# A Review of Machine Learning Techniques to Enhance Vegetable Grading and Spoilage Detection

Zainab Usman
Dept of CSE, PES University
100 Feet Ring Road
Bangalore, India

Vaishnavi C. Kamatagi
Dept of CSE, PES University
100 Feet Ring Road
Bangalore, India

Tilak Matagunde
Dept of CSE, PES University
100 Feet Ring Road
Bangalore, India

Raviprakash
Dept of CSE, PES University
100 Feet Ring Road
Bangalore, India

Revathi G.P.
Dept of CSE, PES University
100 Feet Ring Road
Bangalore, India

## ABSTRACT

Examining the nexus of machine learning and farming, this paper navigates through the landscape of AI-driven fruit and vegetable grading, revealing strides in accuracy alongside ongoing computational and quality related hurdles. YOLOv5, Faster R-CNN, and Vision Transformers (ViT) are some of the techniques reviewed for object detection and classification tasks that have very high accuracy with real-time performance. This survey presents a novel approach for fruit and vegetable classification through modified K-Means, SVM, and ANN techniques with remarkable efficiency and cost reducing abilities. The discussion incorporates traditional edge detection methods like Sobel, Prewitt, and Canny, along with innovative approaches like GrabCut and DeepSORT, for contributions in the field of image segmentation and object counting. Despite improvements with high impact, overfitting, dependence on the quality of images, and computational costs remain, which require continued research. This paper surveys the current methodologies in agricultural Machine Learning and Deep Learning applications and presents some directions of future improvements.

## General Terms

Machine learning, Deep learning, Agricultural technology.

## Keywords

Object detection, Fruit classification, Image segmentation, YOLOv5, Faster R-CNN, Vision Transformers,

## 1. INTRODUCTION

The infusion of agricultural applications with Machine Learning and Deep Learning techniques has revolutionized processes for object detection, edge detection, segmentation, and classification of fruits and vegetables. In this regard, this literature survey explores different state of the art methodologies and their efficacy toward bettering accuracy, efficiency, and robustness in these processes. Focus is mainly on how these technologies have been applied in disease identification, edge detection, classification, and grading of agricultural products with special emphasis on fruits and vegetables. Traditional agriculture methods, like manual grading and disease detection, make the process very time consuming, labor intensive, and subjective. Advances in ML and DL present promising alternative solutions for the same. These methods consist of YOLOv5, Faster R-CNN, Vision Transformers, also known as ViTs, and others, which are very applicable with amazingly good performance in object detection and classification tasks. For instance, faster R-CNN and YOLOv4 are applied in apple grading and have high accuracy with a real-time processing capability.
State of the art methodologies, like GrabCut and Deep-SORT, are discussed in this paper, with major applications in object detection and counting, segmentation of images in various agricultural settings. Yet, even at present, when development is on its way, the persistence of some challenges like overfitting issues, dependence on image quality, and high computational costs does not allow their practical use in real life. Since real-life applicability remains one of the targets, further research and improvement is required continuously. The emphasis on reliable processing data, enhancement techniques, and adaptability to different environmental conditions is needed.This literature survey consolidates the current state of research on the application of ML and DL in the agri-food sector with a focus on classification and grading, particularly of fruits and vegetables. It does so in a manner that displays both potential and continuous challenge, sweeping up the complete landscape of progress and future directions in this field.

## 2. TECHNOLOGIES FOR DETECTING AND RECOGNIZING OBJECTS.

Investigates the utilization of YOLOv5 and Faster R-CNN for autonomous navigation in space debris management and satellite servicing. Motivated by the critical need for precise and fast object detection in space environments, the study compares YOLOv5's real-time capabilities with Faster R-CNN's higher accuracy. Methodologically, the researchers annotated and split 1231 RGB images into training (71%), validation (25%), and testing (4%) sets, augmenting their dataset with videos from the ORION research laboratory. YOLOv5, known for its efficiency in real time applications, was trained with updated pre-trained weights every 100 epochs 4 times, while Faster R-CNN used a pre-trained ResNet101 model in PyTorch with Detectron2. Performance metrics, including mean Average Precision (mAP), highlighted YOLOv5's superior speed and practicality for collision avoidance, achieving an mAP 0.5 of 0.85 compared to Faster R-CNN's 0.91. Although Faster R-CNN exhibited higher accuracy, it was significantly slower, processing at one-tenth the speed of YOLOv5. The study recommends a hybrid approach where Faster R-CNN intermittently validates YOLOv5's detections to balance speed and accuracy. Future directions include integrating these algorithms on spacecraft with limited computational resources and expanding the dataset to encompass more diverse space scenarios. Limitations include dataset constraints and computational challenges, which require further exploration for robust real-world implementation.[1] Focusing on enhancing apple detection in complex orchard environments, presents that the Faster-RFormer model, which features a Swin Transformer backbone, can be used for this without any re-tooling. The study that was able to identify the detection as the apple fruit mainly considers the failures of the traditional detection methods in orchards with complex backgrounds and varied lighting as its focal issue and then brings out a plan to improve the situation via the use of Faster-RFormer. To collect the data, 5081 images, including 1120 from the ACFR Orchard Fruit Dataset, and 2400 from Panjin and 1561 images crawled from the web. The dataset was annotated for the apple detection by the researchers. In the Faster-RFormer model, the Swin Transformer is combined with a Region Proposal Network (RPN) to detect visual features with a high level of precision focusing on the exact object location. Moreover, the concerned problems such as lighting and occlusion were tackled by exercising the data augmentation solution that is responsive to the unique types of challenges in the scene. Models using Faster R-CNN without feature fusion perform worse than even the weakest SSD model. The SSD model has a mean Average Precision (mAP) of 0.552, AP@0.5 of 0.869, and AP@0.75 of 0.619. This highlights the importance of feature fusion. Faster R-CNN with feature fusion: mAP of 0.614, AP@0.5 of 0.917, AP@0.75 of 0.704. YOLOv3: mAP of 0.565, AP@0.5 of 0.896, AP@0.75 of 0.642. YOLOv5: mAP of 0.69, AP@0.5 of 0.933, AP@0.75 of 0.811. YOLOv8: mAP of 0.703, AP@0.5 of 0.938, AP@0.75 of 0.807. The proposed AppleForformer model achieves an mAP of 0.692, with AP@0.5 of 0.941 and AP@0.75 of 0.797, surpassing most of these models in key metrics. Despite its strengths, Potential performance degradation in new orchard environments not represented in training data. Future research aims to enhance adaptability and validate performance across broader datasets to consolidate Faster-RFormer's applicability in diverse orchard settings.[2]

