Transfer-learning based multi-class areca nut image classification under uncontrolled lighting on a conveyor system for automated sorting

Avinash Kumar K M
IIT Dharwad
Dharwad, India
avinash.kumar.20@iitdh.ac.in

Aishwary Singh
IIT Dharwad
Dharwad, India
223031001@iitdh.ac.in

Rahul L. Maurya
IIT Dharwad
Dharwad, India
ME20MS001.alum24@iitdh.ac.in

D. Narasimha
IIT Dharwad
Dharwad, India
d.narasimha@iitdh.ac.in

Samarth S. Raut
IIT Dharwad
Dharwad, India
sraut@iitdh.ac.in

2024 IEEE Conference on Engineering Informatics (ICEI) | 979-8-3315-0577-6/24/$31.00 ©2024 IEEE | DOI: 10.1109/ICEI64305.2024.10912217

***Abstract*— Areca nut is an important cash crop with high cultural and economic significance to the Indian demography. The level of ripeness can be visually assessed based on the color and is an important deciding factor for the market value of the crop and subsequent post-harvest processing such as drying, de-husking, etc. The quality assessment and segregation task, when carried out by manual labor is subjective, inconsistent, and prone to be tedious with inadvertent errors and loss in the desired efficiency. The advent of computer vision (CV) - -based segregation techniques along with deep learning (DL) neural network (NN) models can augment the manual efforts in the segregation process in the agri-food sector. A fast, low-cost conveyor-based system that performs multi-class classification of areca nuts based on color and a preliminary image dataset of areca nut images captured using the same is presented here, along with explorations of some pre-trained neural network models. A feeding mechanism, a conveyor system, and a sorter system were designed and fabricated using traditional machining and 3D printing processes. A dataset of RGB images of harvested areca nuts comprising around 1600 images of areca nuts on the conveyor belt under uncontrolled lighting conditions was generated to simulate the challenging industry-relevant sorting scenarios. Transfer learning using pretrained models like Inception-V3, Resnet50 and others was utilized to implement the classification. Using contour detection and Hough circle-based automated cropping, and manual cropping, four different versions of the original dataset created were used along with two variations in dataset split and two learning rates. Inception-V3 exhibited the highest testing accuracy of 91.98% for the 80-10-10 train-cross-validation-test percentage split with a 0.0001 learning rate, with Resnet50, Inception-Resnetv2 and VGG16 exhibiting 68.52%, 90.74% and 88.27% respectively. All methods of image cropping, including manual, adversely affected all models' classification accuracy and performance metrics, possibly owing to loss of information during cropping or file conversion operations. VGG16 was found to be more sensitive to the cropping of RoI in the image than other models. Thus, for transfer learning, the dataset preprocessing needs careful consideration.**

Keywords- areca nut, fruit quality inspection, transfer learning, image classification, F1 score

I. INTRODUCTION

Machine learning as a tool has become ubiquitous, with its implementation limited not just to engineering research and development applications but also to industrial sectors, process control and automation, logistics, healthcare, etc. Artificial intelligence has also revolutionized manufacturing with smart sensors, IoT devices, and improved data storing and processing techniques [1]. Advances in image processing techniques, coupled with the development of sophisticated electronic hardware and sensing technologies, have drastically improved various domains ranging from healthcare services to agriculture. Machine learning has been instrumental in disease detection, drug development, and, in recent times, robot-assisted surgeries and precision agriculture [2] [3] [4] [5]. Agricultural practices have recently been boosted by integrating computer vision-based techniques for various activities like pest control, disease monitoring, and harvesting. The process of fruit quality control, which comprises picking, sorting, and grading, has been predominantly labour-intensive. The involvement of humans in any process lacks an unbiased outcome due to subjective assessment. The limited shelf life of fruits, coupled with the change in their quality over time, warrants an accurate automated quality control process with high fidelity. For optimal automated post-harvest processing of areca nut, such as lot-specific customised dehusking for improving yield, computer vision-based machinery for segregation is important.

II. LITERATURE SURVEY

Computer vision has brought contact-less run-time consistency and objectivity to the segregation and grading process. Several machine learning models based on Convolution Neural Networks (CNN) have been developed for grading of fruit quality. Torres et al. [5] have extensively reviewed using CNN in agricultural practices, particularly in fruit detection, classification, quality control, and disease detection. They have highlighted the recent uptrend in implementing pre-trained CNN models with a brief

Funding supports from Govt. of India SERB Start-Up Research Grant SRG/2020/002513 and IIT DHARWAD Seed Grant and Networking Fund (SGNF), both awarded to Prof. Samarth S. Raut (SSR), are acknowledged to build the hardware and software set-up. BITS BioCyTiH Foundation (Fruit Quality Assessment) grant jointly awarded to Prof. SSR as a lead PI is gratefully acknowledged for database generation.

