Comparative Analysis for Predicting Shelf life of Fruits Using Advanced Deep Learning Approaches

Sanath Shenoy
Data Scientist
Dept of AI and Learning systems,
Fresko Innovation Labs PVT LTD
Bangalore, India
sanath@gropare.com

Radhika Mishra
Data Scientist
Dept of AI and Learning system,
Fresko Innovation Labs PVT LTD
Bangalore, India
radhika@gropare.com

Ruchi Chaturvedi
Junior AI Researcher
Dept of AI and Learning systems.
Fresko Innovation Labs PVT LTD
Bangalore, India
ruchi@gropare.com

Krushnakant Bhagwat
Junior AI researcher
Dept of AI and Learning systems.
Fresko Innovation Labs PVT LTD
Bangalore, India
krushnakant@gropare.com

***Abstract*—The food industry aims to reduce food waste and ensure the delivery of fresh produce to consumers, making it crucial to predict fruit shelf life accurately. Traditional approaches rely on expensive and time-consuming laboratory testing, which often involves destructive methods. However, recent studies suggested that advanced deep learning techniques can predict fruit shelf life accurately and efficiently. This paper presents a novel approach to predicting fruit shelf life using deep learning models. The study focuses on the application of these advanced techniques to forecast the shelf life of bananas, which can contribute significantly to achieving the food industry's objective.The study tries to develop accurate and efficient models that could predict the maturity of bananas, based on their average shelf-life and appearance. In order to achieve this objective, two object detection algorithms—Faster R-CNN and You Only Look Once (YOLO) are used and their performance is compared in the present research. The dataset has been created by collecting images of the life cycle of bananas and segregating them based on their maturity. Various preprocessing and augmentation techniques have been applied to enhance the features of the training dataset which is useful to get better accuracy.**

The algorithms were trained on the family of Cavendish Bananas dataset and were able to predict the shelf life of bananas with better training accuracy. The YOLO algorithm which is known for efficiency is compared with Faster R-CNN well known for identifying very fine features. This study demonstrates the potential of deep learning algorithms in predicting the shelf life of bananas and can be extended to different fruits.

Keywords—deep Learning, shelf life, Faster R-CNN, YOLO, computer vision, deep Learning, Image processing

I. Introduction

Recently, several studies on the shelf life of bananas and categorization have been conducted using clustering and classification to estimate banana ripeness and shelf life. Good shelf life considers both fruit safety and quality as significant factors. Even so, bacteria have frequently been kept an eye on throughout shelf-life research.

The categorization of fruits and the identification of fruit quality are currently the primary objectives of multiple research projects. Recently, sorted peeled pistachios using computer vision and color characteristics. In addition, hyperspectral imaging has been shown in studies to be able to identify strawberry erosion, which may be used in the online sorting method[1].

Deep Learning has recently been used to reliably recognise fruits. Some studies have discovered that using a faster R-CNN architecture in the research allowed it to recognise fruits and provide remarkably accurate detection capabilities. Fruit maturity is a crucial criterion for determining shelf life. In the traditional methods, it is essentially an individual's subjective judgment depending on their level of individual's experience. It is possible to differentiate between both subjective and objective approaches for evaluating the fruit shelf life, each with advantages and disadvantages.

This research aims to estimate the shelf-life of Cavendish Banana using object detection methods based on Deep Learning models. The objective is to simplify the efforts and to provide a significantly cost-effective method making it easy to implement and affordable throughout the lifecycle of the food supply chain.

II. RELATED WORK

The goal of the paper [2] was to detect intrinsic features of fruits such as internal defects, bruises, texture, and color and classify fruits according to their remaining useful life (RUL). The study uses the data of 'kesar' mango[3]. It uses thermal imaging to determine the intrinsic values of fruits in terms of temperature. Furthermore, a transfer learning approach is combined with thermal imaging techniques to enhance the accuracy of fruit shelf life prediction. The study compares three lightweight CNN-based models, namely SqueezeNet[4], ShuffleNet[2], and MobileNetv2[5]. The results demonstrate that the highest achievable accuracy of up to 98.15% is obtained. It is also observed in the study that using thermal images resulted in a significant reduction in training time.