It is a systematic review of YOLO in agricultural object detection. YOLO is great for real- time object recognition and can change the way agriculture research is synthesized.[3] uses insights from 257 papers to show trends and advancements in YOLO. Methodologically, a strict document selection process was used, focusing on 30 papers demonstrating YOLO in agricultural settings. YOLO versions (v1 to YOLOX) were categorized based on their features, advantages, disadvantages, and suitability for different agricultural tasks. Results showed YOLO's impact on precision agriculture, YOLOv4 achieved mAP of 0.85 and YOLOX performed competitively with different backbones. Challenges are detecting objects against similar backgrounds and continuous algorithm refinement and dataset expansion.It concludes more research is needed to tap YOLO's potential in optimizing agricultural processes through object detection tailored to specific farming needs.[3] Data Acquisition includes using RGB cameras (high speed, action, industrial, surveillance), handheld, robot mounted, drone mounted, or fixed. Real world and simulated environments dictate the use of devices and methods applied in taking photos to obtain the images of the best quality. Synthetic data generation uses image processing or computer graphics software; this method is sometimes necessary when data is limited. Resolution 240×320 cameras is used to detect the stunned broiler chicken on the slaughter lines. High resolution for aerial imagery involves monitoring vast areas in the detection of date palm trees, maize tassels, and wheat spikes. Studies mostly use more than 100 images. Examples like Date palm tree data: 125 images with 8,797 objects. Fish uneaten pellet data: 175 images containing 7,684 labels that attained a detection accuracy of over 90% . With regard to weeding the YOLO model was able to detect more than 95% of 12 weed class issues in a cotton field.

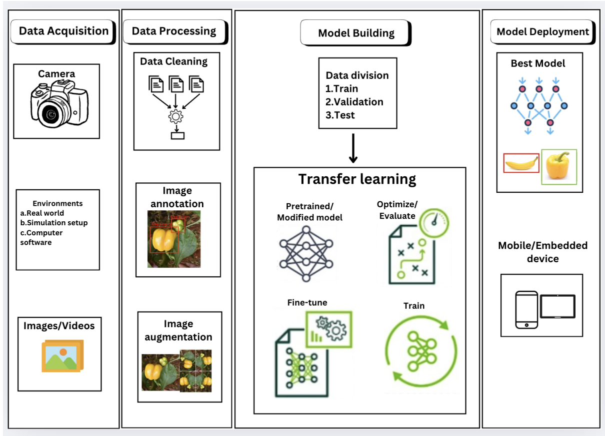


Fig. 1. *Comprehensive overview of an end to end YOLO implementation pipeline*

Data Preprocessing involves removal of low quality and redundant images.The quality of data is affected by machine vibration and environmental conditions like lighting and weather. Experts manually label images with bounding boxes and class labels using various online and offline tools. A set of different transformations to increase the size and diversity of the data. Geometric Transforms like Flip, crop, rotate, translation, noise, pan, blur, scale. Color Space Transforms includes Contrast, brightness, sharpening, histogram equalization, hue, saturation, shadow, sun flare. Image Mixing Transforms are Mixup, cutout, mosaic. Examples include Broiler chicken detection: Improved detection accuracy, Fish behavior detection, Improved robustness of the model, Pest detection in traps with better detection accuracy,

Multi-class cotton weed detection includes a large improvement in performance .
YOLO Model Input ingests images in square format. Resolution affects model detection performance. For Example, Coffee fruit detection improved by increasing the resolution from 416×416 to 896×896 pixels, Small, dense objects Apple inflorescence detection using 1400×1088 resolution had accuracy of 83.4 . The processed images are classified into training, validation and testing subsets with over 70% of the data used to develop the model. The testing subset would be used for unbiased evaluation of the model on unseen data, preferred in case of large datasets. Model Training and Transfer Learning uses Huge, High quality, and Diverse Datasets are Required but Can be Impractical to Collect. Using pre-trained models, for example, COCO or ImageNet, finetuned on specific tasks with smaller datasets. Most YOLO models were modified to aspire to greater accuracy, speed, size, or robustness,schedulers in network architecture, loss functions, or training strategies. Attention mechanisms focus on important regions during training. Channel Attention is concerned with key channels, such as YOLOR for rice row crops. Spatial Attention component focuses on key spatial regions, for example, YOLOweeds for seedling maize weeds. Combined Channel and Spatial Attention approach improves feature extraction; for example, in tomato growth stage detection. Advanced Feature Extraction Backbone Transformers Vision Transformers, ViT, rely on self-attention mechanisms that replace traditional convolutional layers. Examples include the use of improved YOLOv4 with Swin transformer to detect maize tassels by Zhang et al. in 2023.Counting the YOLO output directly from images, for example, maize tassels.Mapping detected objects to understand tree density for crop load management. Combining tracking algorithms avoids multiple counts in moving frames, for example, SORT and DeepSORT. YOLO combined with clustering algorithms for row crop detection. Precision Harvesting, Integrated with Key Point Detection Models: Location of fruit stems for machine picking. Combined with Large Language Models, YOLO predictions are analyzed by GPT-4 to diagnose diseases and pests are a few use cases mentioned.Strengths include Real time Detection: Critical in agriculture, it gives reasonable accuracy and speed. Cost-Effective Deployment: It works well on low-cost, resource constrained devices. Robustness against Environmental Variability: Training on diverse datasets for better robustness.
NTrack is a tool designed to automate the laborious task of counting cotton balls in fields and improve yield estimation. It combines dense optical flow and particle filtering with object detection to address the problems of manual counting errors and cotton ball detection in field conditions, including clustering and foliage occlusion. NTrack integrates object detection with tracking, using optical flow for movement prediction and particle filters for identity management across frames.The paper uses a powerful model called Cascade R-CNN with a ResNet-50 backbone to accurately detect objects. They trained this model extensively on a high-performance computer setup and achieved a detection accuracy of 97% on their test data, outperforming other methods like Tracktor. To track object movement, they calculated optical flow using a standard algorithm and determined object velocities by averaging pixel speeds within detected areas. They also used sophisticated techniques to assign detected objects to predicted ones. For estimating object positions relative to each other, they employed a method based on nearby object neighbors, choosing three neighbors for optimal accuracy in both width and height directions. The dataset and metrics outperform DeepSORT and Tracker, in MOTA and IDF1. Visual descriptors are relied upon for appearance based re-identification, but the dataset lacks