overview of a basic CNN model. Sustika et al. [6] used a dataset of around 1870 RGB images of strawberries and evaluated the CNN performance of several popular architectures like AlexNet, MobileNet, GoogLeNet, VGGNet, and Xception for quality inspection. They reported VGGNet as the most accurate and GoogLeNet as the most computationally efficient. Hani et al. [7] presented an end-to-end system for image-based fruit detection, counting, and yield estimation techniques. They used U-Net, FRCNN-based object detection, and the Gaussian Mixture Model (GMM) for fruit detection and counting tasks. They found that the GMM-based method performed better than others for fruit detection, while the CNN approach was better at fruit counting and yield estimation. Shahi et al. [8] developed a lightweight attention-convolution based-MobileNetV2 model to classify fruit images and found that their model outperforms four other NN models when trained on three fruit-related datasets. The classification accuracy was greater than 95% for all three datasets, closely followed by MobileNetV2. Other metrics like Precision, Recall, and F1 score also indicated that their model outperformed pretrained DL models like DenseNet121, NASNetMobile, VGG-16, MobileNetV1 and Inception-V3. Gulzar et al. [9] developed a customized version of MobileNetV2 NN called TL-MobileNetV2 by modifying the latter's classification layer. They showed an improved accuracy of 3% more than that of MobileNetV2 when trained and tested on the public dataset of 26149 fruit images belonging to 40 classes. TL-MobileNetV2 also outperformed other pre-trained NNs like AlexNet, VGG16, Inception-V3, and ResNet. Dosi et al. [10] implemented the YOLOv3 model with Darknet architecture to segregate areca nuts. They performed three-band photometry under controlled light conditions to classify based on five different classes.

Most of the existing literature reviewed has used the implementation of CNN models based on image datasets of fruits and vegetables captured under ideal lighting and background conditions, such as high-contrast backgrounds and lighting conditions without glare. Image acquisition in an industrially relevant scenario has not been extensively studied. The current work involves designing and fabricating a low-cost areca nut sorting machine aided by transfer learning-based CNN implementation on single-board computers and affordable camera hardware. In addition to generating a preliminary areca nut dataset, this study explores the implications of some of the hyperparameters used in training CNN-based classification algorithms.

III. Methods

A. Experimental set-up

A modular experimental setup for the areca nut sorting was designed and fabricated using traditional machining and FDM-based (Fused Deposition Modelling) 3D printing technology [11]. The entire system, which was mounted on a metal chassis, consisted of a feeder-scooper setup, a conveyor system, an imaging system, and a sorter, as indicated in Fig. 1. The modular design not only aided in providing the necessary flexibility and customization required in fabricating the individual components but also facilitated the ease of maintenance, upgradation, and repairs

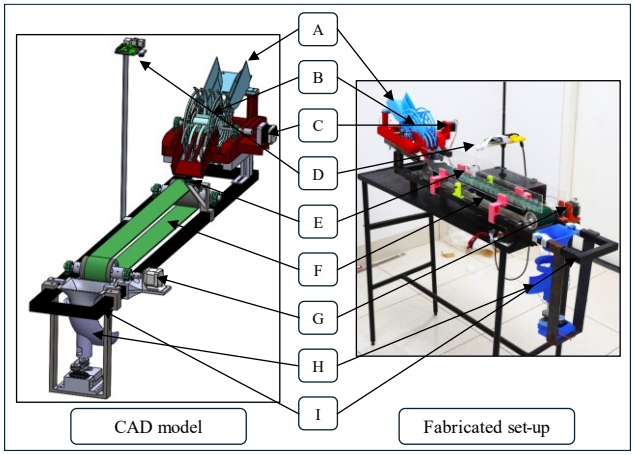

Fig. 1: CAD model and the fabricated sorter machine consisting of following modules: A. Feeder, B. Scooper, C. Stepper motor for scooper, D. Camera Unit, E. IR sensor, F. Conveyor belt, G. Stepper motor for conveyor belt, H. Sorter, I. Metal chassis for mounting all subsystems

without affecting the entire setup. The areca nuts are fed to the feeder-scooper system. The design of the scooper ensures that only one areca nut is dropped onto the conveyor belt at a time using a 5 kg-cm torque stepper motor. The conveyor belt was driven by a 10 kg-cm torque stepper motor controlled by an Arduino-controlled driver. An IR sensor was strategically mounted on the chassis to trigger a pause of the conveyor belt. When the belt is paused for an appropriate time, an 8-bit 704x640 pixel 3-channel (RGB) image is captured by a camera under uncontrolled ambient lighting conditions.

B. Dataset generation

1) Image acquisition

A camera was used to capture the images of the areca nuts against the backdrop of a green conveyor belt. Several black strips were stuck on the belt surface to prevent the areca nut from rolling off and to keep it centred during the belt motion. The motion of the stepper motor was appropriately controlled to achieve a linear belt speed of 20 cm/s. The camera's exposure time, the belt's distance from the camera, and many other parameters were optimized to avoid blurry and out-of-focus images. The areca nut samples of the Mangala variety were sourced from Tumakuru and Hosapete region in Karnataka, India. The camera captured 1615 images of areca nuts belonging to three classes, as indicated in TABLE I. The dataset was manually classified into Green, Green-Yellow, and Orange. Since the prices offered in the market and the methods of dehusking are different for different levels of ripeness, based on domain expert consultations, the following three classes were found to be practically meaningful.

