Professor

Department of Computer Science and Engineering

Independent University, Bangladesh

Subject: Submission of Final Project Report

Dear Sir,

It gives us great pleasure to present our project report, "Machine Vision-based Pepper Breed Classification using YOLOv10 and Transfer Learning Models." The completion of this report is a prerequisite for the Independent University of Bangladesh's Bachelor of Science in Computer Science and Engineering program.

Under your guidance, this research proposes a deep learning way to classification of pepper breeds. The objectives of the study were to create a baseline dataset with convolutional neural networks (CNNs) including VGG16, ResNet50 and YOLOv10 models for comparative performance evaluation in between these three models for early accurate computational classification.

We also want to thank you for the kind help and support of each one of you during the whole Project, your encouragement and advice. Your mentorship made this work possible.

Sincerely,

K.M Tanvir

ID: 2220778

Athina Sarkar Chaitee

ID: 2220940

Mahmuda Akter Moni

ID: 2220183

Table of Contents

List of Figures

List of Tables

Chapter-1: Introduction

One of the most widely grown and consumed crops in the world is pepper with some of the varieties being green chili, red chili, Bombay, Shimla and capsicum. These types are highly divergent with regard to shape, size, color, pungency, chemical composition. In addition to their cooking value, peppers are rich in vitamins, antioxidants and capsaicin that give them not only a pharmaceutical but also an industrial value since they are vital in increasing metabolism and immunity. In Bangladesh, pepper has two folds meaning: it is a fine spice, cash crop that is consumed by the rural people in their livelihood. Big production blocks of Bogura, Rangpur, Jamalpur and Rajshahi can provide fresh and dried chili (local markets), processing plants, and export markets. However, the imports have some inclination to be used as a result of the seasonal deficits. These can be attributed to variations naturally, pest damages, post-harvest operations losses and acceptance of impurity added to the supply chain by inefficient supply chains, and low-quality control. All these restrictions highlight the importance of finding ways of overcoming them as far as pepper production, processing and post-harvest are concerned. The developments of AI, machine vision and robotics are offering a new way of possibilities to enable agriculture to enter the 21st century in such a way that it will increase the aspect of food security. Pepper harvesting is highly developing intensive particularly in grading process which is traditionally graded manually since it involves visual inspection which may lead to variation and error [1].

Recent observations have indicated that convolutional neural networks (CNNs) and other computer vision and deep learning systems can be applicable to detect, identify, and determine the quality of the chili variety in real-time with high accuracy. Existence of said machines in the agricultural supply system can lead to the better accurateness of the grading process, reduction of losses, less reliance on human input and uniformity of exports. Therefore, automated pepper classification is the right way to a smarter, more efficient and sustainable spice industry. Interestingly, most of the other studies have focused on either small datasets or one-model designs; our research paper addresses this gap by comparatively analyzing detection-based and classification-based deep learning models on a balanced data of ten pepper varieties.

1.1 Background

Peppers are some of the economically and nutritionally important crops in the world and different varieties are grown and grown in Asia, Europe as well as the Americas. Peppers are a twofold issue in Bangladesh: they are not only a key ingredient in everyday cuisine as a staple spice, but they are also a valuable cash crop that supports rural livelihoods. Bogura, Rangpur, Jamalpur, and Rajshahi are the key production centers, which produce both fresh and dried chili that are consumed nationwide, processed in the food industries, and also exported to both local and foreign markets. Nevertheless, the pepper business in Bangladesh has continued to encounter numerous problems, such as season and supply shortage, pests, after harvesting, and grading discrepancies. These problems lower profitability, decrease the competitiveness of exports, and enhance dependence of imports during lean periods.

The sophistication of pepper production and export is due to the variability of pepper crop. Peppers are very diverse in terms of color (green, half-red, red, yellow), shape (slender chili, bulky capsicum), texture (smooth, wrinkled), and the intensity of the pungency. Manual classification is also complicated by transitional pigmentation, e.g., the transition between the green and half-red and the full red chili (transitional pigmentation). Historically, grading and sorting have been based on human eye inspection, which is subjective, requires labor-intensive as well as being error prone. The incorrect categorization of similar varieties that look alike does not only compromise the quality assurance, but also, causes economic losses to the farmers and exporters. In the case of international trade, discrepant grading usually leads to repudiation of consignments, which harmed the reputation of Bangladesh in the international spice markets.

The most recent developments in the field of artificial intelligence (AI), machine vision, and robotics offer groundbreaking opportunities to the automation of agriculture. When machine vision systems are involved, it recreates the vision of human inspection by taking digital pictures then analyzing using the computational models. When combined with the deep learning models, like Convolutional Neural Networks (CNNs) and object detection systems, like YOLO, these systems are capable of automatically identifying complex visual features and classifying agricultural products with a high rate of accuracy. In contrast to classical machine learning methods, which are based on manually crafted features (color histograms, texture descriptors, shape indices), deep learning models acquire hierarchical representations starting with raw images, and therefore can work in a variety of crop types and environments.

The application of deep learning in the fields of fruit and vegetable classification, disease detection, maturity assessment and robot harvesting has already been successful worldwide. Peppers machine vision can be particularly useful in a situation where they vary in terms of color, shape and surface texture. Automated classification can contribute to decreasing losses following harvest to a minimum, improving the quality of exports and assist farmers to meet the international standards. Moreover, grading machines, conveyor belts and robot harvesters can be equipped with a real-time classification system that can be scaled to farms and supply chains.

Despite these advances, most of the literature prior to this time is limited, focusing on the research of this or that species of pepper or a limited sample. It results in an imbalance in the development of even data and musical evaluation of different architectures. This divide needs to be bridged in designing sound and generalizable systems that have the capability to handle diversity of pepper breeds which are witnessed in the agricultural environment. Another goal of the given study is to put pepper classification into the perspectives of the general development of smart agriculture by comparatively revealing trade-offs between accuracy, faster inference and deployability through exploring classification-based (VGG16, ResNet50) and detection-based models (YOLOv10).

Lastly, we have the history that underlies this research and that is coming together of farm necessity and technology. Peppers are an essential component of the Bangladeshi diet, along with the rural economy, but automation of peppers can lead to socio-economic payoffs that can be achieved in the short run. The machine vision-based classification is an access point to smarter, more efficient, and more globally competitive agricultural practices, and it aids in eliminating a rift between agricultural practices in the past and the present, which is backed up by AI-powered solutions.

1.2 Motivation

Peppers are not only one of the common spices in the Bangladeshi cuisine, but they form a significant cash crop, on which the population in the countryside depends. Being dependent on pepper farming as an export crop and as a household use crop, such districts as Bogura, Rangpur, Jamalpur and Rahsahi rely on it. However, the industry is still confronted with long-term problems of seasonal deficit, pest infestations, post-harvest losses and un-reliable grading. These issues lower the profitability, decrease the relevance in exports and increase the dependence of imports during lean times.

Manual classification and grading being based on human eye inspection is highly labor-intensive and subjective. The mutual imprecision, which occurs in the classification of similar varieties which are aesthetically equivalent such as half red chili and green chili, is normally misclassified and thus quality assurance is also poor and there are market failures. This will be an international loss of credibility and rejection of the consignments by the exporters. This is bad news to the farmers, who will end up having low prices and wages.

The aim of this work is inspired by the need to bridge the gap between the traditional and contemporary AI-based solutions in agriculture. Machine vision and deep learning can also be used to automate pepper classification, be more precise, have consistency, and scale better. This reduces the human dependency as well as ensuring that the same quality standards are maintained in the supply chains.

In addition, the comparative analysis is propelled by the practicality of deployment (comparative analysis of classification and detection-based models according to VGG16 and ResNet50, respectively). As compared to the classical models, dynamic farm and market environments necessitate application of real time detection models. Practically adopted systems to the farmers, traders and exporters need knowledge of the trade-offs in accuracy versus the speed of inference and the robustness.

Last, the two sources of inspiration of this study are twofold:

- Scientific: To further develop the research on the theme of agricultural machine vision by proposing a balanced sample of ten pepper varieties and applying different deep learning structures in a controlled manner.

- Socio-economic Motivation: To provide the farmers and exporters in Bangladesh with technology that will enhance precision in grading, reduce losses and enhance competitiveness in the global spice markets.

1.3 Objective

The main aim of the work is to reach a new level of machine vision and application of deep learning in recognizing agricultural crops with certain consideration of pepper breeds. The following objectives facilitate this overall objective:

Dataset Development

- Out of every 1,000 pictures you can have 10 kinds of pepper variants.
- The equal classes: Assure that the lack of favoritism is established and be certain that generalization is attained.
- Preprocess to ensure that the samples are consistent like resizing, normalization and RGB standardization.

Model Implementation

- Train classification-based (VGG16, ResNet50) and a detection-based (YOLOv10) architecture.
- Image: follow the scheme of converting the YOLOv10 model to the classification task in order to be equally compared to the classical CNN models.
- Divide models into systems (200 and 800 training images and 200 and 800 validation images respectively).

Performance Evaluation

- Compare the models on the basis of such trivial measurement as the accuracy, precision, recall, F1-score, the false positive rate (FPR) and the false negative rate (FNR).
- Create confusion matrix in order to ascertain trend in misclassifications especially between similarly looking pepper species such as half-red and green chili.

Comparative Analysis

- Discuss advantages and disadvantages of classification based and detection-based methods.
- Display trade-offs in accuracy, speed and deployability of applications in agriculture.

Deployment Considerations

- Finally, Compare Applicability to a practical application of each of the two models like automated grading, sorting and robotic harvesting.
- Comments: What are your realistic suggestions as to how machine vision systems can be used in pepper supply chains to provide quality management and competitiveness to exports

Chapter – 2: Literature Review

Machine vision and machine learning in agriculture sector have shown an exploratory growth over the past few years and have given more attention to the peppers than the others due to their economical and nutritional value. Past procedures were variation of classical concept of processing in digital image and were applied to the manual characteristics (color, texture and shape) which tend to change when the light conditions, the resolution and transitional pigmentation change. To deal with these limitations, increasingly recent literature has been considering a broader, larger ensemble of deep learning models; including, convolutional neural networks (CNNs) with object detectors models, such as YOLO.

The world has seen numerous research projects done regarding various architectures and image-based frame-works in order to improve the level of accuracy, robustness, and adaptability of pepper breed recognition and classification systems.

2.1 Prior Studies

Invented a machine vision smart modelling system to sort bell peppers in-line to meet export specifications. The LDA and MLP classifiers were compared in terms of color, texture and size characteristics of bell peppers. The MLP classifier had a remarkable accuracy of 93.2 at a throughput rate of 0.2s/sample [2]. Proposed a deep learning instance segmentation system to phenotype pepper in the field. Their approach with Mask R-CNN was much higher in performance than classical image processing and reached mAP values of 0.97 on Blocky Bell pepper and La muyo pepper, exhibiting high robustness and generalization capacity across several different pepper varieties [3].

Already give an automatic chili pepper phenotyping system with YOLOv7 to detect fruit and seeds. They had achieved 0.92 and 0.87 precision and mAP respectively. It captures morphological features such as seed numbers, fruit dimensions, color levels, and surface area in support of breeding and genetic resource management of pepper [4]. Suggest a computer vision-based YOLOv5 model that would be used to classify bird’s eye chili in real-time. An example of a Raspberry Pi 4B system employed with a robotic manipulator to sort red/green chilies was shown, with the obtained results being 0.94 and 0.90 mAP, accuracy using 739 labeled images [5]. Recognize and categorize chili fruits (green, red, and rotten) with an accuracy of over 93% using YOLOv5 on a dataset of 300 images. It puts emphasis on YOLOv5 and its effectiveness in real-time use of agricultural robots, to achieve proper localization and grading of chili fruits [6].