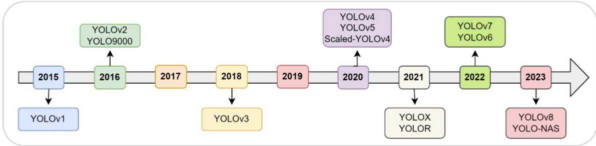


Fig. 2. *Development timeline of YOLO (You Only Look Once) object detection algorithms, showcasing key milestones.*

generalizability to different field conditions and crop stages. Future work will focus on making NTrack more algorithmically flexible and applying it to different agricultural contexts for precision agriculture.[4]
AI based system for tomato disease detection and remedy suggestions through a mobile app.[5] focuses on tomato crops and uses YOLOv5 for its balance between speed and accuracy for real time disease detection. It curates a dataset of 10 tomato diseases, they apply extensive data augmentation techniques like mosaic augmentation, scaling and rotation to make the model more robust.To empower traditional farmers with technology for crop health management, YOLOv5 is integrated into a user friendly Android app that communicates with a server to provide real-time disease detection and localized remedy suggestions. False positives and low accuracy for diseases with limited training samples are addressed through ongoing dataset augmentation and model refinement. YOLOv5 can detect small and dispersed disease areas, the model gets 0.76 mean Average Precision (mAP). The lack of explainability of the dataset for additional vegetable diseases, refine the algorithm based on user feedback.[5]

## 3. ANALYSIS OF EDGE DETECTION TECHNIQUES

Edge detection is an essential tool in the area of digital image processing that forms features for fast and accurate processes such as contemporary object recognition, image segmentation or feature extraction. Edge detection algorithms can be divided into several categories, and each has specialized strengths and corresponding weaknesses. Edge detection has great importance and is considered a very prior task in digital image processing fields as well as wherever the application needs to deal with it such as object recognition, automatic target tracking missile or an image segmentation etc. Edge detection study based on digital image processing class of traditional types of sensing operator including Roberts, Sobel, Prewitt and Laplacian operators as well intensity based methods like Canny also multiscale zooming out type algorithms. It explains the trade off between accuracy and noise immunity, and pays special attention to Gaussian function equations for edge detection, localizations, which helps in smoothing.[6] The Roberts edge operator uses local pixel differences to calculate edges, making it suitable for images with small signal to noise ratios [6]. On the other hand, the Sobel and Prewitt operators require image filtering before edge computation, and differ mainly in the filter weights used [6]. Laplacian operators based on second order derivatives are sensitive to image variation and signal to noise ratio[6] . Among these, the Canny edge operator stands out for its effectiveness, which uses Gaussian smoothing and extreme value computation to enhance edge detection while combating noise [6]. An enhanced Canny Edge Detection algorithm, as described in [9] includes binomial filtering, interpolation for non maximum suppression, and Otsu threshold splitting . This method establishes binomial filtering


Fig. 3. *Detection and segmentation of sweet peppers with colored annotations, highlighting the boundaries of each sweet pepper using distinct colors for better differentiation and analysis. .*

Gaussian filtering is replaced by binomial filtering, preserving edge information and noise reduction [9]. The addition of gradient templates at 45° and 135° reduces false detection loss, and interpolation for sub-pixel gradient calculation increases accuracy [9] Despite these improvements, computational difficulties go high and the speed of the algorithm poses considerable limitations [9]. The paper [8] attempts to highlight how combining the rough set with convolutional neural networks can help in overcoming these challenges about apple image's edge detection [8]. The research methodology involves using Faster- RCNN for object detection, converting RGB images to LAB color space for feature extraction and utilizing morphological processing techniques for refining edges [8]. Noise reduction is one of the salient aspects of this model that also helps in capturing important contours while at the same time being able to handle complex backgrounds as well as occlusions [8]. Nonetheless, it does not deal well with non uniformity edges and depth information remains unsolved yet [8]. Empirical study which shows the use of artificial neural networks and decision trees to rate jujube fruits [10]. The experiment uses feature selection techniques such as PCA and CFS, which gives rise to a high classification accuracy of 98.8% [10]. Moreover, decision tree classifiers, ANN, among other options have been included to demonstrate how machine learning can contribute towards better grading of fruits . However, improving the method's accuracy further will require taking into account features like shape and defects more exhaustively and addressing the limitations with statistical test validations [10]. Enhancements in traditional algorithms that improve detection as well as classification accuracy through MATLAB simulations [7]. Moreover, the algorithm still has some issues associated with threshold data settings and edge detection in noisy and dark backgrounds [7]. On the other hand, extending this technique to different types of fruits and vegetables can enhance its applicability more efficiently [7].

## 4. SEGMENTING AREA OF INTEREST.

The literature review contains exemplary classifications of sweet pepper and crop disease detection, respectively, which show important advances in deep learning approaches for agricultural innovation.
Targeted at sweet pepper with Detection and classification accuracy has increased in complex agricultural environments. Since the traditional manual methods are not effective in managing sweet pepper production, the paper proposes a deep learning based model that allows monitoring and analysis over time in itself is suggested, which showed a significant improvement in welfare and yield for crops. Researchers used a dataset of 1127 images of sweet peppers, scaled to 1024 × 1024 pixels, to train and test the Mask R-CNN model in the detection of sweet pepper fruits and peduncles. The authors divided the dataset into 80% for training and 20% for testing. Therefore, images are chosen at random to include most of the natural conditions likely to affect the performance of the system in different daylight, shadows, overlapping fruits, and a variety of states of ripeness and color. Most of the green sweet peppers were considered since they are a bit difficult to detect due to their color, which nearly matches the environment [11]. This is on the development of an automatic sorting system for bell peppers with machine vision techniques. The main objective is their classification with respect to quality characteristics like color and size. [11] Uses highresolution images of the bell peppers to be captured from a controlled lighting environment. Pre-processing techniques to be done on these images include color normalization and noise reduction to ensure uniformity and clarity in extracting features. Feature extraction plays an important role wherein color values in RGB or HSV are extracted along with geometric properties like contours to different shape and size bell peppers. A Multilayer Perceptron Neural Network shall be used subsequently for the classification. MLPs are chosen for handling nonlinear relationships existent in the data, and due to its effectiveness in tasks of image classification. An MLP is trained using labeled data with a high accuracy rate (95%) in classifying the bell pepper into different grades of quality. It proceeds with automatic fruit identification, including images from 131 different fruit classes based on the Fruit- 360 dataset. This is preceded by extensive preprocessing to increase variety and quality in training datasets through image augmentation and normalization. Extracting features makes use of a modified AlexNet architecture which was pre-trained originally on ImageNet and fine-tuned precisely for fruit classification tasks.
Classification involves the use of neural networks. In this paper, modified AlexNet is used with specified layers and parameters set in accordance with the characteristics of fruit images. The model parameters are to be optimized through the metaheuristic algorithm called the Golden Jackal Optimization Algorithm to attain a relative accuracy of 98% on the Fruit- 360 dataset, benchmarking against state of the art techniques in fruit classification. [12] Advanced techniques in the context of deep learning methodologies have been implemented for the enhancement of instance segmentation relative to sweet peppers. The advanced image preprocessing methods applied that were very important in the enrichment of images for analysis include sharpening, which assisted in enhancing the edges, reduction of salt and pepper noise, and reduction of Gaussian noise. After the integration of Swin Transformer inside the framework of Mask RCNN, it improved feature extraction capability for the identification of incidences of sweet pepper on complex backgrounds and under different lighting conditions. Once UNet3+ is introduced, the segmentation accuracy is further refined due to its efficient integration of shallow and deep features, making the delineation of sweet peppers from occlusions, like leaves and other peppers.
Ablation experiments were done to explicitly denote the effect of enhancements on the model, which made a much bigger improvement in most of the key metrics. It returned a very strong detection AP of 98.1% and AR of 99.4%, underscoring that it is highly resilient to very challenging environmental scenarios to correctly identify sweet pepper instances. The segmentation performance was also really noteworthy, with a very high F1 score, which is indicative of how the model can outline sweet pepper boundaries very accurately in different scenarios.
The applicability of [12] in practice will be real-time monitoring and management of sweet pepper growth in greenhouse environments with the most advanced deep learning techniques and attention mechanisms. Image acquisition, pre-processing, and annotation were carried out comprehensively to ensure varieties