TABLE I: NUMBER OF ARECA NUT SAMPLES USED IN IMAGE DATABASE

Class	Green	Green-Yellow	Orange
Number of areca nuts	517	755	343

2) Image pre-processing

To explore the effect of image cropping (to maximize the RoI in the image) on the accuracy of the classification, the following four versions of datasets were derived using three different cropping methodologies.

A. Original, denoted as D_O
B. Manual cropping, denoted as D_M

C. Cropped using contour detection, denoted as D_1,
D. Cropped using Hough circle and threshold T_1, denoted as D_2
E. Cropped using Hough circle and threshold T_2, denoted as D_3

The following image processing operations were implemented using the Open-source computer vision library OpenCV (version 4.8.1) and ImageJ (version 1.54k).

a) Edge and contour-detection based cropping (D_1)

In this approach, the Canny edge detection function was used to detect the edges of the grey scale image of the areca nut. A three-component Gaussian mixture model [12] was used to identify the thresholds necessary for the edge detection function. Fig. 2 shows the histogram of pixel intensities of the grey-scaled image of an areca nut. The pixel intensities graph exhibits Gaussian distribution with three different peaks, each corresponding to the background (conveyor belt), the areca nut, and the strips. The mean and variance of the Gaussian peak corresponding to areca nut were used to compute the threshold parameters T_1 and T_2 necessary for edge detection using OpenCV [13]. Once the edges were identified, morphological dilation and erosion operations were performed to remove any unintended islands and small edges. The processed image containing edges was fed to a contour detection function to identify the largest contours and store them in descending order. The largest contour, whose barycentre was closest to the image's centroid, was heuristically identified as the areca nut. With sufficient padding of pixels, a bounding box was created to ensure accurate cropping. Fig. 3 illustrates the process of cropping using contour detection.

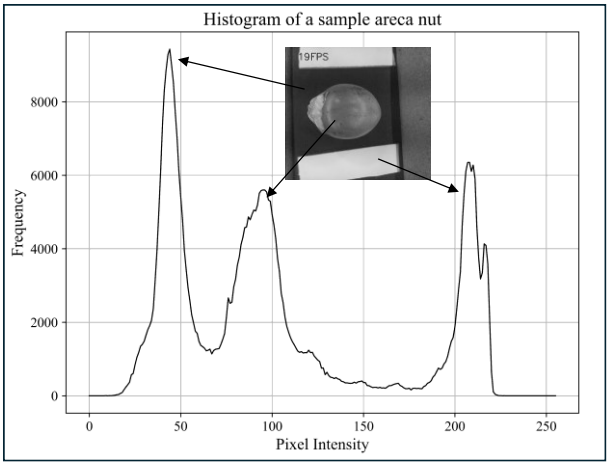

Fig. 2: Histogram of pixel intensity values for a sample image converted to greyscale shows three gaussian peaks

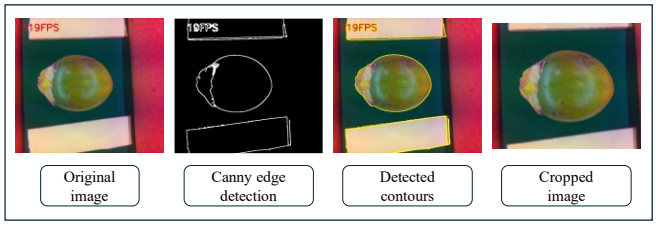

Fig. 3: Cropping using edge and contour detection

b) Hough circle-based cropping (D_2 and D_3)

In this second approach, the Hough Transform was used to identify the areca nut in the image, as illustrated in Fig. 4.

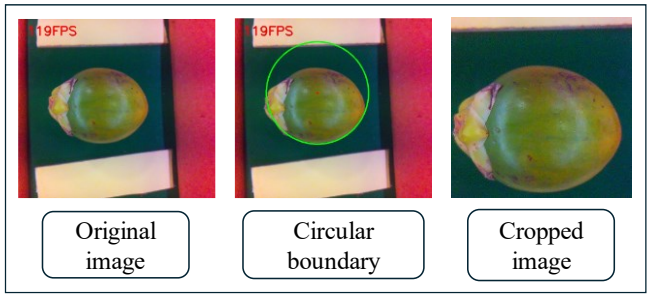

Fig. 4: Image cropping using the Hough Transform

The three-component Gaussian mixture model was used to identify the thresholds necessary for the Hough circle function. The minimum and maximum radii of the circle and its appropriate centre position were arrived at by trial-and-error method on a few randomly chosen images.

If the function could not detect any circle corresponding to the areca nut in the image, no cropping was performed, and the uncropped image was retained. Two thresholds, a lower and a higher (T_1 and T_2) were identified from the GMM for each areca nut image. Accordingly, two cropped datasets, D_2 and D_3, were generated to correspond to the respective thresholds.

C. *Transfer learning*

CNNs are a type of supervised deep neural networks that are used to identify and classify images by analysing features. They are typically fed a set of labelled training images, using which they learn to map the images to the correct labels using feature extraction. The feature extraction is carried out by several convolution and pooling layers, aided by the appropriate choice of activation functions. The convolution layers help identify edges and textures from the images, which are typically followed by the pooling layers, which reduce the dimensionality of the images by condensing the convolved output.