Presents machine vision techniques are used in classifying and grading fruits considering color, texture, size, and shape characteristics. It introduces feature extraction algorithms like SURF, HOG, LBP, and machine learning algorithms like KNN, SVM, ANN, and CNN, with their pros and cons, and issues with automation of the quality of fruit inspection [7]. Autonomously classify pepper varieties using the YOLOv8m deep neural network. A single YOLOv8m model was used to classify eight varieties of Capsicum using 1,476 labelled images with the results showing that data augmentation improved accuracy, recall and mAP [8]. Suggest a robotic gripper solution using transfer learning (VGG-16, ResNet50, GoogleNet) to automate the process of bell pepper quality detection. VGG-16 achieved the highest accuracy of 98.38 and outperformed both ResNet50 and GoogleNet which reached 90 and 97 accuracy, respectively, hence it is capable of classifying fresh and unfresh peppers in post-harvesting sorting [9].

Present YOLOv8-CBSE a lightweight improved model to identify the maturity of chili peppers in natural settings. An initial YOLOv8 CBSE that included the ConvFormer, SRFD, and DRFD blocks and EIoU loss had a mAP of 88.08% on 1390 field images [10]. Compare in detail banana breed recognition with the traditional machine learning and deep learning methods. To evaluate seven deep learning models and five classical ML models we used 5100 banana images. The trained CNN had the highest accuracy (99.37) over the pre-trained nets and classical methods, which supports the strength of the customized CNNs in the recognition of agricultural animals breeds [11]. Trained an improved YOLOv4 model to detect pepper targets during the daytime. The base (88.95) was lower than the improved YOLOv4 Mosaic with CBAM attention of 98.36% mAP on 500 pepper images. In this way, the precision and effectiveness of pepper detection is enhanced, facilitating agri-automation [12].

Introduced a smaller GoogLeNet-EL network that has better results on pepper leaf disease detection. An Inception model that was spatially pyramid pooled achieved 97.87% precision with pepper leaf disease on 9,183 images, 6 percent higher than the prior versions of IncV2. It was also able to be used in a portable application in a large pepper farm with a ResNet-50 as the model due to its small memory requirement (10.3 MB) and quick inference time relative to AlexNet, ResNet-50, and MobileNet-V2 [13]. Introduced Mask R-CNN as the backbone based on ResNet-101 + FPN to locate green sweet pepper fruit and peduncle under greenhouse conditions. Model precision was 84.53% and 71.78% fruit precision and peduncle precision respectively and mAPIoU = 50 of 72.64, which means that there is robust real-time segmentation under occlusion conditions and under complex light intensity conditions [14]. Introduce a papaya disease identification system that uses k-means clustering to partition the data and support vector machines (SVM) to classify the data with a high accuracy of over 90% [15]. Deep learning has the potential of advancing automation in agriculture with emphasis given to pepper varieties classification. With the YOLOv8m model and a set of 1476 annotated images augmented with data augmentation, the researchers reported significant gains in precision, recall and mean average precision. The findings demonstrate the usefulness of YOLOv8m to discriminate pepper in real-time and create a feasible benefit in quality assessment after harvesting and food chain [16].

Red beet study developed a three-phase study with the YOLOv9 and YOLOv10 models that showed that larger sizes and the higher the resolutions, the fewer false negatives and the better the weed is detected to execute an accurate wedeling. To improve the accuracy of the variants of YOLO pose that already exist, the soybean paper suggested YOLO_SSP, a variant of the critical point detector, which has a focus on the detection of the stem node, and can be used to predict yields and breeding. The whitefly paper presented an AR glass segmentation then-detect-based system using an optimized YOLOv11, which enables them to accurately estimate the population of the adult and late-instar nymphs in the complex field settings. Taken together, these papers indicate that sophisticated deep learning models are changing precision agriculture, both in terms of controlling weeds and in terms of monitoring and detecting pests and traits [17], [18], [19].It introduces MS YOLO (better than YOLOv10) on small defects detection to the PCB paper, achieving almost 99% mAP at 1.5 and with edge strengthening, SPDConv, and dual distillation. The citrus paper presents YOLOv11 RDTNet, a compact model which reduces parameters by more than 40 percent and is more effective in recognizing pests and diseases. They combine to demonstrate how industry and agriculture applications of tailored YOLO upgrades achieve high precision in an efficient manner. [20], [21]. The proposed medicinal plant paper is an improved ResNet50 with progressive transfer learning and optimized SVM that attains 96.8 percent accuracy in for a dependable classification with leaf in the future. Cucumber, PCB, and citrus papers demonstrate that lightweight variants of YOLO achieve high-precision, efficient vision models to use in sorting, defect detection, and pest recognition, demonstrating the applicability of deep learning to agriculture and industry [22], [23].

It is observed in the rice disease experiment that a modified ResNet50 is 99.75% accurate in detecting rice blast at an early stage, exceeding Inception V3, VGG16, VGG19, and CNN, and executed through a Gradio web app to enable empirical application. In Malaysia, crop disease detection research compared VGG16, ResNet50 and EfficientNet B1 with tomato and potato leaf samples, with EfficientNet B1 showing the highest utility (99.45 percent accuracy) which were statistically validated and had a web interface to perform real time diagnosis. Combined, they support the idea that state-of-the-art CNN architectures and transfer learning provide high accuracy, useful agricultural disease management tools [24] [25].The mammography images were tested using VGG16 and ResNet50 in the breast cancer study with a reported accuracy of 91.23 per cent and 99.01 per cent respectively, highlighting the importance of the two in early cancer detection. IVGG13, a variant of VGG16, paper on pneumonia, that uses less depth and data augmentation, was found to perform better than LeNet, AlexNet, GoogLeNet, and standard VGG16 on data on chest X ray. The rice disease paper showed that a trained ResNet50 with modifications performed 99.75 percent accuracy on early rice blast detection and was much more accurate than Inception V3, VGG16, VGG19, and CNN, and was implemented using a Gradio web application. Brain tumor model compared AlexNet, VGG16 and ResNet50 on MRI images with Alexnet having an accuracy of 96.10 percent which was higher than the others [26], [27], [28]. ACP Tomato Seg is a better YOLOv8s seg model with AOFRM, CMPrd and PSA modules that improves feature extraction, multi scale detection and attention focus. It made significant increases in both bounding box and mask tasks mAP scores, better than the baseline. The model was also demonstrated to be significantly generalized on a strawberry dataset, which is useful in accurate fruit segmentation and automated fruit harvesting [29].

The potato leaf disease experiment reported a high accuracy (98.22%) with an ensemble of VGG16 and ResNet50, showing better results than the single models (96.16 and 95.68). The affect detection experiment was a comparison of VGG16, ResNet50, and SE ResNet50 against facial images whereby ResNet50 achieved the best validation accuracy of 99.47. Collectively, these papers demonstrate that

CNN assembles and promotes architectures produce highly precise classification with agricultural disease detection and human affect detectors [30] [31].The grapevine study proposed a faster Faster R CNN (Enhanced VGG16) into a UAV where a quadcopter takes pictures of leaves and a hexacopter is used to spray pesticides. It had the highest accuracy of 99.6% and it was more practical than AlexNet, GoogLeNet, ResNet 50, and normal vineyard VGG16 [32].The pulmonary nodule experiment made the ResNet50 Ensemble Voting model with CT images of nodules under 20 mm achieving 94.3% accuracy, sensitivity of 96.4 and specificity of 91.1. This method had great promise of helping clinicians in diagnosing lung cancer early [33].

EfficientRMT Net consists of ResNet 50 and Vision Transformers (ViT), combined with depth wise convolution layers to classify potato leaf diseases, reaching up to 99.12% accuracy, compared with traditional CNN-based techniques. The ResNext50 LSTM hybrid model resolves the problems of vanishing gradient and overfitting by restraining feature growth and storing pertinent features to be learnt over the long term with a 95.44% validation accuracy on the PlantVillage dataset [34] [35].ResNet50 was the most successful in the tomato leaf disease detection task, as it obtained the highest accuracy of 96.51% and exhibited good prospects in real time IoT based crop monitoring. A framework based on YOLO to extract the leaf with ResNet50 to classify the leaf achieved strong performance in 5,600 images in soybean disease detection, surpassing other CNNs. Collectively: The two papers emphasize the accuracy and scalability of ResNet50 to precision agriculture, which can be used to timely take interventions to minimize crop losses and enhance productivity [36] [37].

2.2 Scope of Study

The research will focus on the creation as well as the trial of a machine vision-based system to classify pepper breeds with deep learning models. The paper has provided a balanced set of 1,000 images of ten breed of pepper that has balanced over a slim scope of slightly limited literature, and includes no class imbalance, unlike the literature that used to concentrate on a single breed of pepper. By integrating two classification-based architectures (VGG16 and ResNet50) and a detection one (YOLOv10), the study benefits itself by having the advantage of two different approaches. Those does not only involve simple measures of accuracy estimation but also entails precision, re-call, F1-score, FPR, FNR and the examination of the confusion matrix thus offering an in-depth view of the model's performance. The misclassification issues in particular in the types which are visually close to each other, e.g. half-red and green chili are considered as the issues which appear in the real world. Moreover, the scope will also entail comparative benchmarking of previous agricultural vision research and pepper classification scope will be in a larger picture of smart agriculture research. Lastly, the paper shows the areas of deployment to practice, inferences speed and strength in practical detecting the feasibility of the usage of inferences towards real-time applications as automated grading, sorting and robot harvesting.

Sensu stricto, the scope of the study will be scientific which would provide an opportunity to enhance the future of the agricultural machine vision, socio-economic, which would provide the opportunities to enhance quality control, post-harvest losses, and prepare the pepper supply chain in Bangladesh to be more competitive.

Chapter – 3: Methodology

The chapter explains the methodological workflow to be used in the classification of pepper breeds based on deep learning models. The research was conducted in such a way that it was fair and reproducible due to use of balanced data, systematic pre-processing, and comparative analysis of various architectures. There were three models used: VGG16 and ResNet50 as classification-based CNNs [38] and YOLOv10 as a detection-based model used as classification.

3.1 Proposed Methodology

The pepper breed classification model based on the proposed system is a deep learning pipeline. It is composed of six stages that are related. The 1st phase is capturing pictures of ten varieties of pepper. We control visual clarity and consistency in visual signs such as color, texture and shape. These unprocessed images were then subjected to a series of preprocessing steps. It ensured the homogeneity and diversity of the dataset. In the following step, the data set was labelled and categorized manually and sorted into directories with equal classes of all classes, ensuring that the pepper variety is equally represented to train it more evenly and balanced. We trained the deep learning models such as YOLOv10, ResNet50 and VGG16 using the dataset that we labeled manually. The accuracy, precision, recall, and F1-score were used to measure performance of the model using conventional indicators. The confusion matrix analysis was conducted as well to evaluate the quality of prediction on a case-by-case basis. This tiresome work structure would allow the system to effectively categorize pepper breeds on the basis of looks, and hence, the prospect of the machine would show the vision in the agricultural categorization undertaking. Every stage of the processing as presented in Figure 3.1.1.