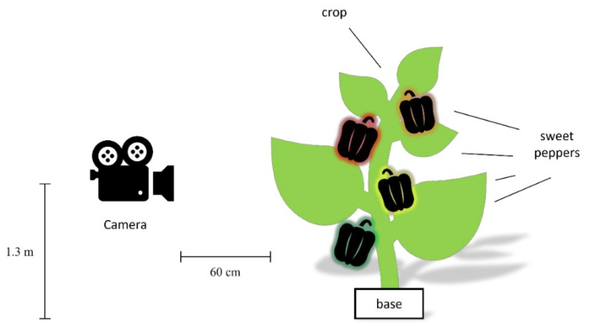


Fig. 4. *Image acquisition setup for capturing capsicum dataset, to ensure consistent image capture for dataset collection.*

of well curated datasets with 2,286 images from sweet peppers. Images were well prepared with Laplacian sharpening, noise reduction techniques, and other methods to ensure the best training conditions for the developed model.
The state of the art in agricultural automation is furthered, along with some important theoretical realizations of how to improve instance segmentation algorithms rather significantly. Integrating Swin Transformer and UNet3+, this work shows some very major strides toward a reliably accurate instance segmentation algorithm for sweet peppers, which is important in precision agriculture and greenhouse management practices.

## 5. EXPLORING CLASSIFICATION METHODS AND STRATEGIC APPROACHES

Earlier, classification of fruits and vegetables has benefited from latest machine learning techniques such as deep learning to improve accuracy and preciseness of the system. For example, the paper[14] explores a fusion of pre-trained convoluted neural networks (CNNs) and traditional Machine Learning algorithms towards tomato classification [14]. It focuses on aspects like size, color, shape, ripeness, and tolerance to defects in accordance with guidelines provided by bodies inclusive of the OECD and USDA as outlined. However, the problem of the tendency of overfitting still persists, as it emerged when training with the images of low intensity and minimal inter class color features [14]. In binary and multiclass datasets, the overall conformance of the proposed model is determined using metrics such as accuracy, recall, precision, specificity, and F1-score while proven to display comparatively higher accuracy with the existing methods used for fruit and vegetable Quality assessment [14].
Tackling the formation of the concatenated CNN model with the aim of classifying pepper diseases. This involves using a deeper network model, which is VGG16 and AlexNet networks for feature extraction, to reduce errors and improve the accuracy rate up to 100% in training and 97% in validation, and 95.82% in testing. This approach eradicates the time wasting aspect that is associated with the conventional visual observation done by farmers .This approach will enable early identification of diseases hence the ability to intervene to save food production in regions such as the Ethiopia agricultural related sector [17]This approach encounters challenges such as limited dataset of the initial study, possibility of over- fitting, sensitivity towards the quality of images, training and inference time being computationally intensive processes [17].
Computer vision and Deep Learning models are used for disease identification and Banana fruit grading . Steps like image

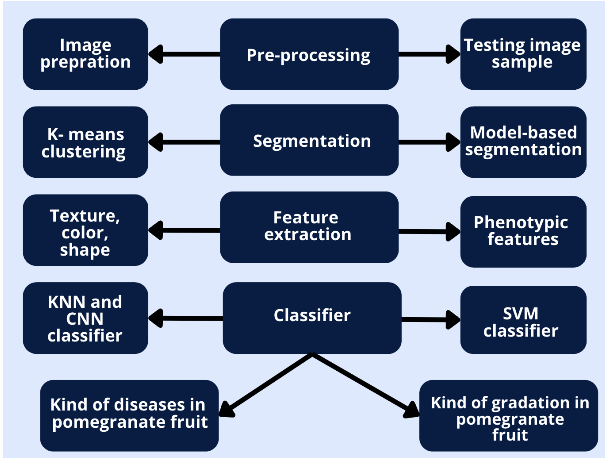


Fig. 5. *Description of End-toEnd Algorithm used in the system.*

acquisition, image preprocessing, image segmentation, feature extraction and finally classification done with the aid of artificial neural networks (ANNs) thus giving a maximum accuracy rate of 99.8% within disease diagnosis and grading as highlighted [16].In this approach, the laborious aspect of the conventional form and likely errors by the human eye are eliminated, making this mechanism efficient by saving time, and at the same time, achieving its goal of eradicating disease affecting banana plantations at a very low cost . The automated way of grading to complement the work done by the personnel is important in ensuring that the market standards and consumer satisfaction is maintained [16]. The procedure used in grading pomegranate fruits entails image capture, assessment of conditions, disease identification, and grading in terms of color and size [13]. These are the steps of image preprocessing, feature extraction, image denoising, and contrast enhancement that can be accomplished by mean filter and histogram equalization [13]. The proposed strategy of using CNN models for disease detection and grading so as to improve the harvesting management practices, increase productivity and reduce losses would go a long way in reaching the vision of sustainable agriculture [13]
Novel approach involving two main steps are Detecting olive fruits and employing a changed variant of K-Means algorithm and categorizing the fruits based on color and defects through the use of SVM and ANN techniques [15]. This method achieves high efficiency rates of 99.26% on a white background and 97.25% on a fruit conveyor, during an applicability test of the developed system for the automatic separation of olive fruits efficiency that other techniques and promising to be applicable in practice. Its high accuracy, potential issues such as generalization to unseen data, robustness to variability, real-world application performance, scalability, and cost effectiveness remain to be addressed [15].