The number of neurons in the convolution and max pooling layers, the number of hidden layers, the type of activation function to choose along with the optimizers and several other decisions are to be carefully and methodically arrived at while constructing a deep CNN to perform the multiclass classification. Transfer learning is usually utilized to perform classification as a workaround to this solution. Transfer learning is the process of utilizing the learnings gained from a well-trained existing deep CNN architecture trained on a similar dataset. This approach uses the knowledge from a task with abundant labelled training data to improve performance on a new task with limited data. This prevents building large CNN models from scratch and instead builds on patterns identified from a similar task. Four well-known CNN architectures were considered in the current study, viz., Inception V3 [14], Resnet50 [15], InceptionResNetV2 [16] and VGG16 [17].

Inception V3 consists of more than 24 million parameters and was trained on the ImageNet dataset [18]. It has 42 layers that are factorised into smaller convolutions and translate into an efficient reduction in grid size. Fig. 5 shows the use of asymmetric convolutions and an auxiliary classifier, along with other improvements, have made it relatively computationally efficient.

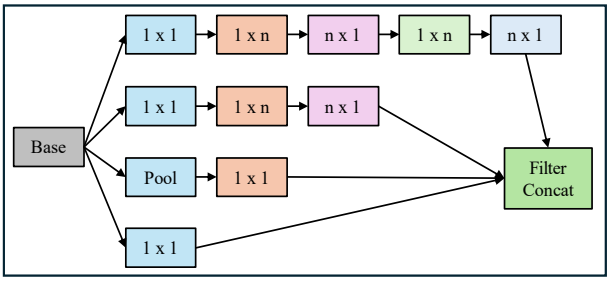

Fig. 5: Schematic diagram representing showing factorisation (n=7) for Inception-V3 architecture [14]

ResNet50 is a mid-sized variant of several deep CNN models belonging to the ResNet (Residual Networks) family. Consisting of 50 layers, it tackles the problem of vanishing gradients by using skip connections to learn the residuals instead of direct mappings. It consists of more than 25 million parameters and was trained on the ImageNet database. Fig. 6 shows the schematic representation of ResNet50.

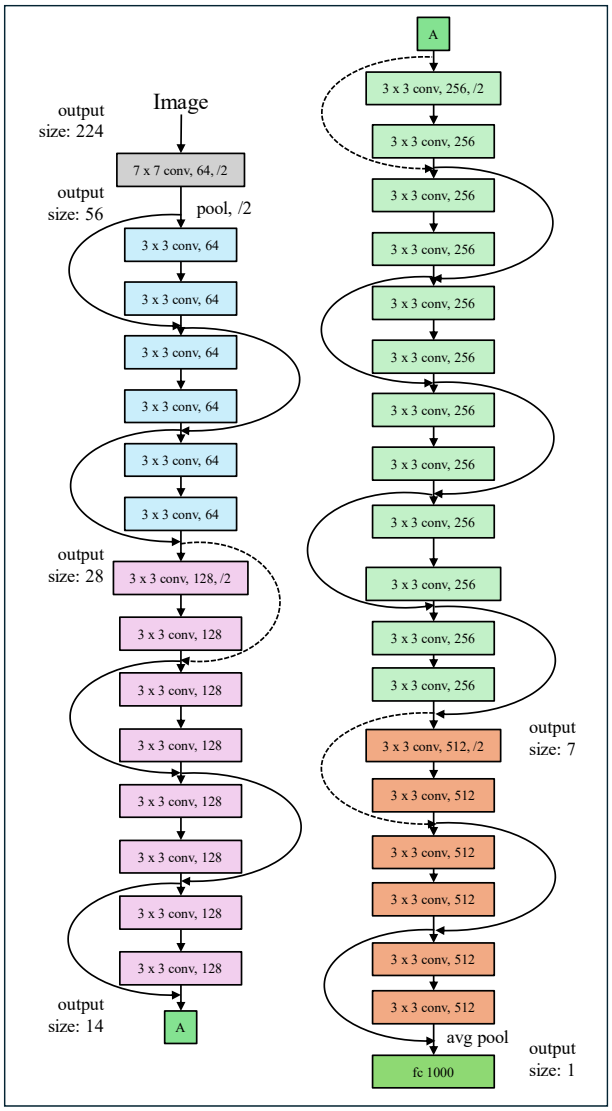

Fig. 6: Schematic diagram representing Resnet50 architecture [15]

InceptionResNetV2 combines the advantages of the original inception architecture with residual connections.

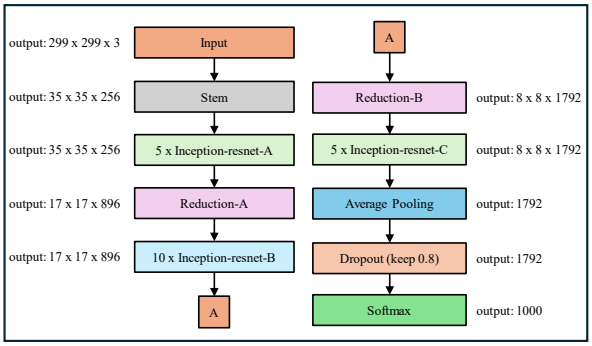

Fig. 7: Schematic diagram representing Inception-ResnetV2 architecture [16]

Trained on the ImageNet database, it has over 55 million parameters. Fig. 7 shows the schematic representation of Inception-ResnetV2.