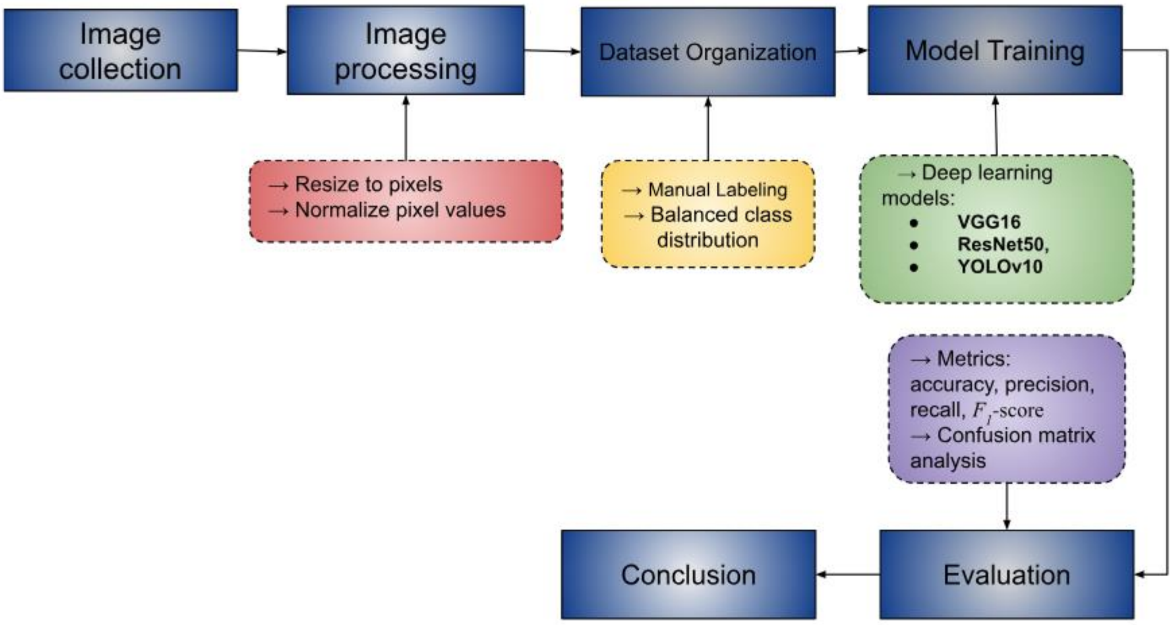


Figure 3.1.1:Diagram of the working process for the classification of pepper

3.2 Data Description

Data on various varieties of peppers was collected using systematic data. The selected peppers figures were taken under controlled conditions then filtered and preprocessed to make the modality the same and of equal size and resolution. The images were in various resolutions, to this end, the images were scaled to give Breeds a uniform result of the dataset. The data set used to identify pepper varieties by use of machine vision is of images of 10 varieties of pepper in Figure 3.2.1, which are used in South Asian markets and agricultural systems.




Red Bomboy Green Bombay Yellow Shimla Red Shimla




Green Deshi Morich Red Deshi Morich Half Red Deshi Morich Green Capsicum


Red Capsicum Yellow Capsicum

Figure 3.2.1: Pepper Variety Samples

The images were all organized into a single file which was referred to as pepper dataset and the images were systematically placed so that they are well-organized of similar distribution. This was needed to utilize the image classification techniques appropriately. Table 3.2.1 accommodated each variety of pepper with the same number of samples, thus resolving the issue of class imbalance. This equal distribution enables the classification system to learn without being biased and enhance its ability. The peppers were organized into 10 categories. Which can be characterized using visual patterns which include color, texture, shape, and grain size.

Table 3.2.1:Samples of each pepper variety

Variety name	Capture Image
Red Bombay	100
Green Bombay	100
Yellow Shimla	100
Red Shimla	100
Green Chili	100
Red Chili	100
Half-Red Chili	100
Green Capsicum	100
Red Capsicum	100
Yellow Capsicum	100
Total	**1000**

3.3 Dataset organization

For effective training and benchmarking of the model, the dataset was split into 2 different group with an equal number of 10 classes in each: 1. Training; 2. Validation.

The structure of the whole dataset is mentioned below: training set- 800 images and validation -200 images. This was conducted in a systematic way where each subset included a balanced sample of all pepper varieties in Table 3.3.1 hence making it achievable to perform unbiased assessments.

Table 3.3.1:Description of pepper classes in the Dataset

Red Bombay	Deep red color; elongated shape; glossy surface with medium grain size.
Green Bombay	Bright green color; similar shape to Red Bomboy; smooth texture.
Yellow Shimla	Bulky bell shape; vibrant yellow color; thick walls and smooth skin.
Red Shimla	Bell-shaped; rich red color; thick and glossy surface.
Green Chili	Slender and curved; dark green hue; rough texture with pointed tip.
Red Chili	Similar to the green variant; deep red color; slightly wrinkled surface.
Half-Red Chili	Transitional color (green to red); mixed pigmentation; uneven texture.
Green Capsicum	Large bell shape; bright green; smooth and shiny surface.
Red Capsicum	Bell-shaped; vivid red color; thick and glossy skin.
Yellow Capsicum	Bell-shaped; golden yellow color; smooth and firm texture.

These pictures of pepper were captured in different resolutions, indicating a difficulty with the deep learning models with a fixed input size. To alleviate this weakness, the images were preprocessed so as to make them consistent throughout the dataset. RGB image format, a three-channel color depth, and heterogeneous initial resolutions are the main peculiarities of the dataset. To match the resolution of YOLOv10, 224x224x3 pixels of ResNet50, and VGG16, images were resized. This resizing saw to it that there was uniformity in height and width. The dataset will be more efficient to train as it can be trained with a compatible and constant input size and enhances the performance [39].

3.4 Methodology model Used

Convolutional Neural Networks (CNNs) are a special class of feed-forward neural networks with convolutional layers that learn features from raw images [40]. In this work, three CNN-based models were applied namely VGG16, ResNet50 and YOLOv10. Traditional architectures of image classification (VGG16 and ResNet50) provide a single label to an image based on the visual features extracted. This study was conducted using YOLOR10, which was originally an object detector model, with slight adaptations to classify pepper breeds. This adaptation allows justifiably comparing with VGG16 and ResNet50 as well as showing the ability of YOLOv10 to detect objects in real-time. The emphasis on this methodological difference explains the synergistic advantages of classification-based and detection-based strategies in the larger problem of pepper breed classification. The overall structure of a CNN model is illustrated in Figure 3.4.1.

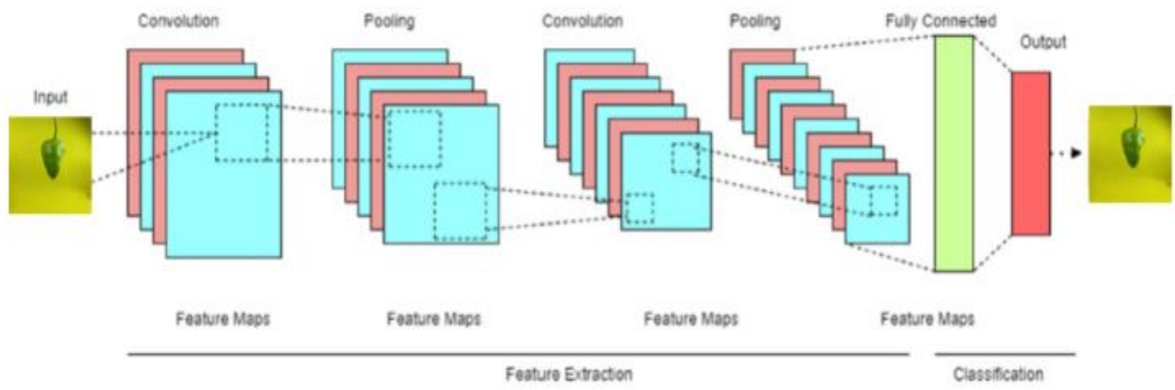


Figure 3.4.1: System Architecture for Pepper Breed Classification *[11]*.

3.4.1 VGG16

Visual Geometry Group - VGG16. VGG-16 is a widely used convolutional neural network structure reported to have outstanding accuracy in image Classification tasks [41]. This model is very deep with the neural network layer being 16. In addition, it consists of convolutional, pooling and fully connected layers. The VGG16 employs small 3x3 filters to acquire particular information on color and texture visual patterns. It is made up of a single input, single output, thirteen convolutional, five pooling, and three fully connected layers [42]. The width decreases and the number of channels grows with the depth of the layers. The convolutional layers that were smaller were trained with varying parameters by VGG 16 over the training period of the boosting process [43], [44].

The significant features of VGG16 are:

- LEVEL and Shallowness: VGG16 contains 16 layers, arranged in a simple and consistent architecture and consequently with straightforward implementation and analyses.
- Small filters: 3 × 3 convolutional filters are used and this can capture finer details of texture and color patterns.
- Stable baseline: A high-accuracy model in image classification tasks and a computer vision benchmark model.
- Limitation: Hard to compute, consumes lots of memory and inferences are slower than newer lightweight models.

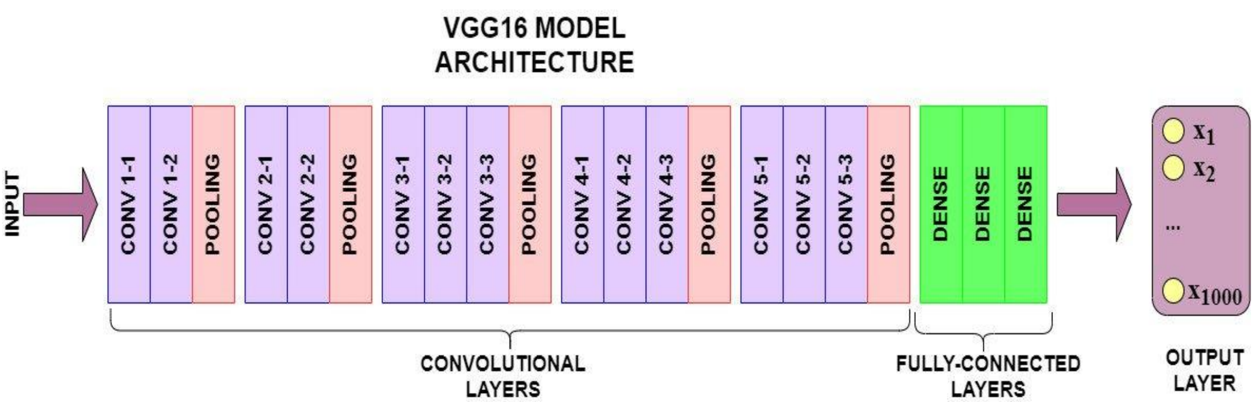


Figure 3.4.2:VGG16 Architecture

3.4.2 ResNet50

Despite this, ResNet50 is a step ahead in the CNNs as they suggested use of residual connections to overcome the issue of vanishing gradient in deep learning models. ResNet50 despite having done so with 50 layers can learn complex feature representations on a hierarchical level without being unstable even during training of the model. Residual layers in ResNet50 are equally important where large value through gradient being passed to its previous layer [45]. This is with skip connections that help the network leverage the past feature representations, therefore the network is better able to classify and avoid misclassifying visually similar less familiar species of pepper, that are red and half red chillis.

The key characteristic features of ResNet50's are:

- Depth and Flexibility: 50 layers, is able to differentiate hierarchical features from simple edges to more complex textures and shapes.
- Residual learning: Skip connections alleviate the vanishing gradient issue and improve convergence in training.
- Good generalization: ResNet-50 pretrained on large datasets like ImageNet generalizes effectively to new areas, and features can be extracted effectively.
- Limitations: Smaller networks take shorter time to train and less computational resources.