## 6. PRECISION GRADING FOR FRUITS AND VEGETABLES

Sophisticated machine vision could make a difference in the sorting procedure of bell peppers by shifting from labor intensive, error prone manual methods to real-time, high accuracy sorting. The dataset contains images of bell peppers varying in color, size, greenness, damages, and diseases are collected one month before harvesting. Noise reduction, color normalization, and segmentation are some of the preprocessing steps taken to ensure the accurate extraction of features. The methodology integrates a variety of image processing methods for the improvement of quality and isolation of the bell pepper. The extracted key visual characteristics, which include color histograms, texture patterns and shape descriptors from contour analysis, guarantee efficient classification. Classification tasks include classifying into 5 categories using a multilayer perceptron artificial neural network with weights optimized through the backpropagation algorithm using gradient descent. The performance parameters, including accuracy, precision, sensitivity, and specificity, were 93.2%,86.4%, 84%, and 95.7%, respectively. This presents how well this model works in putting bell peppers into ripe, unripe or damaged categories.the MLP network of [18] after several hidden layer sizes had been tested, the best performance was obtained with 8 neurons, so its optimal structure was 6–8–5. This means there are 6 input factors and 5 output classes—GS1, GS2, RS1, RS2, UR—and 8 hidden neurons. Comparing the MLP network to an LDA model by ROC curves, one can see that the MLP's curve is closer to the ideal one, which means better prediction capability. The in depth analysis of the confusion matrix revealed that compared to the LDA model, which showed 90.4%, the MLP network showed a better result at correctly identifying GS1, with 92.5 percent, and RS1, with 93.4 percent; both did poorly for the immature class. MLP also did best in some other important metrics: accuracy 94.6%, precision 83.5%, sensitivity 85.4%, specificity 96.7%, compared to LDA with 93.9%, 75.5%, 76%, and 96% respectively. For real-time application, the MLP processing time of 0.2 seconds per sample was efficient enough to handle the speed of the conveyor belt and thus suitable for the proposed machine vision based system for bell pepper sorting. Therefore, considering higher accuracy and more efficient processing, the MLP network was chosen as the preferred classifier. Comparative analysis with respect to already existing approaches confirms the superiority of this approach, under pinned by visualizations through confusion matrices and ROC curves. Not withstanding, despite the challenges in handling environmental variability, the proposed technique improves sorting accuracy and efficiency of operations in agriculture. Future studies on even deeper integration of Convolutional Neural Networks should also be conducted to further enhance learning of features and increase accuracy in classification. This study, therefore, marks the potential for transformative machine vision in terms of agricultural automation and quality control, hence contributing towards global food security.

They improve the YOLOv5 architecture which is one of the most efficient object detection models with high real-time performance. The primary changes include the Mish activation function for the extraction of features and generalization, the DIoULoss for the bounding box regression to obtain higher precision, and the Squeeze Excitation (SE) feature map to increase the detection accuracy. In [20], large and thorough tests are performed with the given dataset with Samples of

red Fuji apples obtained from the markets and orchards of certain areas. The measures like accuracy, recall, and mean Average Precision (mAP) are used to measure the effectiveness of the improved YOLOv5 model. Following the results of the model, it can be concluded that its effectiveness is very high with 90.6% percent accuracy. Over the earlier versions of YOLO, such as YOLOv4 and YOLOv5s and have shown improved results than even SSD. The mAP score also stands as further proof that it can identify and rank apples according to standards such as ripeness, size, defects, and diameter. The model gives a grading accuracy of 93% in differentiating apples into grades that are recognized industrially including Grade-1, Grade-2, and Grade-3 [20]. In particular, the above mentioned system functions with the speed of 59.63 frames per second (FPS). This increases organization's operation effectiveness and decreases labor costs of manual grading methods. The experiment of adopting the enhanced YOLOv5 model has been illustrating considerable improvements in the automated apple grading process; at the same time, there are some drawbacks due to changes in apple types and changes in environmental conditions. Possible directions for future work can be directed toward increasing the efficiency of the model for the grading of other varieties of fruits, increasing the model's resilience to changes in lighting conditions as well as background, and using advanced methods in the visualization of both the model and the fruits.

Exploring to improve the current state of automated fruit identification by integrating deep learning techniques oriented towards high accuracy and efficient classification based on visual attributes[19]. Specifically, this approach is motivated by the fact that the proposed Fruit-360 dataset comprises a large collection of images accounting for 131 different classes of fruit. This dataset shall be very important in training and validating the proposed models to handle all shapes, sizes, and textures that are inherent in real images of fruits. In that regard, the research is trying to allow, through such a comprehensive data set, the needed robustness to tackle the problem resulting from variation in the fruit appearance within an automated classification system.

In Paper [19] The methodology involves tuning the deep network, AlexNet, which is an established CNN for image classification, to meet the specific requirements expected of fruit image classification. In particular, this will involve layer configuration changes and activation functions to ensure the best feature extraction. Besides, techniques like dropout layers will be used in order to avoid overfitting and guarantee generalization on new data. This paper uses the Golden Jackal Optimization Algorithm in attempting to optimize the parameters of the network so as to improve its capability of extracting discriminative features from fruit images.

Novelty includes introduction of the Fruit Shift Self Attention Transform Mechanism, in which self-attention mechanisms are fused with spectral transformation in such a way that both spatial and spectral features can be effectively captured by the proposed model. By focusing on both the spatial and spectral dimensions, FSSATM helps enhance the model's ability to cope with appearance variations between similar looking fruits and changes in illumination or backgrounds. Further, such strategies substantially improve overall accuracy and reliability according to [19] for fruit classification, making the system more practically useful in agriculture, food processing, and quality control.

The core model architecture envelops the modified AlexNet's strong feature extraction capability with the enhanced discriminative power that FSSATM made. Afterward, the performance of the trained model is rigorously tested using standard metrics such as accuracy, precision, recall, and F1-scores. Accomplishing an accuracy of 92.71%.This set of metrics provides insight into how the model identifies and classifies fruits across almost all scenarios. [19] uses normal instances with high accuracy rates, this research passes a strong message of just how effective the suggested methodology could be when embedded in any real-world applications requiring exact fruit classification for quality assessment and sorting.