VGG16 is a CNN proposed by Visual Geometry Group (VGG). Well-known for its simplicity, it consists of 16 layers trained on the ImageNet database, as schematically represented by Fig. 8.

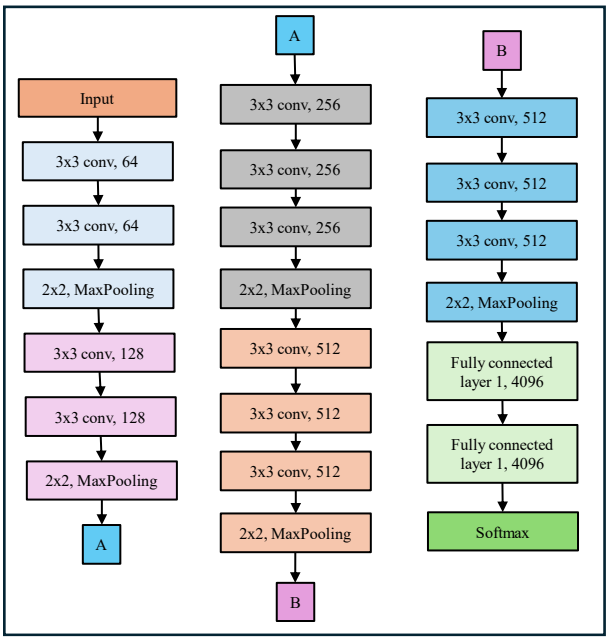

Fig. 8: Schematic diagram representing VGG16 architecture [17]

The fully connected layers of all the pretrained models were removed and the convolutional layers were frozen to retain the feature extraction capabilities gained from the ImageNet dataset. A new fully connected dense layer is added to enable the multiclass classification. Softmax activation function and categorical cross-entropy loss functions are used. Rescaling, shearing, flipping and other data augmentation techniques are implemented. A batch size of 32 is used for training with 500 epochs.

D. Performance metrics

The current work aims to study the performance of different pre-trained CNNs in classifying the areca nuts into three different classes. Several metrics are derived from a confusion matrix.

$$F1\ score = 2 * TP / (2*TP + FP + FN)$$

where, TP: Number of True Positives, FP: Number of False positives, FN: Number of False Negatives.
Since the current problem is that of a multi-class classification, the above metrics are computed on a micro level, i.e., averaged across all test samples. The TP, FP and FN values for each class are summed up and then used to compute the performance metric.

E. Parameters of the current study

The presented work also explores the effect of learning rate and data set split on the performance metrics of different CNNs. Learning rates of 0.001 and 0.0001 are considered along with a 60-20-20 and 80-10-10 split in the dataset. Three variations of the original dataset are generated by cropping the images. A canny edge detector followed by contour detection is used for cropping in one case, and Hough-circle-based cropping is used in the other two cases. The current work aims to study the performance of different pre-trained CNNs in classifying the areca nuts into three different classes. A given CNN is trained on five different datasets, each with two cases of learning rates and two dataset splits. All the images in the dataset were resized to 224x224 pixels and fed to the image data generator module.

1) Parameters summary

A. Learning rates: 2 (0.001, 0.0001)
B. Dataset split: 2 (60-20-20, 80-10-10)
C. Datasets: 4
D. CNN architectures: 4 (Inception V3, Resnet50, InceptionResNetV2 and VGG16)
E. Batch size: 32
F. Number of epochs: 500
G. Optimizer: Adam
H. Loss: categorical cross entropy

IV. RESULTS AND DISCUSSIONS

Fig. 9 shows a schematic representation of all variations of datasets, different deep CNN models chosen, and the parameters selected in the current study, i.e., learning rate and dataset split. It is worth noting that 20 models are trained for each case of a given learning rate and dataset split. For two variations, each for the learning rate and the dataset split, 80 models were trained.

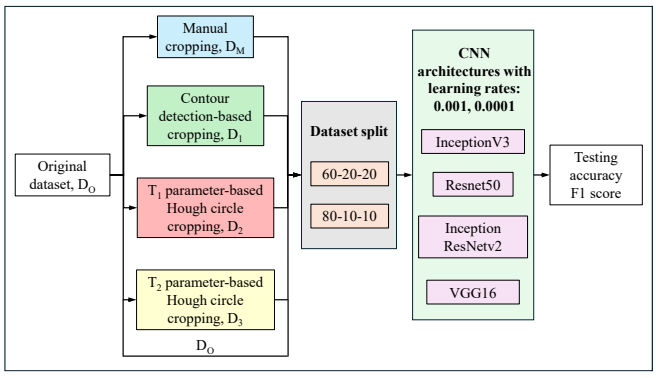

Fig. 9: Schematic representation of explorations in this study - image preprocessing, training of CNN model (total 20) followed by a performance assessment

The results like training time, training and testing accuracies, and performance metrics like F1 score have been collated. Since it is not feasible to showcase all results for all the cases, a few of them have been presented in the current paper.