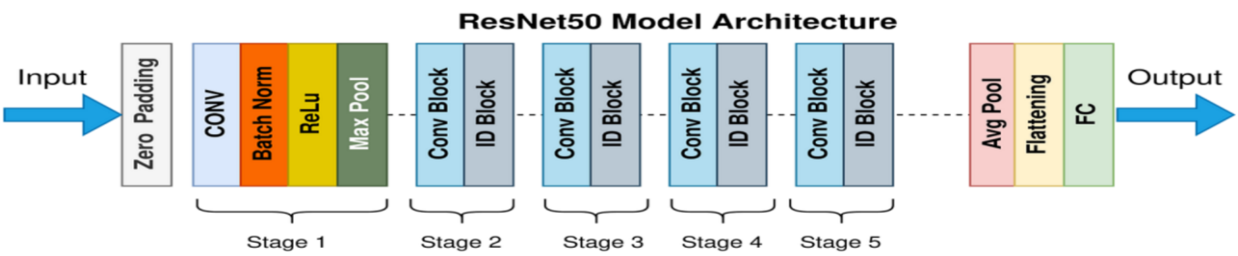


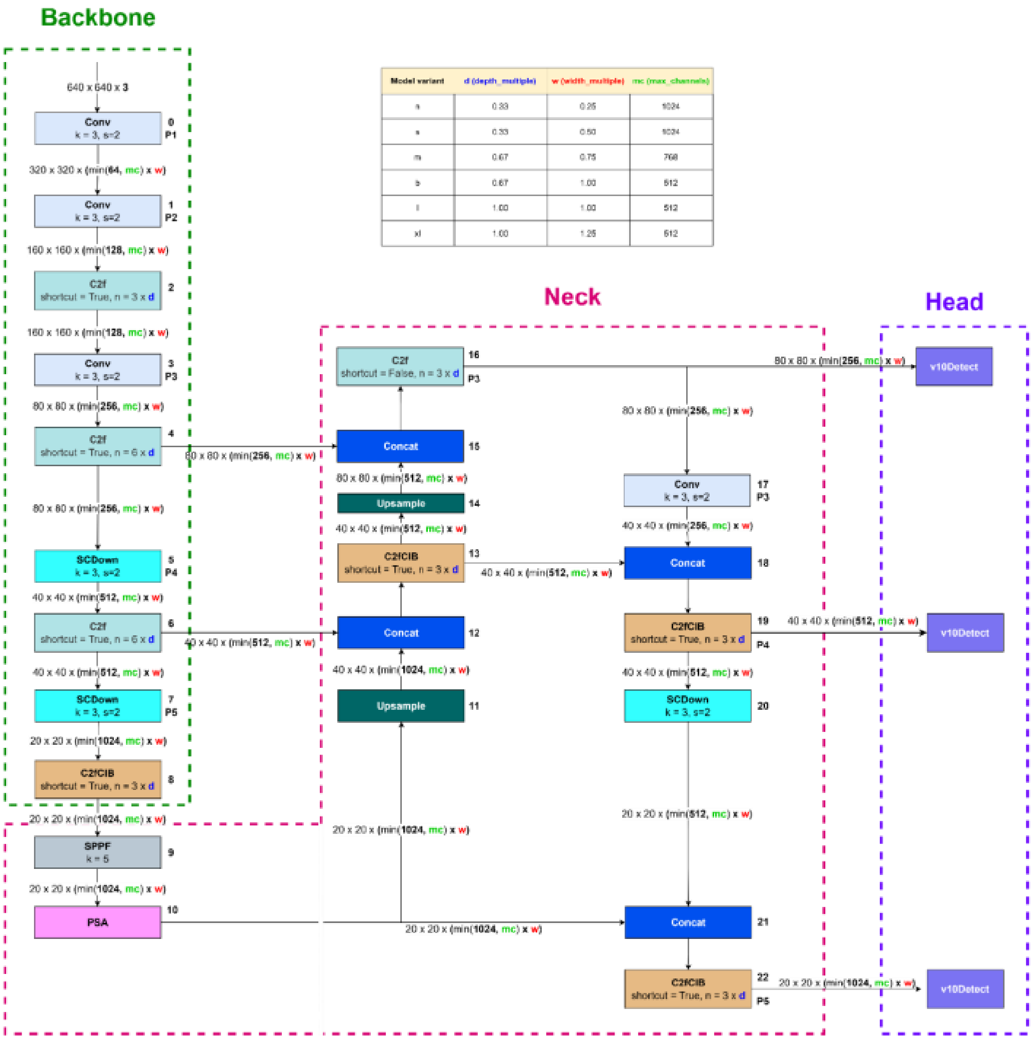
Figure 3.4.3:ResNet50 Architecture

3.4.3 YOLOv10

YOLOv10 is an advanced real-time object the detection model, which has delivered exceptional performance in terms of endpoint latency and model size [46]. Unlike the other two models, ResNet50 and VGG-16, which only aim at classifying information, YOLOv10 is structured to do both localization and classification of objects in a single forward walk, thus making it a good model to sort and grade various peppers in an image. It can detect various peppers in spite of lighting variations as well as variations in background due to its capacity of addressing the challenges of real-time object detection [47] in addition to having an efficient architecture.

The key features of YOLOv10 are:

- Real-time detection: Localization and classification are performed in a single forward pass, which is suitable to real-time inference in a dynamic agricultural setting.
- Efficiency: Has been optimized to be low-latency and scalable, and is thus feasible to deploy in real-world grading and sorting operations.
- Robustness: Works well with image variations in lighting, background clutter, and occlusion.
- Restriction: It is slower, but its classification accuracy might be a little less than that of classical CNNs when used in controlled datasets.



Model variant	d (depth_multiple)	w (width_multiple)	mc (max_channels)
n	0.33	0.25	1024
s	0.33	0.50	1024
m	0.67	0.75	768
b	0.67	1.00	512
l	1.00	1.00	512
xl	1.00	1.25	512

Figure 3.4.4:YOLOv10 Architecture

3.5 Performance Metrics

A separate test dataset was used to measure the performance of the proposed pepper breed recognition models. Accuracy is not necessarily a good measure to indicate the performance of a model and other measures that may be applied include Precision, Recall and F1-score. Whereas in the multiclass case the confusion matrix is a matrices of n x n, that is, a matrix of numbers, and n is the number of classes. The individual 2x 2 matrices are calculated to obtain the average performance measures and precision, recall, and F1-score and accuracy of each model can be obtained.

Precision refers to the direct quantification of the accuracy of positive predictions of a model. It gives the probability of the model being right in predicting a sample to fall under a particular category. The high accuracy in the context of pepper breed classification means that on the occasion the system recognizes a pepper as red chili, it will be highly accurate. This minimizes false alarms and improves trust of automated grading systems. Precision is the ratio of positive samples which are called positive of the total samples that are called positive, division gives:

$$Precision = \frac{TP}{TP+FP} \tag{1}$$

Recall or sensitivity is a measure of the model to accurately classify all the actual positive samples. It plays a vital role in the agricultural applications since false cases may be missed and the incorrect classification will be made and quality will not be regulated. In pepper recognition, high recall is used to ensure majority of peppers of a certain breed are identified, eliminating chances of missing varieties such as half-red chili that are physically similar to the green chili. Recall is a measure of the fraction of true positives that are identified correctly, and is defined as:

$$Recall = \frac{TP}{TP+FN} \tag{2}$$

F1-score is a balanced measure of both precision and recall since they are harmonic to each other. In contrast to the arithmetic mean, the harmonic mean punishes drastic variations between precision and recall [48], so the F1-score is especially useful when the data are skewed, or when faults on the false side and inaccuracies in the true side have practical implications. The F1-score of the pepper classification is the score that indicates the ability of the model to determine which errors to reduce to detect as many as possible. Harmonic mean of precision and recall is offered by F1-score, which is defined as:

$$F_1 - \text{score} \ = \frac{2 \times Precision \times Recall}{Precision \times Recall} \tag{3}$$

One of the most popular metrics in classifier evaluation is accuracy, which gives a very simple measure of the overall accuracy. It can be defined as the ratio of the right predictions of the samples to the total number of samples. Mathematically, the accuracy can be said to be:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{4}$$

False Positive Rate (FPR) is a percentage that characterizes the percentage of negative samples wrongly identified as positive. Practically, this shows the frequency of false alarms by the model. As an example, when a green chili is incorrectly identified as a red chili, this will contribute to the FPR, a factor that can lower confidence in automated systems. It is a measure of the proportion of negative items that were wrongly identified as positive:

$$FPR = \frac{FP}{FP+TN} \times 100\% \tag{5}$$

The rate of the positive samples that got falsely identified as negative is known as False Negative Rate (FNR). It shows up missed detects that can be a very important factor in the agricultural setting. In order to demonstrate this, FNR is large when a red chili is mistaken to be a green chili to show that the model did not recognize that there was a true positive case. Collectively, FPR and FNR can help to better understand the nature of errors that a model is committing, and can be extended to accuracy, recall and F1-score to construct a general assessment model. It is a proportion between cases of positive samples that have been termed as negative:

$$FNR = \frac{FN}{FN+TP} \times 100\% \tag{6}$$

As a combination, the recall and the F1-score are improved scores under deep learning models as compared to the accuracy score without the combination of both scores since examining each score alone excludes the other, which shows that deep learning models have a score of 100 and similar performance across each section [49].

Chapter -4: Experimental Evaluation

This chapter describes the experiment, which makes use of the proposed deep learning models, to classify peppers based on their breed. It had performed the experiments on the basis of equalized dataset (1,000 images of ten type of pepper) split into one publication (800 training images) and into another publication (200 validation images). VGG16, ResNet50 and YOLOv10 models were used to at least give a comparative a semblance of equity by being implemented on the model. The performance was measured using the accuracy, precision, recall, F1-score and the confusion matrices and the results of the batch labeling and prediction were further processed. With these analyses, the strengths and weaknesses of each of the architectures are quantitatively and qualitatively analyzed in terms of the tradeoff between accuracy, inference speed and performance to the real-world agricultural conditions.

4.1 Comparative Analysis

Table 4.1.1 shows the performance of three deep learning models (i.e., VGG16, ResNet50, and YOLOv10) in terms of accuracy, precision, recall and F1-score. The overall accuracy of VGG16 and ResNet50 was very similar (96 percent). VGG16 an image-wise achievement was more stable compared to the accuracy (96.3) and recall (96.0) and F1-score (96.0) which shows the strength of the system based on its predictive faculty. ResNet50 As well, had a better accuracy (96.8), yet a good trade-off (95.9) in F1-score showing that the relationship is good with fewer false negative. YOLOv10 possessed the lowest accuracy (94.25) yet equal precision (96.32) and F1-score (94.23) and a satisfying performance in terms of what classes of pepper.

VGG16 and ResNet50 were particularly accurate on all pepper varieties which means that they can be applied in controlled classification. This has been found to be of great importance to the agricultural applications since it minimizes the chances of systematic misclassification and offers confidence that the results would be very accurate in terms of quality control and conservation. In the meantime, YOLOv10, with slightly lower overall accuracy, had a higher inference time and was also more stable when it came to visually comparable breeds like half-red versus green chili.