[21] proposes a fully automatic system that applies the state of art deep learning techniques for classifying fruit images. The results of the studies demonstrate huge improvements over conventional

CNN approaches, thereby putting advanced algorithms and attention mechanisms in a bright light on studies that might improve the accuracy of classification tasks. Besides, the research also shows avenues for further studies, by the use of more complex architectures or adding of datasets, to leverage the proposed system herein in the development for generalization. Overall, the research points to continuous innovation in automated image recognition systems for real-life applications across various industries. ViTs are significantly better in benchmarking on datasets with little training data as ViTs achieve higher accuracy percentages (such as 90%) within considerably smaller sample numbers than the ones needed for CNNs, which in this case are 1500 samples for comparable benchmarks [21] . As for enhancement capabilities, ViTs' capacity to generate embeddings from the input images reduces the training complexity of the downstream classifiers such as SVM, logistic regression compared to the CNN-based counterparts. This approach does not only eliminate significant amounts of calculations but also helps to improve the model usability and deploy it in the limited access environments that can be characteristic for decentralized supply chains.

Thijs Defraeye in [21] has made a work on Vision Transformers (ViTs) for image based sorting and grading of agricultural products. Established the best way through which solutions to such issues could be developed in decentralized post harvest supply chain circumstances where access to such facilities is a limiting factor besides having limited access to large training datasets. The main incentive for this research is to make state of art sorting technologies more available to fruit producers and farmers and other small buyers with restricted means for typical deep learning solutions.[21] ViTs are proposed to replace CNNs in image-based tasks, which is explained in the paper's context. Here, ViTs are primarily developed for natural language processing (NLP) problems; however, they are applied to image data by encoding global dependencies. Unlike CNNs that work with local features, ViTs divide the input image into the patches being projected into embeddings through learnable linear transformations, and then apply Transformers with self attention mechanisms that allow it to capture long-range dependencies. Any image processing feature extractor will identify relevant details around the image, which will be convolutional layers of patterns such as edges and textures, and pooling layers that will actually simplify the image by retaining key parts. This requires a lot of data to train; therefore, pretrained feature extractors are used. The classifier takes such key details extracted to decide what the image represents. It uses simple layers for processing the information to make a decision, hence making it quicker and less data-intensive to train. Together, the feature extractor and the classifier will allow a computer to understand an image by identifying key patterns, allowing the determination of what this image shows. Benchmark their approach on two datasets:Fayoum University Banana dataset for its classification based on ripeness and another dataset known as CASC IFW for classifying apples based on the defects present on it. These datasets include the problems like complicated light conditions, noise, and non-consecutive image capture that are characteristic in the agricultural environment. They contrast ViTs' performance against traditional CNNs based on the classification achieved, especially in unique tasks including banana ripeness determination and apple defect identification. It offers more specific performance parameters in which ViTs perform as well as or better than CNNs while using a far smaller training sample. For example, with 500 samples ViTs have test accuracies over 90%, while CNNs may need about 1500 samples [21]. This serves as proof of ViTs' ability to perform well in situations with less data which is a characteristic of a decentralized agricultural setting.

ViTs' strength in capturing global dependencies and scene layouts, as pointed out in the study, are particularly beneficial for fruit defect detection and ripeness estimation. As it can be seen using the visualizations such as Principal Component Analysis [PCA] on the generated embeddings, the way ViTs operate ensures that different fruit classes are well separated and that the model is less sensitive to noise and variations from the environment. Shortcomings like, the imperfections in labeling the dataset, and ViT's basic and straightforward engagements in niche tasks without the opportunity to be fine tuned. Further work should compare the ViT's performance in more detailed classification problems, work on extending the ViT approach to problems that require defect localization or image segmentation, and incorporate the ViT embeddings with other data types for improving quality prediction in food supply chains.

## 7. OPTIMIZED IMAGE PROCESSING WITH MACHINE LEARNING

Sorting of fruits and vegetables on the edge and their classification are basic and very significant operations in modern agriculture that exist due to the development of new machine learning and image processing.

Chili pepper disease diagnosis via image reconstruction using GrabCut and generative Adversarial Serial Autoencoder, elaborates a new approach to diagnose chili pepper diseases through image reconstruction and scoring[22]. The proposed method achieves promising accuracy in disease diagnosis based on the error of the reconstructed images where various evaluation metrics are exceeded . They also demonstrate that improvement in normal as well as diseased pepper images is possible using this approach which helps in identifying diseased chilli pepper images [22]. For the background removal, the paper uses the Grabcut algorithm, and for scoring based on the errors between images, it applies the image reconstruction methods . Social issue of generating more data and improving the automation of the object-detection process for smart farming by applying SRGAN to increase the amount of diseased leaves data and improve performance is focused[22].

Emphasis is given to the enhancement of the crop harvesting process by successful peduncle recognition before capsicum removal from plant [23]. The detection system makes use of color and geometry descriptors that are obtained from the HSV color space and the PFH . Time a prediction takes to be produced, whereby it took approximately 40 seconds hence may not suit practical applications that would require instant results. To counter this, parallel processing is used to enhance on the pace of prediction. The system attains a qualitative success rate, specifically has an AUC of 0.71 for point-level segmentation, which proves that depth information could be utilized in the agricultural field [23]. DeepSORT algorithm to track sweet peppers with improved precision especially in the greenhouse setting [24]. This algorithm is used for following the sweet peppers that are found by the YOLOv5 model and identifying the object from frame to frame, dealing with occlusions and overlaps as well [24]. The paper presents promising results in detecting objects and counting the objects with confidence level of 0.973 for the object detection. DeepSORT algorithm was successfully applied for counting sweet peppers, demonstrating an accuracy level of 85.7% in two simulated environments such as varied lighting and inaccuracies in maturity level assessment.[24]. Nevertheless, from the farmer's prospective it is revealed that the system demands

| Maturity | Description | Image |
|---|---|---|
| Immature | Entirely green |  |
| Early-Mid | From 10–50% ripeness but no more |  |
| Mid-Late | From 50–90% ripeness but no more |  |
| Mature | Over 90% ripeness |  |

Fig. 6. *Grading Capsicum into four categories based on maturity.*

extensive computational power, thereby being unsuitable for small holders for instance [24].