A. Model training time

It was found that the training time for Inception-V3 was the least across all datasets, dataset splits and learning rates and that of VGG16 was the highest. As expected, it was also found that the training time increased with a decrease in learning rate across all CNN models, datasets, and dataset splits. Cropping using different approaches reduced the training time, albeit marginally, except for the VGG16 model where the time increased with cropping. TABLE II shows the time taken by different models to complete 500 epochs of training for different datasets for the case of 60-20-20 dataset split and a learning rate of 0.0001.

TABLE II: TRAINING TIMES (S) FOR 60 - 20 - 20 SPLIT AND LEARNING RATE 0.0001

NN model / Dataset	D_O	D_M	D_1	D_2	D_3
Inception	513.3	416.3	458.2	444.1	436.3
Resnet50	725.5	688.7	706.5	701.2	696.8
InceptionResNetV2	1017.2	1000.2	1018.5	1018.1	1007.4
VGG16	1729.6	1659.5	1711.7	1715.7	1706.9

B. Model accuracies

For a given dataset split, the training accuracy increases with a decrease in the learning rate for all datasets and CNN models. For a given learning rate, the greater the number of samples in the training dataset, the greater the training accuracy. It is evident from Fig. 10 that the accuracy of the InceptionResNetV2 model is almost equal to that of the Inception-V3 model, leading the VGG16 model by a small value. However, the Resnet50 model exhibited poor training accuracy across all cases of learning rates and splits.

TABLE III shows the testing accuracies for the case of 80-10-10 split and a learning rate of 0.0001. Even though InceptionResNetV2 had a slightly better training accuracy, Inception-V3 has the best testing accuracy among all the models and for different datasets.

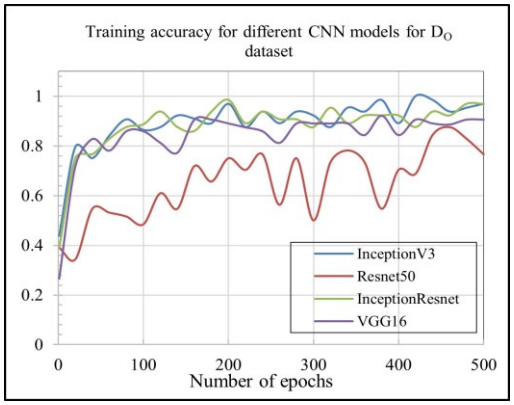

Fig. 10: Training accuracy for D_O dataset for different CNN models

TABLE III: TABLE 3: TESTING ACCURACY FOR 80 - 10 - 10 DATASET SPLIT AND LEARNING RATE 0.0001

NN model / Dataset	D_O	D_M	D_1	D_2	D_3
Inception	0.9198	0.8447	0.8571	0.9068	0.8634
Resnet50	0.6852	0.6522	0.6584	0.7391	0.6398
InceptionResNetV2	0.9074	0.8758	0.8634	0.8696	0.8199
VGG16	0.8827	0.8199	0.8447	0.8820	0.8820

C. Performance metrics

Confusion matrices are generated for each case, and the micro-averaged performance metric F1 score is evaluated. Fig. 11 shows the confusion matrix for the Inception-V3 model trained on D_O dataset, with a learning rate of 0.0001 and a dataset split of 80-10-10 as an example.

Since it is a multiclass classification, the current work employs micro averaging to compute F1 score. The values of TP, FP, FN (as shown being calculated for the confusion matrix in Fig. 11), are summed up for all the classes as shown in TABLE IV. The value of F1 score computed using these summed values is called micro averaged performance metric.

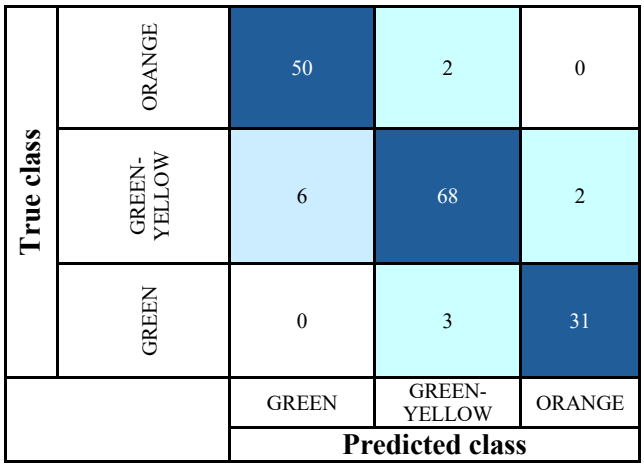

Fig. 11: Confusion matrix for InceptionV3 model for 80-10-10 and learning rate 0.0001

Fig. 12 shows the F1 score calculated for the case of an 80-10-10 split and a learning rate of 0.0001. It was found that the Inception-V3 model performed the best across all the combinations of datasets, dataset splits, and learning rates. Resnet50 model fared poorly across all datasets, with InceptionResNetV2 being better than VGG16 and slightly underperforming compared to Inception-V3.