In the paper [6], the aim is to predict the ripeness level and CO2 respiration rate (RRCO2)[7] of the 'kesar' family of mangoes using Artificial Intelligence. To achieve this goal, The study uses a deep learning algorithm that was trained on 1524 images of the fruit. The data used was divided into four classes: 'unripe', 'early-ripe', 'partially-ripe' and 'ideally-ripe'. In progression to this, the research correlates 'RRCO2' and the ripeness level of the Mango. The prediction accuracy using 'VGG-16[8]' on the training dataset was 99%, and the test data set was 96.2%.

Near-infrared spectroscopy was used in [9] to analyze the quality of apples at different stages. Unfortunately, there are very limited literatures on fruit shelf life prediction. This research is focused on building a real-time, self-learning model that considers changing information from observations made at the unit level of a fruit throughout every step of the supply chain. The research aims to assess how well a self-learning model predicts storage life and how the existing supply chain can be made more efficient. The accuracy of prediction is close to 98.15%.

The paper [10] aims to predict the maturity and quality in terms of the shelf life of fruit. The fruit used was a Banana. The study used a total of 2100 images that were divided into 3 classes: ripe, unripe, and over-ripe, with each containing 700 images. Additionally, it used two sets of datasets. Convolutional neural networks (CNN) and AlexNet[11] algorithms were used to achieve the goal, and the study concluded that CNN was a more suitable algorithm for the dataset used in the research, and its highest accuracy obtained was 99.36%.

It can be concluded that various methods proposed in the above-mentioned research papers differ from those used in this study. This research involves estimating of shelf life of banana using object detection techniques, namely Faster RCNN and YOLOv5. It compares the performance of both the models on various grounds, while the above methods proposed by different research studies involve VGG-16, SqueezeNet ShuffleNet, MobileNetv2, CNN and AlexNet.

III. PROPOSED METHODOLOGY

The methodology section of our paper presents an in-depth discussion of our proposed methodology employed in the study, which involved the following steps: data collection, image annotation, preprocessing, model selection, model training, model testing, result visualization, and analysis.
Each step was designed to ensure that the study was conducted systematically.

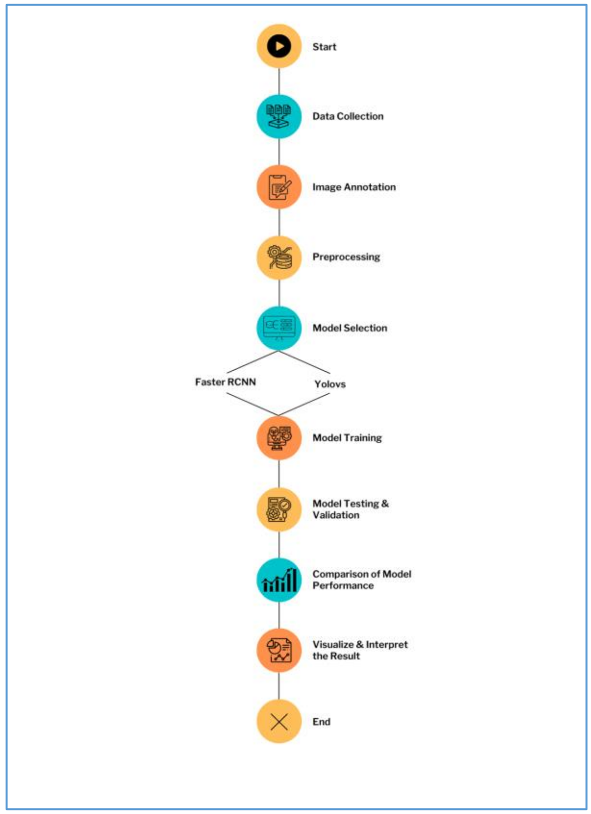

Figure 1:Flowchart of the proposed methodology

A. *Dataset Collection:*

The data acquisition process for this research involved taking images of bananas daily for a period of six days from day 0 to day 5. The images were taken under a standard white light at 4:00 pm GMT+5.5H each day to ensure consistent lighting conditions. The purpose of this data acquisition was to study the changes in the bananas over time and use advanced deep-learning approaches to predict the shelf life of the banana.

Preparation of the set-up involved using a standard white light source in a room and a table was placed underneath it. The bananas were placed on the table in a single layer. Videos of the bananas were captured daily using a digital camera. The camera was set up in a fixed position to ensure consistent video capture from the same angle. The videos were 30 seconds long. Image Extraction: From the recorded video, images were extracted at an interval of 15 seconds.