Describes the possibility of using image processing techniques, PLC, and OPC server software for sorting spoiled apples on a conveyor belt. It manifests effectiveness in terms of organization, yet it has medium limitations with the differentiation of apples by light, for green and yellow apples, which might call for extra filters or modifications [25]. As a consequence of the study, they recommend that each unit should obtain high-definition cameras or else use parallel computing to increase the total efficiency of the system's performance [25]. Automated fruit classification is a modern application of algorithms and AI that help to identify fruits through their appearances and classify them based on their sizes, colors, textures, and shapes. It has an important role in agriculture and food industries, which use automatic sorting and grading for products, increasing accuracy with lower labor costs and delivering homogeneous quality products. Deep learning, which is a subset of AI, has been in the forefront across varied fields in recent times, involving agriculture.[26] Out of the several deep learning models, Convolutional Neural Networks are outstanding in carrying out tasks of fruit classification since they efficiently learn complex patterns from visual data. These models have effectively improved operations such as precision farming, crop disease detection, yield prediction, and resource optimization. The data used in [26] were sourced from the Web of Science database by searching with specific keywords, which produced 52 studies in the form of research articles and conference papers. Of these, 21 articles were selected for detailed analysis based on relevance related to the fruit classification task using deep learning. Researchers of [26] have applied a host of specialized deep learning model architectures in the fruit classification task, from some very well known structures, such as VGG16, ResNet, and MobileNetV2, to some very specialized models like Mask R-CNN, which was proposed for instance level segmentation. Each of these architectures allows the addressing of certain aspects of fruit classification better than others, hence making available a wide array of approaches that can perform tasks within this area. Public datasets, already existing ones, like Fruits-360 and privately self-built datasets by researchers, which can really perform deep learning model is used for training. Such datasets differ in size and degree of variety, thus affecting the generalization ability of models towards new fruits and different environmental conditions.

Data augmentation strategies enhance the robustness and generalization of deep learning models trained for fruit classification. They diversify training data through rotation, scaling, flipping, and addition of noise, avoiding overfitting and enabling the handling of variations in fruit appearance effectively.A common strategy in fruit classification is transfer learning, where these large-scale image datasets, such as ImageNet, are pre-trained with weight and used to initialize the models. This is majorly for the purpose of accelerating model training and improving performance by using previously learned features to accomplish a high accuracy rate in many studies.

Model performance measures in fruit classification studies are mainly done using accuracy metrics. Several reports return high accuracy scores of more than 95%, moreover, even higher than 99%. Following that, what appears to affect the model's performance is the dataset size, generally the more significant, the better, and the choice of the model architecture. Specialized architectures in specific fruit types show very high accuracy in classification tasks.

The challenges to fruit classification using deep learning relate to differences in dataset size, over-reliance on private datasets that permit no reproducibility, and class imbalance, where some fruits are more prevalent than others. These challenges impact the development and deployment of effective models in real-world applications. Deep learning based classification of fruit finds application in agriculture, food processing, and consumer services, and even in crop diseases and biodiversity conservation. These technologies improve operational efficiency by automating sorting processes and quality controls, including decision making tasks associated with the production and consumption of fruits.

Emerging trends and future directions in the field include the integration of multimodal sensing data, such as visual and spectroscopic data, for accuracy; the development of interpretable AI models that give transparency to decision making; and the enhancement of collaboration between AI systems and human experts in fruit classification tasks. These developments will further stabilize, adapt, and extend deep learning models to varied domains, solving current challenges and opening up future innovations [26].

## 8. CONCLUSION

Integrating Machine Learning and Deep Learning techniques has reshaped the fruit and vegetable classification landscape, yielding high accuracy, efficiency, and robustness compared to traditional methods. This review highlighted key algorithms such as YOLOv5, Faster R-CNN, and Vision Transformers , as well as segmentation methods like GrabCut and DeepSORT, and underscored their successes in real time detection, grading, counting, and disease diagnosis. Despite these advances, critical issues persist. Overfitting still undermines model generalization, computational costs pose challenge especially on resource limited device and data quality remains pivotal: large, diverse, well labeled datasets are expensive yet indispensable for robust performance.

Recent improvements, such as attention mechanisms, hybrid algorithms (e.g., combining Faster R-CNN with YOLO methods), and specialized feature-extraction approaches (e.g., ViTs with smaller datasets) have opened promising avenues for improving object detection accuracy and scalability. Nonetheless, further refinements are needed for small and dense object detection, addressing lighting and environmental variability, and enhancing speed without sacrificing accuracy. A more extensive evaluation on broader datasets and under diverse conditions would also strengthen the findings. Future directions therefore includes:

(1) Scalability**:** Developing low-computation or edge friendly architectures that can be deployed cost effectively in real world farm environments.
(2) Data Diversity and Quality**:** Building larger, carefully annotated datasets that reflect real world complexities various crops, diseases, and environmental conditions to reduce overfitting and boost generalizability.
(3) Hybrid Approaches: Combining ML/DL methods with domain knowledge or sensor data (e.g., multimodal sensing) to capture nuanced features like texture and subtle defects that purely visual systems might miss.
(4) Explainability and Usability**:** Translating model outputs into practical insights for farmers, integrating interpretability frameworks, and building user-friendly interfaces.

Overall, the continuous evolution of ML and DL in agricultural applications, informed by new algorithms, attention mechanisms, and more rigorous evaluations, will bring about reliable, scalable, and cost-efficient solutions. By addressing existing limitations and expanding testing scenarios, the community can further pave the way for smarter, more sustainable farming practices worldwide.

## 9. REFERENCES

[1] Trupti Mahendrakar, Andrew Ekblad, Nathan Fischer, Brian Kish, Markus Wilde, Ryan T. White, Isaac Silver "Performance Study of YOLOv5 and Faster R-CNN for Autonomous Navigation around Non-Cooperative Targets". In 2022 IEEE Aerospace Conference (AERO), Big Sky, MT, USA, 2022, pp. 1-12, doi: 10.1109/AERO53065.2022.984353 [March 2022].

[2] Xiaohong Kong Xinjian Li, Xinxin Zhu, Ziman Guo, Linpeng Zeng "Detection model based on improved faster-RCNN in apple orchard environment" Intelligent Systems with Applications 21(9):200325 [March 2024].

[3] Chetan M Badgujar, Alwin Poulose, Hao Gan "Agricultural Object Detection with You Look Only Once (YOLO) Algorithm: A Bibliometric and Systematic Literature Review" Computers and Electronics in Agriculture Volume 223, August 2024, 109090 [May 2024].

[4] Md Ahmed Al Muzaddid and William J. Beksi "NTrack: A Multiple-Object Tracker and Dataset for Infield Cotton Boll Counting" in IEEE Transactions on Automation Science and Engineering, doi: 10.1109/TASE.2023.3342791 [December 2023].

[5] Yagya Raj Pandeya, Samin Karki, Ishan Dangol, Nitesh Rajbanshi "Deep Learning based Tomato Disease Detection and Remedy Suggestions using Mobile Application" [August 2023].