Table 4.1.1:Classification Performance of All Deep Learning Models

Model	Accuracy (%)	Precision (%)	Recall (%)	F_1-*score*	FPR (%)	FNR (%)
VGG16	96	96.3	96	96	0.44	4
ResNet50	96	96.8	96	95.9	0.44	4
YOLOv10	94.25	96.32	92.22	94.23	3.68	7.78

In addition to precision-oriented metrics, speed of inference was also considered to assess feasibility of implementation. To illustrate, the VGG16 and ResNet50 models are highly precise, but consume much more computing power and are slower than networks outlined in this paper with reduced convolution stages. Conversely, YOLOv10 is an efficiency-focused architecture that promises to run faster inference and be able to perform real-time detection. Altogether, VGG16 is the most stable model in all metrics of stability whereas ResNet50 offers high accuracy and equitable performance. Though YOLOv10 is not very precise, it has some practical benefits when used in detection within a real time. This approach can be an interesting option, speed, and flexibility are crucial.

4.2 Visualization of Pepper Classification Across Models

The images of the batch labels depict the step of dataset preparation as each of the pepper varieties was labeled with the bounding boxes and the class name. This visualization affirms that every ten of the classes such as Green Chili, Red Bombay, Half Red Chili, Shimla and Capsicum were always represented in training. Analysis of these named batches allows seeing the clarity of the dataset organization and the clarity of the boundaries of the classes. This was necessary in order to achieve a balanced learning and bias reduction, whereby the models were given equal representation of each pepper variety.

(Figure 4.2.1) the results of the trained models in the inference. All images feature peppers with bounding boxes and predicted labels as an indication of how VGG16, ResNet50, and YOLOv10 categorised the various varieties of peppers in various visual setups. The strong performance of the architectures is indicated by correct predictions, whereas the occasional misclassifications especially between transitional types like half red chili and green chili are related to the findings of the confusion matrix. Such visualizations affirm that CNN based models (VGG16 and ResNet50) were superior on overall accuracy, but that YOLOv10 was superior in a faster inference and more stable in processing visually overlapping classes.

Confidence scores and the bounding boxes and labels (Figure 4.2.2). The values offer a further analysis level to show the reliability of every forecast. When the confidence score is high (nearing 1.0), the model reliability is good whereas a lower score reveals instances in which the visual similarity posed difficulties to the classifier. This visualization can be especially relevant when deployed to the real world, as it will tell not only what was predicted by the model but also how certain the model was of those predictions.

Together, these visualizations work to bridge the divide between numerical measures and implementation in the real world. They demonstrate how classifier-based models and detector-based models complement with each other. CNNs are precise in well-regulated settings involving the controlled conditions domain, but YOLOv10 allows flexibility and speed at an applied real time agriculture scenario. Under this heading, labeled and predicted batches are presented; such as create an integrated view of how effective the system is and solidified the role of machine vision in automated assessment, classification and monitoring of peppers supply chains.

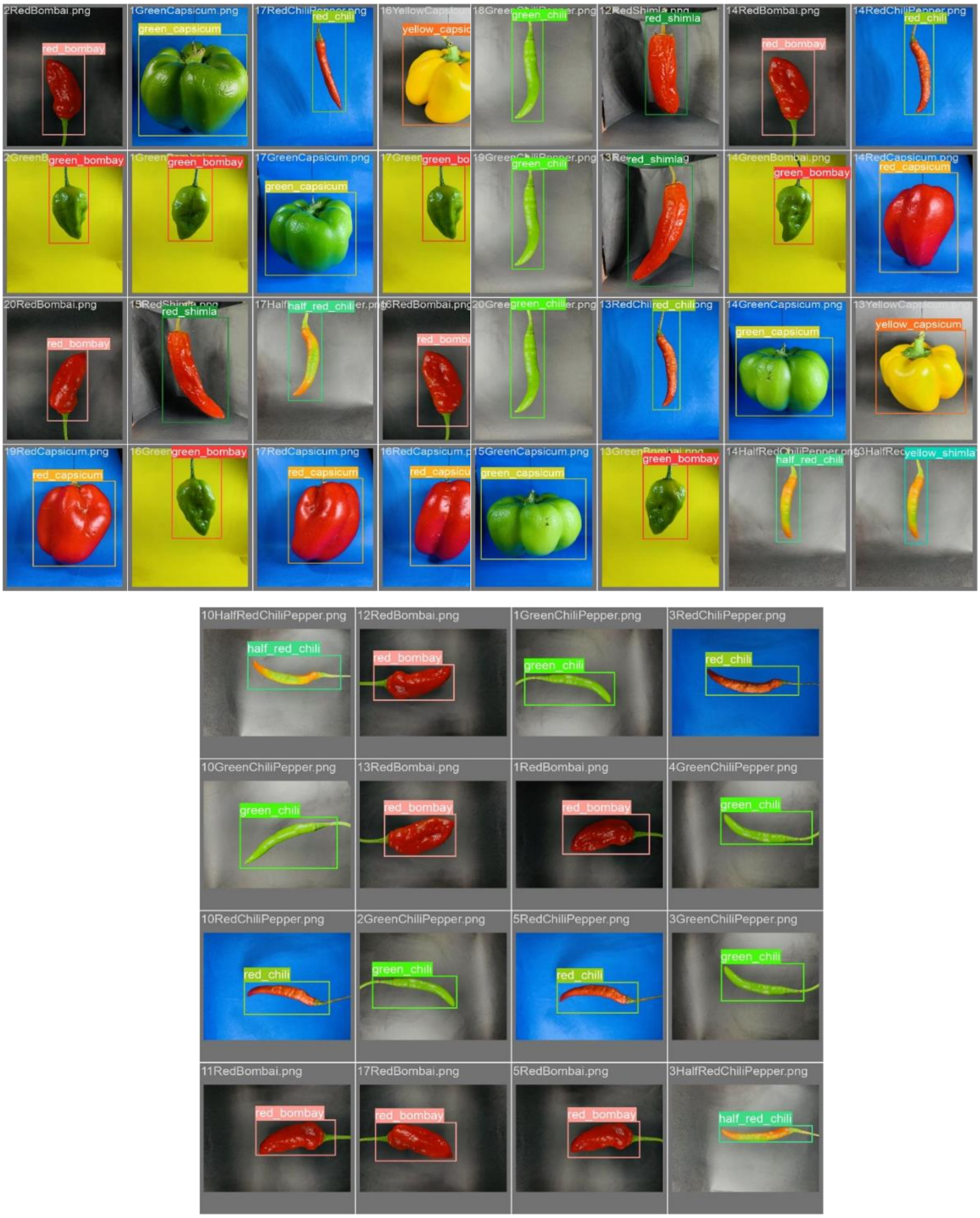


Figure 4.2.1:Batch prediction outputs with bounding boxes and labels, showing correct classifications.

Figure 4.2.2:Batch prediction results with confidence scores, indicating prediction reliability.

.

4.3 Performance Curve Analysis of YOLOv10 Model

The performance curve analysis helps to better understand how YOLOv10 behaves in terms of classification behavior with the variant of changing confidence thresholds. The precision-confidence statement indicates that most pepper varieties perform near-perfect in terms of precision whereas transitional interbreeds like half-red chili are tricky. The F1-confidence curve indicates that the model is most effective at moderate levels of confidence, where precision and recall have a good balance. The recall confidence curve shows the trade-off that exists between total detection at low confidence and conservatism at higher levels of predictions. Lastly, the accuracy-recall curve confirms the strength of YOLOv10 whose total mean of 0.5 mAP is 0.975 which proves that it is suitable to use in real time in the agricultural context.

4.3.1 Precision–Confidence Curve

The precision confidence curve shows the effect of the confidence level on the model precision, at each of the different pepper classes. The more confidence is gained, the accuracy is usually enhanced and most classes including green_bombay, red_bombay and red_capsicum have values near 1.0. Nevertheless, half_red_chili reveals a significant decrease, which indicates the challenge of differentiating transitional pigmentation. The general curve of all classes reaches maximum of 1.00 precision with a confidence threshold of 0.778, which means that YOLOv10 has an ideal precision when it is highly confident with its prediction. This indicates that the model is very dependable at the increased confidence levels but transitional types are difficult.

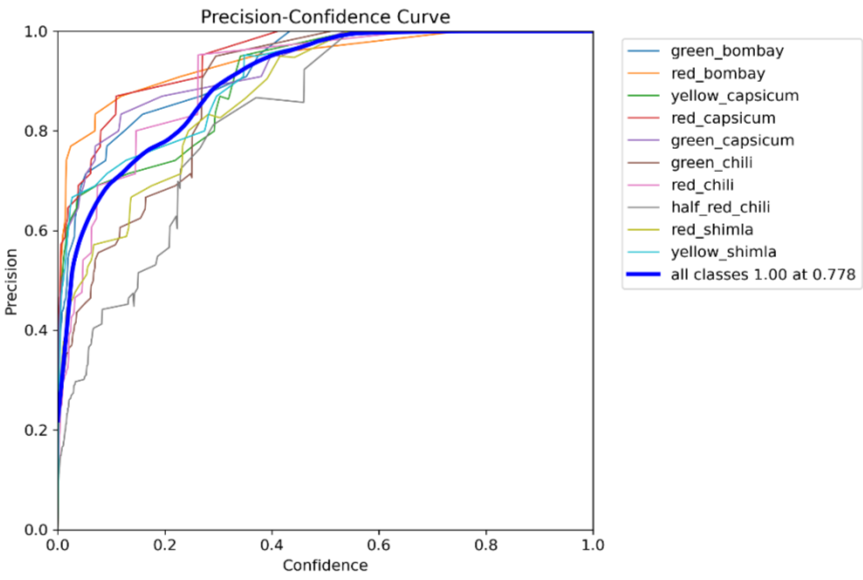
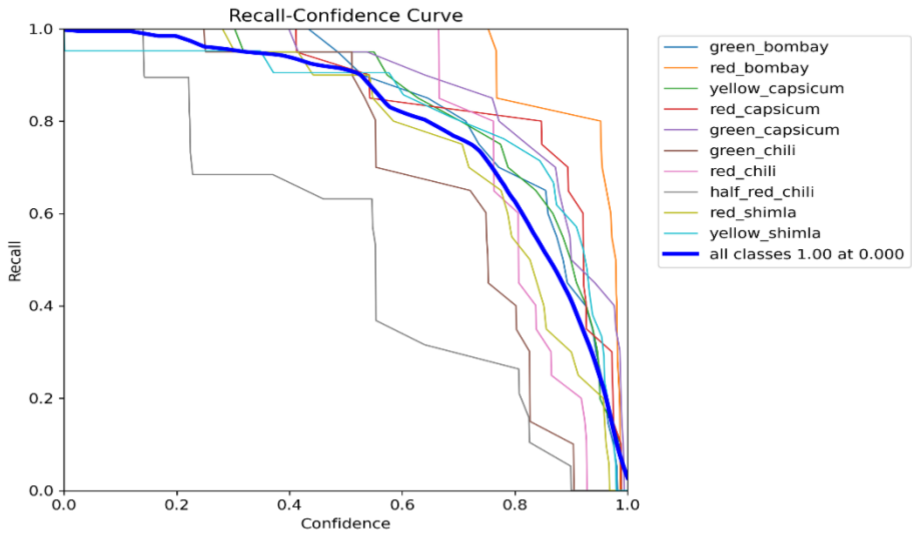


Figure 4.3.1Precision–Confidence Curve for YOLOv10

4.3.2 Recall–Confidence Curve

The recall confidence curve indicates that when confidence levels are high, recall declines. The recall is 1.00 on a confidence threshold of 0.0, i.e. all peppers are recalled irrespective of confidence. As confidence however, the tradeoff between complete detection and conservative prediction is reflected by a steadily decreasing recall. This observation demonstrates that YOLOv10 can recall almost all peppers with low confidence, but to implement it in practice, it is necessary to balance the recall and precision to prevent misclassification. In the case of agricultural automation, a mid-confidence level is important to achieve sufficient coverage as well as trustworthy classification.