By following these steps, we acquired a high-quality dataset suitable for use in our comparative analysis for predicting the shelf life of bananas using advanced deep-learning approaches.

Figure 2:Snapshots of the Cavendish Banana Dataset

B. Annotation

After acquiring the images, we conducted image annotation for a Cavendish banana set, using five labels to categorize the bananas based on their shelf life. The labels used were d1, d2, d3, d4, and d5.

TABLE 1 DESCRIPTION OF SHELF LIFE LABELS

label name	description
d1	fruit with 1 day shelf life remaining
d2	fruit with 2 days shelf life remaining
d3	fruit with 3 days shelf life remaining
d4	fruit with 4 days shelf life remaining
d5	fruit with 5 days shelf life remaining

The annotation process was conducted using bounding boxes, which allowed for accurate identification and labelling of the bananas in the images.
The d1 label was used to indicate bananas that had one day left before becoming overripe, while the d2 label was used for bananas that had two days left, and so on. After the annotation process was complete, the annotated data was exported to COCO[12] and PyTorch[13] labelling formats, making it accessible to a wide range of object detection models. The COCO format was used for the Faster R-CNN[14] model, while the PyTorch format was used for the YOLOv5 model[15].

C. Preprocessing

Post data annotation, we conducted image preprocessing to prepare the annotated dataset. The images were first resized to a standardized size of 640x640 pixels to ensure consistency and facilitate processing. The annotated dataset was then divided into train, test, and validation sets to enable the training and evaluation of models. The trainset was used to train the models and the test set was used to evaluate their performance on unseen data. Finally, the validation set was used for unbiased evaluation of the model. The image preprocessing step helped to ensure that the CNN models had access to consistent and appropriately-sized data, which is essential for achieving reliable and accurate results.

D. Proposed Architecture

In this study, we have utilized two popular object detection models, YOLOv5s and Faster R-CNN. In this section, we discuss the comprehensive overview of these two models' architecture and the specific features that make them suitable for our use case of shelf detection of a fruit. Each aspect of the architecture is explained in detail in the future sections.

E. YOLOv5 Algorithm:

YOLOv5 is the version of the YOLO family of algorithms, which has been shown to achieve state-of-the-art performance on several benchmark datasets. This section provides an overview of the YOLOv5s architecture, a smaller and faster variant of the YOLOv5 algorithm. This section will discuss the backbone, neck, and head architecture of YOLOv5s and how it can be used for the detection of Cavendish bananas.

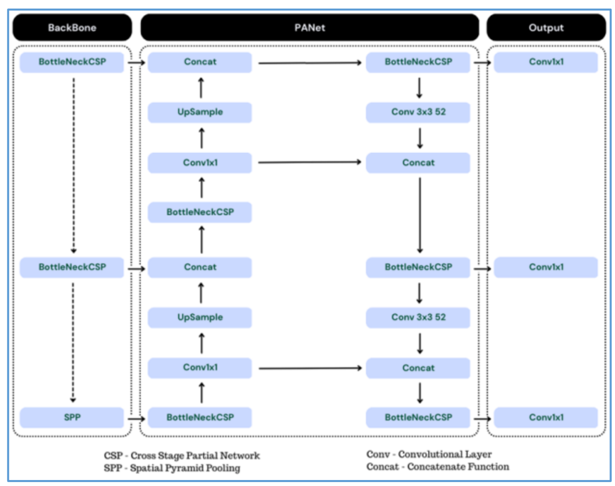

Figure 2: Flowchart depicting Neural Network

a) CSPDarknet53 Backbone

The CSPDarknet53[16] backbone is the feature extraction part of the YOLOv5s object detection model, which is a modified version of the Darknet network architecture. It is composed of a series of convolutional layers, with the addition of a cross-stage partial connection (CSP) module at each stage of the network. The CSP module splits the feature maps from each stage into two paths, one for processing and the other for bypassing. This allows for better information flow and faster training. In addition, the backbone generates high-level feature maps of the input image, which are then passed to the neck.