[6] Zheng Wen-kui, Liu kun "Research on Edge Detection Algorithm in Digital Image Processing" Advances in Engineering Research, volume 123. 2nd International Conference on Materials Science, Machinery and Energy Engineering (MSMEE 2017) [May 2017].

[7] Md Khurram Monir Rabby; Brinta Chowdhury; Jung H. Kim "A Modified Canny Edge Detection Algorithm for Fruit Detection Classification" 2018 10th International Conference on Electrical and Computer Engineering (ICECE), Dhaka, Bangladesh, 2018, pp. 237-240, doi: 10.1109/ICECE.2018.8636811 [December 2018].

[8] Junqing Li,Ruiyi Han,Fangyi Li,Guoao Dong,Yu Ma,Wei Yang ORCID,Guanghui Qi and Liang Zhang "Apple Fruit Edge Detection Model Using a Rough Set and Convolutional Neural Network" Sensors (Basel). 2024 Apr 3;24(7):2283. doi: 10.3390/s24072283. PMID: 38610494; PMCID: PMC11014221 [ January 2024].

[9] Chuan Xu, Jianhuan Su, Teng Wang, Yinguang Zhang, Ao Yu "Improved Canny Edge Detection Operator" 2023 7th International Conference on Communication and Information Systems (ICCIS), Chongqing, China, 2023, pp. 135-140, doi: 10.1109/ICCIS59958.2023.10453881. [October 2023].

[10] Abdollah Zakeri, Ruhollah Hedayati, Mohammad Khedmati, Mehran Taghipour-Gorjikolaie "Classification of jujube fruit based on several pricing factors using machine learning methods" October 2021.

[11] Jesús Dassaef López-Barrios, Jesús Arturo Escobedo Cabello, Alfonso Gómez-Espinosa, Luis-Enrique Montoya-Cavero "Green Sweet Pepper Fruit and Peduncle Detection Using Mask R-CNN in Greenhouses" Appl. Sci. 2023, 13(10), 6296; https://doi.org/10.3390/app13106296 [March 2023].

[12] Peichao Cong, Shanda Li, Jiachao Zhou, Kunfeng Lv, Hao Feng "Research on Instance Segmentation Algorithm of Greenhouse Sweet Pepper Detection Based on Improved Mask RCNN" Agronomy 2023, 13, 196. https://doi.org/10.3390/agronomy13010196 [January 2023].

[13] Sultanabanu Kazi, Kazi Kutubuddin "Fruit Grading, Disease Detection, and an Image Processing Strategy" Journal of Image processing and Artificial Intelligence e-ISSN 2581-3803 Volume-9 issue-2 [June 2023].

[14] Hassan Shabani, Ahmed Abdel-Mawgood, Atsushi Shimada "Tomato Quality Classification Based on Transfer Learning Feature Extraction and Machine Learning Algorithm Classifiers"in IEEE Access, vol. 12, pp. 8283-8295, 2024, doi: 10.1109/ACCESS.2024.3352745 [January 2024].

[15] Nashaat M. Hussain Hassan, A.A Donkol, M. Mourad Mabrouk, A. M. Mabrouk "Highly Efficient Machine Learning Approach for Automatic Disease and Color Classification of Olive Fruits" IEEE Access, vol. 12, pp. 35683-35699 10.1109/ACCESS.2024.3362294 [February 2024].

[16] Himanshu B. Patel, Nitin J. Patil "An Intelligent Grading System for Automated Identification and Classification of Banana Fruit Diseases Using Deep Neural Network" International Journal of Computing and Digital Systems ISSN (2210-142X) [ February 2024].

[17] Yohannes Agegnehu Bezabh, Ayodeji Olalekan Salau, Biniyam Mulugeta Abuhayi, Abdela Ahmed Mussa, Aleka Melese Ayalew "CPD-CCNN: classification of pepper disease using a concatenation of convolutional neural network models" Scientific Reports 13(1) [September 2023].

[18] Khaled Mohi-Alden, Mahmoud Omid, Mahmoud Soltani Firouz, Amin Nasiri “A machine vision-intelligent modeling based technique for in-line bell pepper sorting” information processing in agriculture 10 (2023) 491– 503 [December 2023].

[19] Mrs Arunadevi Thirumalraj, B. Rajalakshmi, B Santosh Kumar, S. Stephe “Automated Fruit Identification using Modified AlexNet Feature Extraction based FSSATM Classifier” 10.21203/rs.3.rs-4074664/v1. [March 2024].

[20] Bo Xu, Xiang Cui, Wei Ji, Hao Yuan, Juncheng Wang “Apple Grading Method Design and Implementation for Automatic Grader Based on Improved YOLOv5” Agriculture, vol. 13, no. 1, p. 124, Jan. 2023, doi: 10.3390/agriculture13010124 [December 2022].

[21] Manuel Knott, Fernando Perez-Cruz, Thijs Defraeye “Facilitated machine learning for image-based fruit quality assessment” Journal of Food Engineering 345 (2023) 111401 [May 2023].

[22] Jongwook Si, Sungyoung Kim ”Chili Pepper Disease Diagnosis via Image Reconstruction Using GrabCut and Generative Adversarial Serial Autoencoder” [June 2023].

[23] Inkyu Sa, Chris Lehnert, Andrew English, Chris McCool, Feras Dayoub, Ben Upcroft, and Tristan Perez “Peduncle Detection of Sweet Pepper for Autonomous Crop Harvesting - Combined Colour and 3D Information”IEEE Robotics and Automation Letters PP(99):1-1 [ January 2017].

[24] Luis David Viveros Escamilla, Alfonso Gómez-Espinosa, Jesús Arturo Escobedo Cabello, José Antonio Cantoral-Ceballos “Maturity Recognition and Fruit Counting for Sweet Peppers in Greenhouses Using Deep Learning Neural Networks” Agriculture 2024, 14(3), 331; https://doi.org/10.3390/agriculture14030331 [June 2020]

[25] Onur AYAN, D. Zeynep DEMİREZ, H. Kagan KIZILOZ, Gizem INCI, Seckin ISLEYEN, Semih ERGIN “The Detection of Spoiled Fruits on a Conveyor Belt Using Image Processing Techniques and OPC Server Software” International Journal of Computational and Experimental Science and Engineering 4(1):11-15 [March 2018].

[26] Yonis Gulzar, Zeynep Unal, Shahnawaz Ayoub, Faheem Ahmad Reegu, Alhanouf Altulihan “Adaptability of deep learning: datasets and strategies in fruit classification” BIO Web of Conferences 85, 01020 (2024) [January 2024].