4.3.2:Recall–Confidence Curve for YOLOv10

4.3.3 F_1–Confidence Curve

The F1-confidence curve is a balanced representation of the precision and recall that illustrates the trade-off between the two. The curve indicates that the model reaches the optimum output at moderate levels of confidence. In particular, the overall F1-score is the highest at 0.94 with a confidence threshold of 0.485. The same applies to individual classes, where it is clear that F1 increases rapidly at low confidence levels, peaks at middle levels of confidence and decreases as confidence levels increase.. This implies that YOLOv10 works optimally at moderate levels of confidence, and therefore, this compared range is the most suitable to be used in a real-time grading system where accuracy and coverage are essential factors.

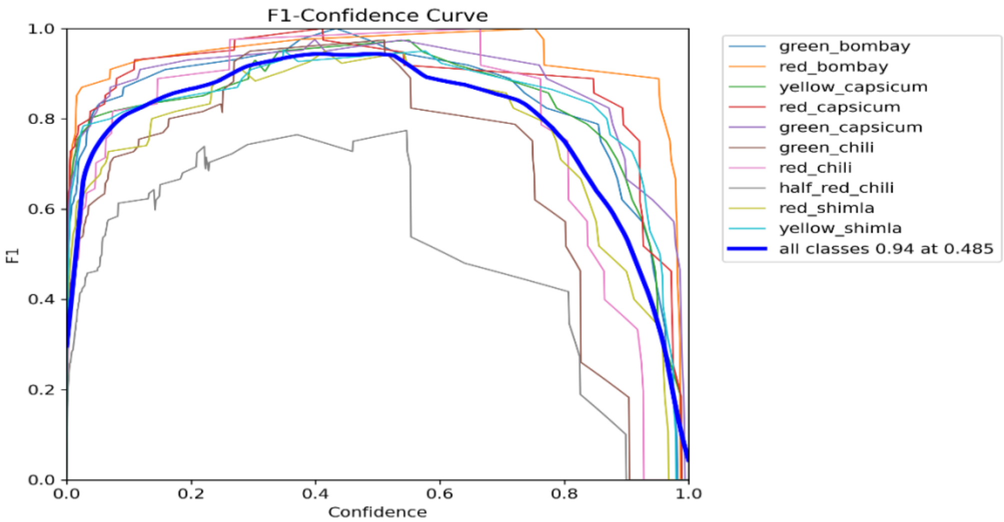


Figure 4.3.3:F_1–Confidence Curve for YOLOv10

4.3.4 Precision–Recall Curve

Precision-recall are directly compared in the precision–recall curve with all classes of pepper. The remainder of classes have very high values of precision at over 0.98 such as: green Bombay, red Bombay and red capsicum. By contrast, half-red chili has a smaller precision of 0.867, which proves its classification challenge based on the transition color properties. The general average precision (mAP 0.5) is 0.975, which is a strong score and indicates the stability of YOLOv10 in relation to different varieties of pepper. Such curves confirm the effectiveness of the model in the real-world application to agriculture, and only transitional varieties would need further optimization.

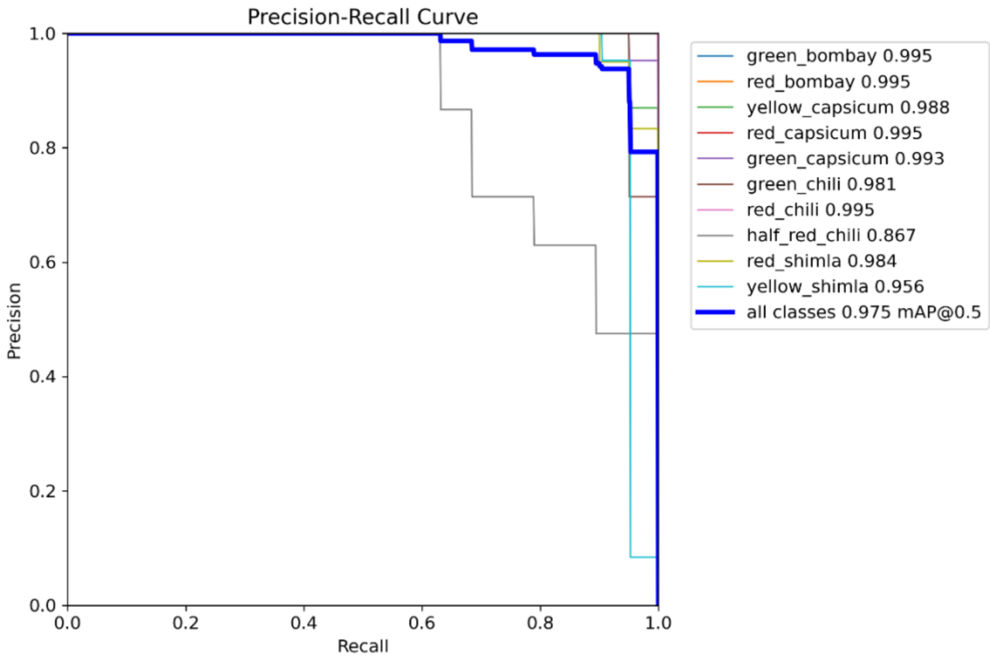
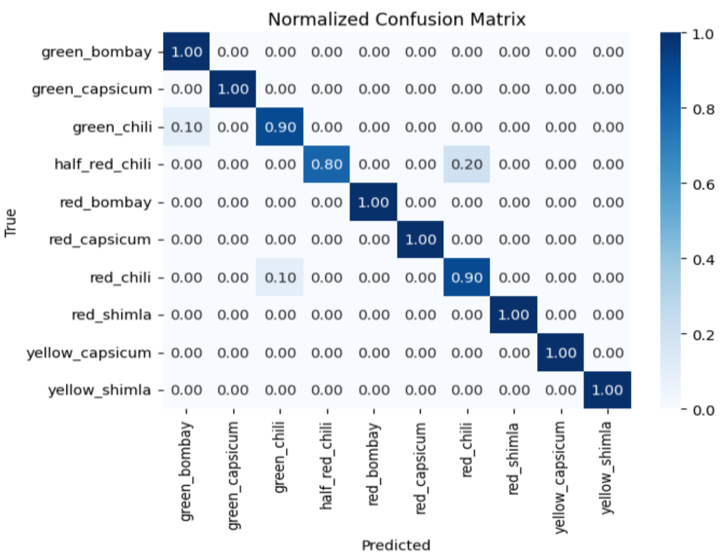


Figure 4.3.4:Precision–Recall Curve for YOLOv10

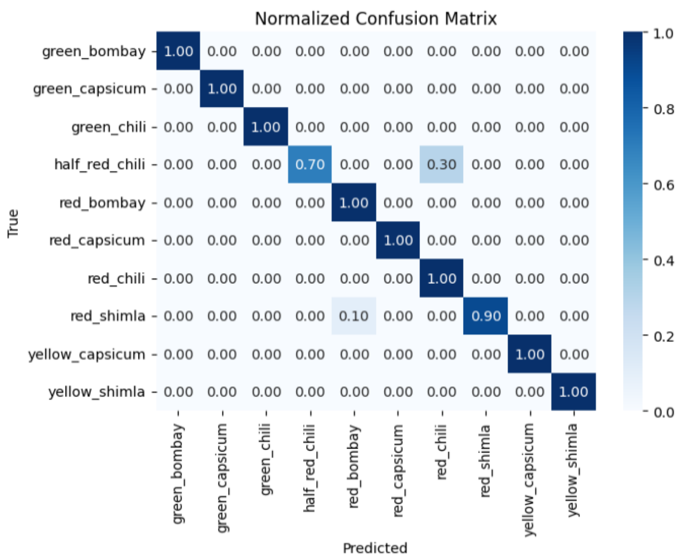
4.4 Confusion Matrices

In the experiments (Figure 3.2.1) we tested how to identify pepper varieties. In line with the holdout method, the dataset is further split into two subsets: 800 images to train, and 200 images to validate to prevent biased evaluation. In the case of the YOLOv10 model (Figure 4.4.1 (c)) the normalized confusion matrix indicates the strength of detection systems. Even half red chili, though, causes considerable confusion within the classification to background (32%), and the green chili too, has confusion with the background (21%).

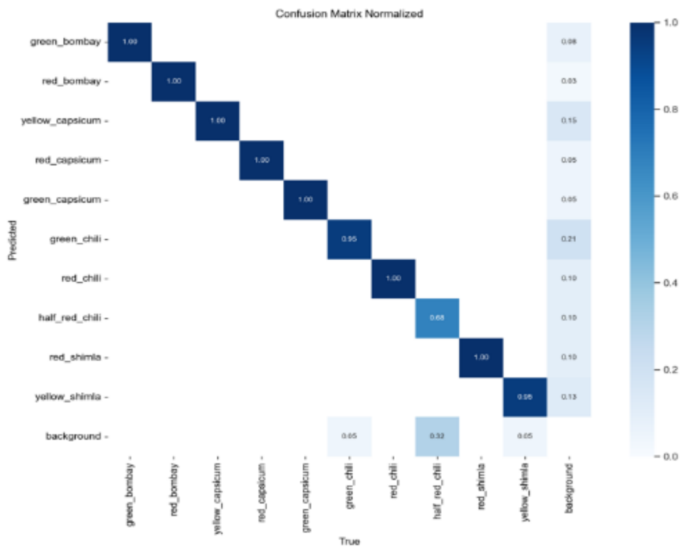
Overall, these errors indicate that the detection accuracy is impacted by background noise and visual similarity. Transitional phenotype half red chili has some characteristics of both green and red phenotypes; intermediate pigmentation. In contrast, pure classification-based classifiers such as VGG16 and ResNet50 (Figure 4.4.1 (a) and (b)) do not contain a background class and display more clear diagonal domination. We found VGG16 to work quite well with very few mislabeling on majority of the classes, most confusion between half red vs. green chili. Also, ResNet50 is off to a good performance with some confusion happening between half red chili and red Shimla. This kind of mistake can be explained by the similarities between varieties in their morphological characters such as overlapping color range and shape. On the whole VGG16 is slightly more accurate and YOLOv10 performs more consistently in difficult object detection environments.



(a) VGG16



(b) ResNet50



(c) YOLOv10

Figure 4.4.1:Normalized confusion matrices of (a) VGG16, (b) ResNet50, and (c) YOLOv10

First, the fact that the values of the confusion matrix of YOLOv10 are off diagonal indicates that object detection models are sensitive to the complexity of the environment. In contrast to a classification model which takes cropped or processed inputs, YOLOv10 is required to localize and classify objects in a scene all at once. Consequently, the background clutter, variation of illumination, and partial obscuration have a substantial impact on its predictions. The fact that there is a relatively higher confusion with the background class is an indication that sometimes the model does not differentiate fine object boundaries particularly when the pepper is a small part of the image or those that have a similar color hue of the environment.

Also, the mis-labeling of half red chili highlights the difficulty of addressing the issue of transitional or hybrid classes in real world data. The learned representations of features can overlap in the latent space since half red chili overlaps both the case of green and fully red versions. This implies that the discriminative features (patterns of texture, finer grained color histograms or even time progression) would be beneficial to the classification strength with their inclusion as well.

In cases where this variability is close to the differences among the classes, then the model might not be able to be generalized well. This is very low in instances where green chili is mixed with the ground, meaning that some of the samples might not have any strong distinguishing features or that the sample was taken in unfavorable conditions. Adding more representative and diverse samples to the data may address this problem.