TABLE IV: CLASSIFICATION OUTCOME - CONFUSION MATRIX

Label	TP	FP	FN
GREEN	50	6	2
GREEN-YELLOW	68	5	8
ORANGE	31	2	3

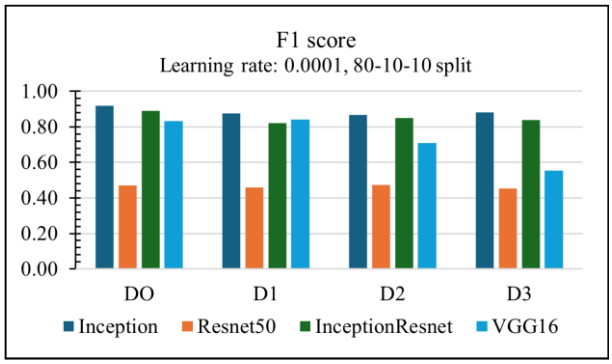

Fig. 12: Effect of manual and automated cropping techniques on F1 score (80-10-10 split and 0.0001 learning rate)

TABLE V shows F1 scores for different learning rates and dataset splits employed in the current study. The Inception-V3 models exhibit the best F1 score across all datasets and parameter cases. Resnet50 is at the other extreme, showing the lowest score for all the cases. It is observed that the F1 score increases with a decrease in learning rate and an increase in the number of samples available for training.

TABLE V: F1 SCORE FOR DIFFERENT CNN MODELS FOR DIFFERENT LEARNING RATES AND DATASET SPLITS

NN Model / Dataset	D_O	D_M	D_1	D_2	D_3
Dataset split: 80-10-10, Learning rate: 0.0001					
Inception	0.92	0.826	0.876	0.867	0.882
Resnet50	0.469	0.484	0.46	0.472	0.453
InceptionResNetV2	0.889	0.863	0.82	0.851	0.839
VGG16	0.833	0.516	0.842	0.708	0.553
Dataset split: 80-10-10, Learning rate: 0.001					
Inception	0.907	0.857	0.839	0.882	0.801
Resnet50	0.451	0.472	0.441	0.472	0.422
InceptionResNetV2	0.87	0.826	0.776	0.845	0.826
VGG16	0.784	0.472	0.689	0.528	0.516
Dataset split: 60-20-20, Learning rate: 0.0001					
Inception	0.877	0.854	0.868	0.831	0.849
Resnet50	0.465	0.466	0.452	0.462	0.431
InceptionResNetV2	0.883	0.851	0.88	0.806	0.837
VGG16	0.825	0.522	0.815	0.671	0.578
Dataset split: 60-20-20, Learning rate: 0.001					
Inception	0.782	0.854	0.855	0.871	0.772
Resnet50	0.465	0.466	0.422	0.465	0.44
InceptionResNetV2	0.889	0.851	0.849	0.834	0.874
VGG16	0.766	0.522	0.658	0.523	0.566

D. Effect of cropping

Cropping using the two approaches mentioned in the image processing section (III.B.2) had a positive impact on the training time. However, it was also found that the process of cropping worsened the training and testing accuracies along with the performance metrics. Global automated cropping for images of all classes and datasets based on contour detection and Hough circles was performed. Image capture in ambient lighting conditions (without controlled lighting conditions) in tandem with inconsistent position of areca nut in the image and lack of contrast between the background and areca nut (object of interest) made it difficult to select the same cropping parameters for over 1600 images.

Fig. 13 shows a couple examples of different cropping methodologies employed in the current study. It is seen each method performs the cropping process with sufficient fidelity. Fig. 14 shows a few sample cases of incorrect cropping that were encountered during image pre-processing, highlighting the challenge posed by the patches in the background in the traditional RoI detection. As seen in TABLE V, the model VGG16 appears to be highly sensitive to the different cropping methodologies.

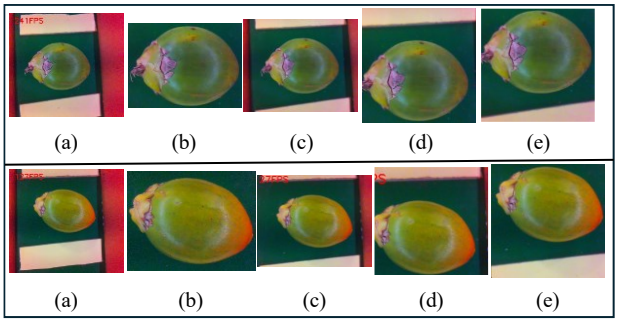

Fig. 13: A couple examples of different cropping methodologies used; (a) Original image, (b) Manual cropping, (c) Cropping using Canny edge detection, (d) GMM-based cropping with threshold T_1, (e) GMM-based cropping with threshold T_2

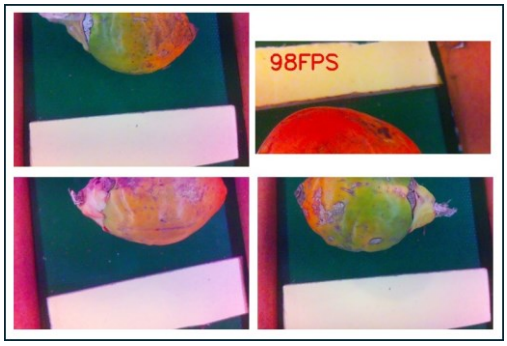

Fig. 14: A few sample cases of incorrect cropping

Taking manual cropping as a reference without any error, we can conclude that performance decreases with a cropped image compared to the original image with extra background information. This might have originated from data loss during file conversion and compression processes or reduction in neighbouring areas of the RoI. This could also originate from the dissimilarity of the RoI relative to the total image area between the cropped dataset and the dataset (e.g., ImageNet) used to train all pre-trained models.