The CSPDarknet53 backbone in YOLOv5s can be used to extract high-level features from the input image of Cavendish bananas, such as the shape and texture of the bananas. This information can help the model to distinguish between ripe and unripe bananas, which is important for determining the shelf life of the banana. The backbone can also capture the color features of the bananas, such as hue and saturation.

b) SPP and PANet in the Neck

The neck of YOLOv5s is composed of two main components: spatial pyramid pooling (SPP)[17] and path aggregation network (PANet)[18]. SPP is used to extract features from different regions of the feature maps generated by the backbone at different scales to improve detection accuracy. The SPP component in YOLOv5s uses three levels of pooling with different scales (1x1, 2x2, and 3x3). PANet is used to aggregate features across different levels of the feature pyramid to further improve detection accuracy. The PANet component in YOLOv5s uses a series of lateral connections and feature fusion modules to aggregate features from different levels of the feature pyramid.

These components in the neck of YOLOv5s can be used to extract features from different regions and scales of the input image of Cavendish bananas, which can help to detect subtle changes in color and texture that may indicate the ripeness of the banana. For example, the SPP component can extract features from small regions of the bananas, such as the tips, which may exhibit different colors or textures compared to the rest of the banana. The PANet component can then aggregate these features from different scales to make a more accurate prediction of the shelf life of the bananas.

c) Head

The head in YOLOv5s is the final stage of the neural network that produces the output of the algorithm, i.e., the predicted bounding boxes and associated class probabilities for the objects detected in the input image. The head in YOLOv5s is a single convolutional layer that takes the feature maps generated by the neck as input and applies a series of convolutions to generate a set of anchor boxes, each of which is associated with a set of class probabilities. The head layer is designed to balance the trade-off between accuracy and speed.

The head layer in YOLOv5s can be used to generate the final predictions of the shelf life of the Cavendish bananas. The head layer can take the feature maps generated by the neck as input and apply a series of convolutions to generate a set of anchor boxes for the bananas. Each anchor box can be associated with a set of class probabilities, such as d1, d2, d3, d4 and d5, based on the color and texture features extracted from the backbone and neck layers.

In our research we experiment with the YOLOv5 architecture to detect the shelf life of Cavendish bananas. The backbone of this architecture can extract properties such as high-level features, color and contrast settings from the input image. The neck layers can then extract features from different regions and scales of the input image, which can help to detect subtle changes in color and texture that may indicate the ripeness of the bananas. Finally, the head layer can generate the final predictions of the shelf life of the bananas based on the features extracted from the backbone and neck layers.

F. Faster RCNN Algorithm:

Faster R-CNN is a state-of-the-art object detection algorithm that can accurately detect and classify objects within an image.

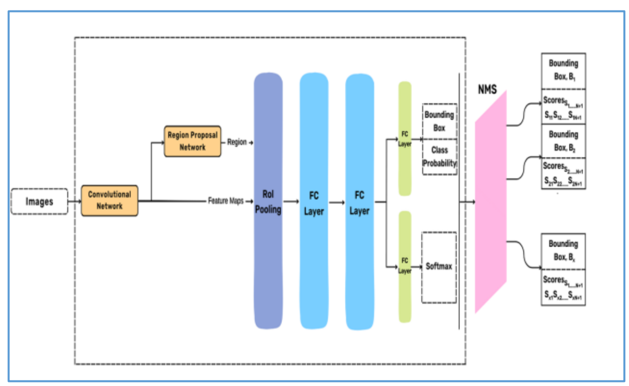

Figure 3: Architecture of Faster RCNN

It uses a two-stage approach that first generates region proposals and then predicts the class and location of objects within those proposals. This approach is faster and more accurate, hence a very good choice for a wide range of computer vision tasks.

a) Feature Network

A pre-trained image classification network, such as VGG[19], is used as the Feature Network. The network's purpose is to extract useful features from the input image. The Feature Network takes in the raw image as input and outputs a feature map with the same shape and structure as the original image. The output feature map can be considered a compressed representation of the original image that is more informative and easier to process.

b) *Region Proposal Network (RPN)*

The RPN[20] is a simple neural network with 3 convolutional layers. The network has a shared layer that is fed into two separate layers for classification and bounding box regression. The RPN's goal is to generate Region of Interest (ROI)[21] proposals, which are a set of bounding boxes that are most likely to contain the objects of interest. The RPN outputs a set of bounding boxes with pixel coordinates of two diagonal corners and a classification score (-1 for ignored, 0 for background, or 1 for object).

c) *Detection Network (Fast RCNN)*

The Detection Network takes input from the Feature Network and RPN to generates the final class and bounding box predictions. The network consists of 4 fully connected or dense layers. There are two shared layers followed by classification layer and a bounding box regression layer. In order to classify only the inside of the bounding boxes, the features are cropped according to the bounding boxes. The output of the Detection Network includes the class of the detected object and the coordinates of the bounding box.