The larger diagonal dominance in VGG16 and ResNet50 signifies that both models have learned more robust and discriminative feature representation in performing classification tasks. By mapping over already stored preprocessed images, since objects of interest are already centered and salient in the preprocessed image, these architectures are less vulnerable to background noise. The controlled input environment enables the models to concentrate on the learning of class-specific features, and therefore makes fewer misclassifications. It should be observed though that this benefit was acquired at the expense of real-life practicality since these models by themselves do not inherently execute object localization.

Comparison between VGG16 and ResNet50 can also give us insight into differences between architecture. In case such techniques were used in the training process, they could have helped in the enhanced performance of classification models. But in cases with detection model such as YOLOv10, augmentation has to be put into moderation to maintain spatial associations and bounding box precision.

Although classification models are very accurate in a controlled environment, they lack the capabilities to identify an object in a complex environment. It is in this trade-off that the significance of picking of models based on the desired application is emphasized and not necessarily the accuracy measures.

Besides, the confusion matrix may be utilized as a diagnostic measure, which can be further enhanced in the model. As an example, more half red chili samples might be collected under different conditions or new auxiliary labels capturing transitional points of ripeness might be used to minimize ambiguity. Although deep learning models are formidable, their ability is naturally connected with the quality and the representativeness of the data models on which they are trained. Future research can investigate the possibility of combining the merits of classification and detection models such as a detection model to recognize objects and then a specialized classifier to do a fine-grained classification. This can assist in surmounting the shortcomings noted in the present study and result in more concrete and credible pixel applications to identify pepper varieties.

Chapter – 5: Comparative Results Analysis

In order to give a clear picture of the practical implications of the performance of the deep learning models implemented in breeding pepper we must give their results against the larger picture of the so-called state-of-the-art research in the field of agricultural machine vision. Such comparisons help in putting the accuracy, robustness and efficiency into perspective the architectures used in this research are compared to the previous methods in terms of classification capability, and their adaptation to different environmental conditions and their computing capabilities.

With the VGG16, ResNet50, and YOLOv10 being fitted in this larger framework, it becomes not just possible to gauge their numerical peculiarities in regard to the accuracy, the precision, recall, and F1-score but to also assess their relevance to the real-life agricultural environment. It has been demonstrated in this analysis that classical classification architectures have both stability and high accuracy, or that detection-based techniques such as YOLOv10 can result in new measures of efficiency and practicality to real-time grading and sorting problems.

5.1 Result Comparison Against External Studies

Table 5.1 can be considered as a point of reference to demonstrate the position of our research concerning pepper breed recognition against other research which is specific in nature in agricultural and food vision. The maximum accuracy of our work was 96% using 1000 images, although the other articles (like Xuan et al.) reported high 95% F1-score when using 849 images to detect apples.

Furthermore, our own results are also quite competitive when compared with the classification of coffee beans study of Korkmaz et al., whose best model (InceptionV3) had an accuracy of 93% on a larger data set of 1,554 images. Such a difference is a sign that our present model, which was built on the basis of ResNet50, is very efficient when it comes to the specific task connected to naming various types of peppers.

Table 5.1.1:Comparative Overview of Transformer-Based Studies

Study	Languages & domain	Object(s) Deal With	Dataset	Model Evaluated	Best Performance
This Study	English, Agricultural Machine Vision	Pepper Breed Classification	1000	YOLOv10, ResNet50, VGG16 Faster	ResNet50: 96%Accuracy
Xuan et al. [50]	English, Agricultural Machine Vision	Apple Detection	849	R-CNN, YOLOv3, Improved YOLOv3	Improved YOLOv3:F1 up to 95%
Korkmaz et al. [51]	English, Food Vision	Coffee Bean Species Classification	1554	Xception, DenseNet201, InceptionV3, ResNetV2, DenseNet121	InceptionV3: 93% Accuracy

5.2 Benchmarking Against External Studies

Comparison of the findings of this research with the work of other individuals is crucial in order to confirm the soundness and originality of the suggested framework. Earlier studies on agricultural machine vision have shown good results, but most frequently, with a small dataset size, generalization in the model or practicality in deployment. For example,

Xuan et al. obtained a 95% F1-score in the task of detecting apples with 849 images and an enhanced YOLOv3 model, which demonstrates the promise of detection-based architectures but at the cost of fairly small datasets. Likewise, Korkmaz et al. studied the classification of coffee bean species with InceptionV3, reporting the results of 1,554 images with an accuracy of 93 percent, which highlights the feasibility of CNN-based transfer learning but in a food vision system, not in the agricultural crop classification.

Conversely, the current study presented a balanced dataset of 1,000 pepper images of ten varieties and balance the data in the study, removing any imbalance between classes. ResNet50 with this dataset accuracy 96 percent, VGG16 used the dataset 96%, and YOLOv10 used the data 94.25 % accuracy with a faster inference rate. Furthermore, YOLOv10 had an overall mAP 0.5 of 0.975, which is higher than most of the previous YOLO-based agricultural research that reported between 0.92-0.94 as their precision value. These performances prove that the proposed framework is in agreement with and extends beyond the state-of-the-art by merging high accuracy and the flexibility of the solution in real-time. Notably, the benchmarking points out that CNN models are effective under controlled conditions, but YOLOv10 offers a viable tradeoff between model quality and speed and can apply in dynamic agricultural situations like grading, sorting, and harvesting with robots. The comparative analysis, therefore, places this study as an important advancement to agricultural machine vision and closing the gap between the theoretical performance and the real implementation.

Chapter-6: Limitations and Future Works

This chapter is a summary of the main limitations of the study and an indication of the future research. Although promising results were obtained with the proposed methods, there are certain limitations that restrict their application to a broader range. Such limitations outline the ways in which this work can be improved, whereas the future work section shows ways in which this research can be expanded and become more solid.

6.1 Limitations

This study has a number of significant limitations in spite of promising results. The initial limitation is related to the size and diversity of the data. Despite the dataset being evenly distributed in ten pepper varieties, the size of the dataset in general is relatively small when compared to large benchmark datasets that are applied in computer vision research. Such small size can reduce the generalization ability of the models, especially when used in the wide variety of agricultural settings beyond the experimental dataset.

The other limitation is the model calculations of the models used. Though VGG16 is equally as well as ResNet50, both resource-intensive, and high-accuracy and both can be easily matched to real- because of their high memory and processing power requirements applications to life in low-resource or edge systems. YOLOv10 was more and faster not easily affected by noises in the surrounding, but in a controlled environment, it was quite accurate a little less, and signifying a trade-offs between speed and accuracy. This trade-off underlines the difficulty of achieving the optimal combination of the high degree of accuracy and prompt sense making in the farm vision systems. Moreover, the pre-trained models have a weakness with regards to their usage originally intended to do general image classification. Despite the improved transfer learning, these architectures are not designed to be efficient to take benefit of the special visual characteristics of the agricultural samples. Lastly the they only conducted experiments with peppers, which limited the use in terms of application. the more agricultural produce recognition jobs.

6.2 Future Works

We will also consider the future research to enhance the dataset size and variety by including more images taken in different environmental conditions. Multimodal aspects like spectral images and texture descriptors will be included to enhance the material so that the models can be able to detect subtle variations in pigmentation and surface architecture. Also, hyperparameter optimization, such as learning rate adjustments, will be considered to improve convergence and performance in general.

In this paper, three heavy pre-trained models (VGG16, ResNet50, and YOLOv10) were used; however, they have high computational requirements, making them impractical to use in more resource-limited settings. To solve this, the future work will involve research the design and implementation of the custom CNN architecture, specifically designed to recognize pepper breeds. These light but specialized models are likely to be more accurate and less complex, thus easier to implement in practice in agriculture.

Chapter-7: Conclusion

This paper is a comparison of three deep learning models to the task of pepper breed classification the models VGG16, ResNet50, and YOLOv10. As witnessed in the methodological differences between classification and detection they are based methods and the study has shown the strengths and weaknesses of the methods each models. Classical image recognition models (VGG16 and ResNet50) had a good performance in regard to accuracy, precision, recall and F1-score with VGG16. The consistency of being repeated in all measures and the most accurate in its ResNet50 reduces to being the most accurate capacity to create finesse and harmonious characteristics. YOLOv10, which was which was originally designed to be applied in a real time context to detecting objects, was converted to classification, but with only a little inferior accuracy, the model is, up to a great extent, great practical advantages regarding the inference speed, the capability to adjust to the environment modifications, and the possibility to use in real-time agriculture. The measurement metrics as accuracy, accuracy, recall, F1-score, false positive rate (FPR) and false negative rate (FNR) were a more detailed methodology of ascertaining the performance of a model.VGG16 model is an effective breed classifier that classifies pepper breed with better accuracy and generalization [52]. The findings showed that classification-based models are very reliable and robust in controlled conditions where detection-based models like YOLOv10 have special advantages in real-world situations, especially when it comes to operating in dynamic field conditions where speed and flexibility are most important. This comparative study highlights the trade-off between accuracy and efficiency, which should be put into serious consideration during the design of machine vision system to be used in agriculture. In addition to the technical results, the research shows that how deep learning can be utilized to recognize agricultural breed, especially with peppers. It demonstrates that pre-trained architectures can provide powerful performance, but they have computational and size requirements, as well as being unoptimized on agricultural datasets, which makes them difficult to use in the real world applications. The results of the research contribute not only to the confirmation of the effectiveness of deep learning in crop classification, but also to the identification of the necessity of further research in the expansion of datasets, integration of multimodal features, and the creation of lightweight custom CNNs adapted to agricultural tasks [53].

To summarize, the study has shown that deep learning models have a great potential to develop automated agricultural systems. Classification-based models are more accurate and robust whereas detection-based models are adaptable in real-time.

Bibliography

[1] M. I. M. N. N. W. M. M. S. K. S. M. a. R. M. M.N.Shah Zainudin, "Analysis of Detection, Recognition and Distance Estimation of Chili using YOLOv5," *IAENG International Journal of Computer Science ,* pp. 4245-4256, 2025.

[2] M. O. S. F. N. Khaled Mohi-Alden, "A machine vision-intelligent modelling based technique for in-line bell pepper sorting," *Information Processing in Agriculture,* vol. 10, no. 4, pp. 491-503, 2023.

[3] P. G. B.-P. ,. P. M. M. M. B. E. Laura Gomez-Zamanillo, "Deep learning-based instance segmentation for improved pepper phenotyping," *Smart Agricultural Technology,* vol. 9, p. 100555, 2024.

[4] T. N. P. T. T. T. L.-D. M. S. C. Trong Tai Ha, "Applying machine learning for chili pepper phenotyping and feature extraction," *Smart Agricultural Technology,* vol. 12, p. 101458, 2025.

[5] N. K. ,. K. R. Alexey Kutyrev, "Recognition and Classification Apple Fruits Based on a Convolutional Neural Network Model," *Information Technology and Implementation ,* pp. 90-101, 2022.

[6] M. Z. S. M.A.Farhan Ibrahim, "Detection and Classifying of Chili Fruits Variations using YOLOv5," *KZYJC,* vol. 38, no. 04, pp. 1363-1376, 2023.

[7] B. P. Sapan Naik, "Machine Vision based Fruit Classification and Grading - A Review," *International Journal of Computer Applications ,* vol. 170, no. 9, pp. 22-34, 2017.