V. CONCLUSIONS

Machine learning and computer vision-based agricultural practices have been instrumental in reducing bias, subjectivity, and unintended human errors in classification and segregation along with the quality assessment process. Implementation of machine learning-based multiclass classification under industry-ready conditions (far away from the cushion of controlled lighting and contrast rich background conditions) has been attempted in the current study. A large areca nut dataset comprising over 1600 images spread across three classes has been created. Transfer learning-based CNN models have been used to perform classification on five different versions of datasets for two different cases each of learning rates and dataset splits. The Inception-V3 model was found to perform the best among the chosen CNNs with a testing accuracy of over 91% for the case of 80-10-10 dataset split and a learning rate of 0.0001. F1 score was computed for all combinations of parameters and datasets and Inception-V3 was found to exhibit a highest score of 0.92, followed by InceptionResNetV2, VGG16 and Resnet50 models with scores of 0.889, 0.833 and 0.469 respectively for the case of 80-10-10 dataset split and learning rate 0.0001. Image cropping for data with complex background adversely affects the CNN models' accuracy and performance metrics, possibly owing to file conversions or data reduction during the cropping process.

ACKNOWLEDGMENT

We would like to thank Prof. Rajshekhar Bhat for his valuable time, insights and discussions regarding various aspects of dataset preparation and machine learning. We thank Mr. Vignesh PS and Mr. Chethan Y R for sharing their help and expertise for creating the database. We acknowledge Mr. Mohd Shafi Kumar and Mr. Shrikant Badi for their assistance in manual image pre-processing.

REFERENCES

[1] Rai, R., Tiwari, M. K., Ivanov, D., & Dolgui, A., Machine learning in manufacturing and industry 4.0 applications., International Journal of Production Research, 59 (16), 2021, pp. 4773-4778.

[2] Panch, T., Szolovits, P., & Atun, R., Artificial intelligence, machine learning and health systems., Journal of global health, 2018, 8 (2).

[3] Deo, R. C., Machine learning in medicine. Circulation, 132 (20), 2015, 1920-1930.

[4] Rajkomar, A., Dean, J., & Kohane, I., Machine learning in medicine., New England Journal of Medicine, 380 (14), 2019, 1347-1358.

[5] Naranjo-Torres, J., Mora, M., Hernández-García, R., Barrientos, R. J., Fredes, C., & Valenzuela, A., A review of convolutional neural network applied to fruit image processing., Applied Sciences, 10 (10), 2020, 3443.

[6] Sustika, R., Subekti, A., Pardede, H. F., Suryawati, E., Mahendra, O., & Yuwana, S., Evaluation of deep convolutional neural network architectures for strawberry quality inspection., Int. J. Eng. Technol, 7 (4), pp. 75-80.

[7] Hani, N., Roy, P., & Isler, V., A comparative study of fruit detection and counting methods for yield mapping in apple orchards., Journal of Field Robotics, 37 (2), 2018, pp. 263-282.

[8] Shahi, T. B., Sitaula, C., Neupane, A., & Guo, W., Fruit classification using attention-based MobileNetV2 for industrial applications., Plos one, 17 (2), 2022, e0264586.

[9] Gulzar, Y., Fruit image classification model based on MobileNetV2 with deep transfer learning technique., Sustainability, 15 (3), 2023, 1906.

[10] Dosi, S., Vamsi, B., Raut, S. S., & Narasimha, D., Segregation of Areca Nuts Using Three Band Photometry and Deep Neural Network, International Conference on Soft Computing and its Engineering Applications, 2021, pp. 15-27.

[11] Aishwary, Singh, Modularized Design of Computer Vision Based Fruit Segregation System, MS Thesis, Dept. of MMAE, IIT Dharwad, Dharwad, 2024

[12] Reynolds, D. A., Gaussian mixture models., Encyclopedia of biometrics, 741, 2009, pp. 659-663.

[13] Bradski, G., The OpenCV Library, Dr. Dobb's Journal of Software Tools, 2000.

[14] Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., & Wojna, Z., Rethinking the inception architecture for computer vision., Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 2818-2826.

[15] He, K., Zhang, X., Ren, S., & Sun, J., Deep residual learning for image recognition., Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 770-778.

[16] Szegedy, C., Ioffe, S., Vanhoucke, V., & Alemi, A., Inception-v4, inception-resnet and the impact of residual connections on learning., Proceedings of the AAAI conference on artificial intelligence, 31 (1), 2017.

[17] Simonyan, K., & Zisserman, A., Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*., 2014

[18] Deng, J., Dong, W., Socher, R., Li, L. J., Li, K., & Fei-Fei, L., Imagenet: A large-scale hierarchical image database, IEEE conference on computer vision and pattern recognition, 2009, pp. 248-255.

[19] Tammina, S., Transfer learning using vgg-16 with deep convolutional neural network for classifying images, International Journal of Scientific and Research Publications (IJSRP), 9(10), 143-150, 2019