To predict the shelf life of Cavendish bananas, we are using Faster-RCNN. This involves using a pre-trained image classification network to extract useful features from the images, followed by a Region Proposal Network to generate bounding boxes of regions of interest of the banana where color changes and feature changes have occurred. Finally, a detection network crops the features according to the bounding boxes and uses them to predict the remaining shelf life of the banana. Training this model on a dataset of images of bananas at different stages of shelf life, we could accurately predict the shelf life of the banana.

G. *Model Training:*

In our study, we are using two architectures for training. Here are the details of the parameters considered for both models. Two object detection models, YOLOv5s and Faster R-CNN, were trained on Google Colab using NVIDIA T4 Tensor Core GPU with 15 GB memory. Both models were trained on 164 images with a resolution of 640x640. Both models were evaluated using mAP50 and mAP50-95 metrics. YOLOv5s was trained for 250, 300, 500, and 1000 epochs, whereas Faster R-CNN was trained for 250, 300, 500, and 1000 iterations. The weights used for YOLOv5s were YOLOv5 weights, and for Faster R-CNN, COCO Faster R-CNN weights were used. The models' performances were evaluated based on correctly classified, misclassified, and unclassified data for each day.

IV. Result and Discussion

In our experiments YOLOv5 and FasterRCNN were custom trained. The dataset consisted of 254 images. Each model was trained using 164 images distributed among 5 classes for predicting the shelf life.Additionally, 45 each were selected for validation and test .

Table 2 Experimental results of shelf-life detection for YOLOv5 and Faster RCNN

model	**epoch/iter**	**mAP50**	**mAP50-95**
YOLOv5	250	99.5	95.8
	300	99.5	95.6
	500	99.5	96.3
	1000	*Optimizer stopped	*Optimizer stopped
Faster Rcnn	250	9.6	3.4
	300	9.06	6.31
	500	96.3	82.3
	1000	97.1	81.3

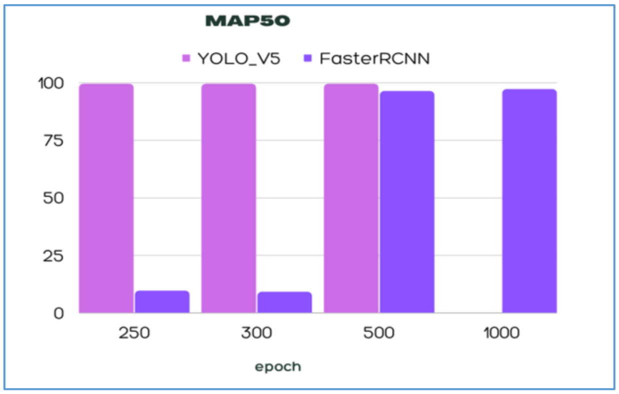

Figure 4: Comparison of mAP 50 metric for YoloV5 and Faster RCNN

A detailed explanation of the classes for the prediction of shelf life has been explained in the previous section. Both models were trained with a constant batch size of 16 images. A set of 4 experiments were performed for epoch or iteration of 250, 300, 500, and 1000 for YOLOv5 and FasterRCNN, respectively. The performance metrics we choose are mAP50 and mAP50-95. These are the standard metrics used to measure the detection performance. In our study, we measure the performance using inference.