[8] F. P. L. A. d. S. S. d. S. C. d. A. N. ,. T. O. Madalena de Oliveira Barbosa, "YOLOv8m for Automated Pepper Variety Identification: Improving Accuracy with Data Augmentation," *Applied Sciences ,* vol. 15, no. 13, p. 7024, 2025.

[9] G. A. C. A. M. J. R. Christian Joel Lazo, "Automated Bell Pepper Quality Assessment: Robotic Gripper Sorting System with Transfer Learning," *วารสารงานวิจัยและพัฒนาเชิงประยุกต์โดยสมาคม ECTI,* vol. 5, no. 1, pp. 1-13, 2025.

[10] S. Z. Yane Ma, "An Enhanced Computer Vision Model for Detecting the Maturity of Chili Pepper in the Natural Environment.," *agronomy,* vol. 15, no. 3, p. 37, 2025.

[11] M. A. R. M. S. U. M. H. M. T. H. Md. Ataur Rahman, " A Comprehensive Analysis of Classic Machine Learning and Deep Learning Modeling for Breed Recognition of Bananas," *IEEE Access,* vol. 13, pp. 194562-194579, 2025.

[12] B. C. S. X. Z. C. Zhiyuan Tan, "Pepper Target Recognition and Detection Based on Improved YOLO v4," *Information Technology and Control,* vol. 52, no. 4, pp. 878-886, 2023.

[13] W. S. L. W. M. H. D. S. Z. H. M. H. X. Z. W. Min Dai, " Pepper leaf disease recognition based on enhanced lightweight convolutional neural networks," *Frontiers in Plant Science,* vol. 14, p. 1230886, 2023.

[14] J. A. E. C. ,. G.-E.-E. M.-C. Jesús Dassaef López-Barrios, " Green sweet pepper fruit and peduncle detection using mask R-CNN in greenhouses," *Applied Sciences,* vol. 13, no. 10, p. 6296, 2023.

[15] A. M. J. M. A. S. U. A. Md. Tarek Habib, "Machine vision-based papaya disease recognition," *Journal of King Saud University-Computer and Information Sciences,* vol. 32, no. 3, pp. 300-309, 2020.

[16] F. P. L. A. ,. S. d. S. S. L. d. S. C. d. A. N. M. T. O. Madalena de Oliveira Barbosa, "YOLOv8mfor Automated Pepper Variety Identification: Improving Accuracy with Data Augmentation," *applied Sciences,* vol. 15, p. 7024, 2025.

[17] A. B. B. ,. L. M. N.-G. Oscar Leonardo García-Navarrete, "Automated Weed Detection in Red Beet (Beta vulgaris L.,Conditiva Group, cv. Kestrel F1) Using Deep Learning Models," *agronomy,* vol. 16, p. 167, 2026.

[18] H. L. ,. Z. ,. W. ,. W. H. Z. L. Qiong Wu, "YOLO_SSP: An Auto-Algorithm to Detect Mature Soybean Stem Nodes Based on Keypoint Detection," *MDPI,* vol. 15, no. 5, p. 1128, 2025.

[19] J. L. L. F. L. J. N. Y. Y. L. Shilong Zhao, "A counting method of whiteflies on crop leave images captured by AR glasses based on segmentation and improved YOLOv11 models," *frontiers,* vol. 16, 2025.

[20] X. K. ,. M. Weixun Chen, "Small defect detection in printed circuit boards based on the multiscale edge strengthening and an improved YOLOv10," *Scientific Reports,* vol. 15, no. 36445, 2025.

[21] S. L. L. L. X. S. H. Z. Qiufang Dai, "YOLOv11-RDTNet: A Lightweight Model for Citrus Pest and Disease Identification Based on an Improved YOLOv11n," *agronomy,* vol. 15, no. 5, p. 1252, 2025.

[22] H. S. w. Z. Z. C. L. j. Fupeng Li, "Lightweight optimization of YOLOv8m for robotic vision-based snack cucumber sorting and palletizing," *Journal of Agriculture and Food Research,* vol. 23, p. 102248, 2025.

[23] A. M. ,. A. K. T. Himanshu Kumar Diwedi, "CNN-based medicinal plant identification and classification using optimized SVM," *Springer Nature ,* vol. 83, p. 33823–33853, 2023.

[24] Z. A. S. M. G. B. P. V. S. Danish Syazani Mohd Zakir, "Comparative Analysis of VGG-16, ResNet50andEfficientNet-B1 with Optimization Techniques for Crop Disease Detection," *Journal of Advanced Research Design,* vol. 143, no. 1, pp. 55-64, 2026.

[25] S. R. a. Q. S. a. B. H. a. S. S. M. W. a. S. M. I. a. M. F. Shah, "Comparing inception V3, VGG 16, VGG 19, CNN, and ResNet 50: a case study on early detection of a rice disease," *Agronomy,* vol. 13, no. 6, p. 1633, 2023.

[26] A. P., S. N. and V. N., "Detection and classification of breast cancer types using VGG16 and ResNet50 deep learning techniques.," *International Journal of Electrical and Computer Engineering,* vol. 14, no. 5, pp. 5481-5488., 2024.

[27] Z.-P. a. L. Y.-Y. a. S. Z.-E. a. H. K.-W. Jiang, "An Improved VGG16 Model for Pneumonia Image Classification," *Applied Sciences,* vol. 11, p. 23, 2021.

[28] T. K. T. A. A. K. M. S. S. S. W. M. Azaharan, "Investigation of VGG-16, ResNet-50 and AlexNet Performance for Brain Tumor Detection," *International Journal of Online & Biomedical Engineering,* vol. 19, no. 8, 2023.

[29] Y. S. L. L. L. N. C. M. Y. Y. M. H. G. Jinfan Wei, "Tomato ripeness detection and fruit segmentation based on instance segmentation," *Frontiers in Plant Science,* vol. 16 , 2025 .

[30] R. R. S. Dhananjay Theckedath, "Detecting Affect States Using VGG16, ResNet50 and SE-ResNet50 Networks," *Springer Nature,* vol. 1, 2020.

[31] T. M. T. M. R. M. ,. S. T. Amit Kumar Trivedi, "Leveraging feature fusion ensemble of VGG16 and ResNet-50 for automated potato leaf abnormality detection in precision agriculture," *Springer Nature,* vol. 29, p. 2263–2277, 2025.

[32] G. F. Seyedamirhossein Mousavi, "A Novel Enhanced VGG16 Model to Tackle Grapevine Leaves Diseases With Automatic Method," *IEEE Access,* vol. 10, pp. 111564 - 111578, 2022.

[33] S. Y. ,. Y. ,. T. Z. L. J. Z. L. T. G. L. Z. G. Weiming Li, "Machine Learning Model of ResNet50-Ensemble Voting for Malignant–Benign Small Pulmonary Nodule Classification on Computed Tomography Images," *Cansers,* vol. 15, no. 22, p. 5417, 2023.

[34] J. S. Shashi Tanwar, "ResNext50 based convolution neural network-long short term memory model for plant disease classification," *Springer Nature,* vol. 82, p. 29527–29545, 2023.

[35] I. Q. ,. A. ,. J. A. A. ,. Z. S. Kashif Shaheed, "EfficientRMT-Net—An Efficient ResNet-50 and Vision Transformers Approach for Classifying Potato Plant Leaf Diseases," *sensors,* vol. 23, no. 23, p. 9516, 2023.

[36] S. S. G. R. E. ,. T. S. S. Srijah, "Tomato Leaf Disease Classification Using an Improved ResNet50 Architecture," *Springer Nature,* pp. 529-539, 2024.

[37] M. B. S. A. B. M. S. G. A. K. S. S. M. S. Nilesh B. Korade, "Proactive Soybean Disease Detection through YOLO Leaf Extraction and ResNet-50 Classification to Reduce Crop

Loss and Boost Productivity," *International Journal of Engineering Trends and Technology,* vol. 73, no. 1, pp. 385-396, 2025.

[38] M. S. Z. a. A. A. N. A. a. L. H. S. a. G. A. K. a. A. M. a. A. F. a. L. A. A. a. A. A. K. Ahmad, "A Trustable Spine Abnormalities Classification System Using ResNet50 and VGG16 Supported by Explainable Artificial Intelligence," *Biomimetics,* vol. 11, no. 3, 2026.

[39] I. S. G. E. H. Alex Krizhevsky, "Imagenet classification with deep convolutional neural networks," *Advances in Neural Information Processing Systems ,* vol. 25, 2012.

[40] A. Z. Karen Simonyan, " Very deep convolutional networks for large-scale image recognition," *arxiv ,* pp. 1-14, 2015.

[41] K. S. S. R. B Leelavathy, "Plant Disease Classification using VGG-16: Performance Evaluation and Comparative Analysis," *International Journal of Engineering Science and Advanced Technology,* vol. 23, no. 08, pp. 219-229, 2023.

[42] C. a. Z. D. a. S. Y. a. Z. X. a. L. S. a. C. L. Zhang, "Research on the authenticity of mutton based on machine vision technology," *Foods,* vol. 11, no. 22, p. 3732, 2022.

[43] S. S. ,. M. S. R. ,. S. N. ,. M. T. H. Iffat Firozy, "Cnn Modeling for Bengali Sweet Recognition," *SSRN,* p. 14, 2023.

[44] S. Z. ,. S. G. Rui Ren, "Research on Pepper External Quality Detection Based on Transfer Learning Integrated with Convolutional Neural Network," *sensors,* vol. 21, no. 16, p. 5305, 2021.

[45] M. S. Nishant Behar, "ResNet50-Based Effective Model for Breast Cancer Classification Using Histopathology Images," *Computer Modeling in Engineering \& Sciences (CMES),* vol. 130, no. 2, pp. 823-839, 2022.

[46] H. C. L. L. K. C. Z. L. J. H. G. D. Ao Wang, " Yolov10: Real-time end-to-end object detection," *Advances in Neural Information Processing Systems,* vol. 37, pp. 107984--108011, 2024.

[47] M. H. Mujadded Al Rabbani Alif, "YOLOv1 to YOLOv10: A comprehensive review of YOLO variants and their application in the agricultural domain," *arXiv preprint arXiv:2406.10139,* 2024.

[48] W. J. G. G. L. Y. Z. C. W. Y. B. H. L. Y. G. X. ,. L. Q. Li Z, "SGSNet: a lightweight deep learning model for strawberry growth stage detection.," *Front. Plant Sci,* vol. 15, p. 1491706, 2024.

[49] J. J. D. L.-B. A. E. C. ,. G.-E.-E. M.-C. Jesús Dassaef López-BarriosORCID, "Green sweet pepper fruit and peduncle detection using mask R-CNN in greenhouses," *Applied Sciences,* vol. 13, no. 10, p. 6296, 2023.

[50] C. G. Y. S. M. Z. Y. W. J. Z. Guantao Xuan, "Apple detection in natural environment using deep learning algorithms.," *IEEE Access,* vol. 8, pp. 216772-216780, 2020.

[51] T. T. S. K. T. I. Adem Korkmaz, " Comparison of deep learning models in automatic classification of coffee bean species," *PeerJ Computer Science,* vol. 11, no. e2759, 2025.

[52] S. B. H. E. U. R. S. H. Ritu Rani, "VGG-EffAttnNet: Hybrid Deep Learning Model forAutomated Chili Plant Disease Classification Using VGG16and EfficientNetB0 With Attention Mechanism," *Food Science & Nutrition,* vol. 13, no. 7, 2025.

[53] W. Z. Y. W. W. Q. Y. ,. Y. L. Xiaoyao Yang, "Lightweight and efficient deep learning models for fruit detection in orchards," *scientific reports,* vol. 14, p. 26086 , 2024.