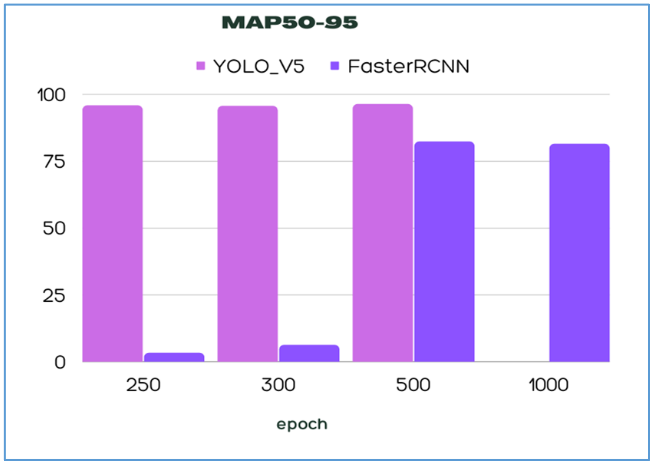

Figure 5: Comparison of mAP 50-95 metric for YoloV5 and Faster RCNN

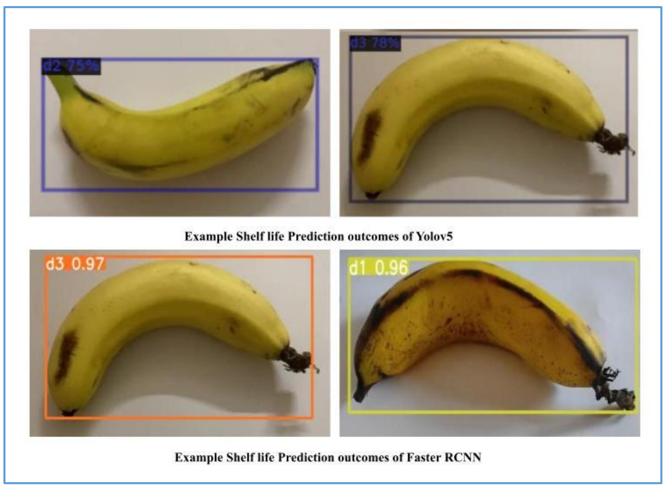

Figure 6 Shelf life prediction samples from YOLO and Faster RCNN

V. CONCLUSION

The experimental results of both the models show that object detection models with customization can definitely help in real-life use cases of shelf life detection of bananas or fruit of interest with an acceptable level of accuracy. Further, the results also show that both Yolov5 and Faster RCNN were able to predict the shelf life with both mAP50 and mAP50-95 greater than 80%.

It is also observed that faster RCNN performs better at a higher number of iterations and requires higher computing hardware. On the other hand, Yolo provides model variants which do not demand high computer hardware. We also see that the first attempt towards the use of object detection in fruit shelf life detection is greatly satisfactory this also opens new opportunities for further research in other use cases of the Fruit and Vegetable domain.

VI. FUTURE WORK

Potential future research in this field may include:

- Exploration of alternative deep learning architectures: Although this study compared the performance of two deep learning approaches, there are other architectures that could be explored to predict the shelf life of Cavendish bananas.
- Incorporation of external data sources: This study exclusively utilized internal banana quality metrics to predict shelf life, but incorporating external data sources such as temperature, humidity, or transportation time could enhance the accuracy of the predictions.
- Inclusion of ripeness level: Assessing the ripeness level of bananas at the time of testing through manual inspection or computer vision techniques and incorporating this feature into the prediction model could be explored to determine its impact on prediction accuracy.
- Enhancement of prediction accuracy: Despite the good performance of deep learning models in predicting fruit shelf life, researchers can strive to refine the model architecture or incorporate additional data parameters to further enhance prediction accuracy.
- Real-time monitoring: To minimize food waste and improve food safety, predicting the shelf life of fruits at the point of purchase or during storage through real-time monitoring systems utilizing deep learning techniques could be developed.
- Integration with supply chain management: To optimize inventory management and reduce food waste, integrating deep learning models into supply chain management systems to predict fruit shelf life may be investigated in future studies.

ACKNOWLEDGMENT

We offer our sincere gratitude to all people who have contributed to this research project. We also thank Greeshika Sharma for her unwavering support in this study. We also thank the staff members who assisted us in recruiting participants and conducting the study.

We thank our families, friends, and colleagues who have provided us with encouragement and support throughout the

research process. Their unwavering belief in us has been invaluable, and we are grateful for their constant support.

References

[1] Omid, Mahmoud & Soltani Firouz, Mahmoud & Nouri-Ahmadabadi, Hosein & Mohtasebi, Seyed. (2017). Classification of peeled pistachio kernels using computer vision and color features. Journal of the Science of Food and Agriculture. 10. 259–265. 10.1016/j.eaef.2017.04.002.

[2] Bhole, V., & Kumar, A. (2021). A Transfer Learning-based Approach to Predict the Shelf life of fruit. Inteligencia Artificial, 24(67), 102–120. https://doi.org/10.4114/intartif.vol24iss67pp102-120. 5

[3] Ramani, M & Tarpara, Vrajlal & Swaminathan, B. & Manasi, P & Pokiya, N. (2019). COST OF CULTIVATION AND PROFITABILITY OF KESAR MANGO CULTIVATION IN SAURASHTRA REGION OF GUJARAT, INDIA.

[4] Iandola, Forrest N., et al. "SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and< 0.5 MB model size." arXiv preprint arXiv:1602.07360 (2016).

[5] Shahi TB, Sitaula C, Neupane A, Guo W (2022) Fruit classification using attention-based MobileNetV2 for industrial applications. PLoS ONE 17(2): e0264586. https://doi.org/10.1371/journal.pone.0264586

[6] Dutta, J., Deshpande, P. & Rai, B. AI-based soft-sensor for shelf life prediction of 'Kesar' mango. SN Appl. Sci. 3, 657 (2021). https://doi.org/10.1007/s42452-021-04657-7 5

[7] Bhande, S.D. & Ravindra, M.R. & Goswami, T.K.. (2008). Respiration rate of banana fruit under aerobic conditions at different storage temperatures. Journal of Food Engineering. 87. 116-123. 10.1016/j.jfoodeng.2007.11.019.

[18] S. Liu, et al., "Path aggregation network for instance segmentation" in Proc. IEEE Conference on Computer Vision and Pattern Recognition, 2018, 8759-8768 [doi:10.1109/CVPR.2018.00913].

[19] K. Simonyan and A. Zisserman, 'Very deep convolutional networks for large-scale image recognition.' arXiv Preprint ArXiv:1409.1556, 2014.

[20] S. Ren, et al., "Faster r-cnn: Towards real-time object detection with region proposal networks," Adv. Neural Inf. Process. Syst., vol. 28, 2015.

[21] L. Rossi et al., "A novel region of interest extraction layer for instance segmentation" 25th international conference on pattern recognition (ICPR). IEEE, 2020, p. 2203-2209 [doi:10.1109/ICPR48806.2021.9412258].

[8] Tammina, Srikanth. (2019). Transfer Learning using VGG-16 with Deep Convolutional Neural Network for Classifying Images. International Journal of Scientific and Research Publications (IJSRP). 9. p9420. 10.29322/IJSRP.9.10.2019.p9420.

[9] J.C. Fan, G.M. Zhou, "A Near infrared spectroscopy qualitative analysis of apple shelf life", Food and Nutrition in China, 17:47-49, 2011. doi: https://doi.org/10.3390/molecules200813603.

[10] Aherwadi, N.; Mittal, U.; Singla, J.; Jhanjhi, N.Z.; Yassine, A.; Hossain, M.S. Prediction of Fruit Maturity, Quality, and Its Life Using Deep Learning Algorithms. Electronics 2022, 11, 4100. https://doi.org/10.3390/electronics11244100 1

[11] Heinrich, Kai & Janiesch, Christian & Möller, Björn & Zschech, Patrick. (2019). Is Bigger Always Better? Lessons Learnt from the Evolution of Deep Learning Architectures for Image Classification.

[12] D. Nagle, "An implementation of coco,", SIGPLAN Fortran Forum. New York, NY, USA: ACM, vol. 20, no. 1, 2-5, 2001 [doi:10.1145/375738.375740].

[13] V. Subramanian, Deep Learning With PyTorch: A Practical Approach to Building Neural Network Models Using PyTorch. Packt Publishing Ltd, 2018.

[14] R. Girshick, "Fast r-cnn" in Proc. IEEE International Conference on Computer Vision, 2015, 1440-1448 [doi:10.1109/ICCV.2015.169].

[15] J. Redmon, et al., "You only look once: Unified, real-time object detection" in Proc. IEEE Conference on Computer Vision and Pattern Recognition, 2016, 779-788 [doi:10.1109/CVPR.2016.91].

[16] A. Bochkovskiy et al., 'Yolov4: Optimal speed and accuracy of object detection.' arXiv Preprint ArXiv:2004.10934, 2020.

[17] K. He, et al., "Spatial pyramid pooling in deep convolutional networks for visual recognition," IEEE Trans. Pattern Anal. Mach. Intell., vol. 37, no. 9, pp. 1904-1916, 2015 [doi:10.1109/TPAMI.2015.2389824].