

# Optimizing Dehusked Arecanut Quality Segregation: CNN-Based Approach with Contrast Enhancement and Data Augmentation\*

Sameer Patil<sup>1,†</sup>, Aparajita Naik<sup>2,†</sup>, Marlon Sequeira<sup>3,†</sup>, Sulaxana Vernekar<sup>4,†</sup> and Jivan Parab<sup>5,\*</sup>

<sup>1a</sup>Electronics Programme, School of Physical and Applied Science, Goa University, Goa, India

<sup>1b</sup>Department of Electronics, Dnyanprassarak Mandal's College and Research Centre, Goa, India

<sup>2</sup>Electrical Engineering, Cambridge University, Cambridge, UK

<sup>3</sup>Electronics Programme, School of Physical and Applied Science, Goa University, Goa, India

<sup>4</sup>GVM'S GGPR College of Commerce and Economics, Goa, India

<sup>5</sup>Research Supervisor, Electronics Programme, School of Physical and Applied Science, Goa University, Goa, India

## Abstract

In the production process of Areca nut, the segregation stage is of prime importance. As of now, most commercial retailers use skilled workers for quality segregation, which means a lot of time is required for finalizing the product costing. Based on the inputs received from marketing executives, it was observed that if any method for automatic segmentation has to be meaningful, then the quality segregation should not have errors of more than 5% standard deviation. In this study, we propose a methodology based on 10-fold cross-validation training of Convolutional Neural Network (CNN) using contrast enhancement and no data augmentation of the images. Also, in this paper, we compare the results attained on the quality segregation using numerous processing methods, for instance, data augmentation for images with and without cropping and also for images with and without contrast enhancement. The database developed here uses Areca nut cultivated in the Western Ghats region of the Indian Peninsula, particularly focused on the Konkan belt. In our paper, we achieved the lowest standard deviation of 4.1% for cropped images with contrast enhancement.

## Keywords

Areca nut, Segregation, Convolutional Neural Networks (CNN), Contrast Limited Adaptive Histogram Equalization (CLAHE)

## 1. Introduction

Areca palm (*Areca catechu L.*) is grown for its kernel, popularly known as Areca nut (or Betel nut or Supari) in India. It is grown commercially along the western coast of India (Maharashtra, Goa, Karnataka), Tamil Nadu, Kerala, Assam and West Bengal [1]. It is a tropical crop grown commercially in Southeast and South Asian countries, parts of East Africa, and some tropical Pacific nations [2,3]. Since ancient times, it has been used as a masticator wherein a slice of Areca nut, along with slaked lime (aqueous Calcium hydroxide paste) and some aromatics (clove, cardamom, saffron) for extra flavor, is wrapped in a betel leaf [4]. In ancient Indian scripts, it is mentioned that Areca nut can be utilized as a therapeutic agent for leucoderma, leprosy, anemia, and obesity and has de-worming properties due to which it is used in gastrointestinal disorders [5]. Areca nut has many pharmacological properties and is extensively used in medicine. It has anti-allergic, anti-parasitic, anti-microbial, and anti-aging properties. Areca nuts are also used to produce adhesives, non-woven

---

SCCTT-2024: International Symposium on Smart Cities, Challenges, Technologies and Trends, 29th Nov 2024, Delhi, India

\* Corresponding author.

† These authors contributed equally.

✉ sameer@dmscollege.ac.in (S. Patil); an656@cam.ac.uk (A. Naik); marlon@unigoa.ac.in (M. Sequeira); sulaxgoa@gmail.com (S. Vernekar); jsparab@unigoa.ac.in (J. Parab)

ORCID 0009-0007-3674-1942 (S. Patil); 0009-0002-0332-1654 (A. Naik); 0000-0002-1444-5428 (M. Sequeira); 0000-0002-0848-7349 (S. Vernekar); 0000-0002-7462-3492 (J. Parab)

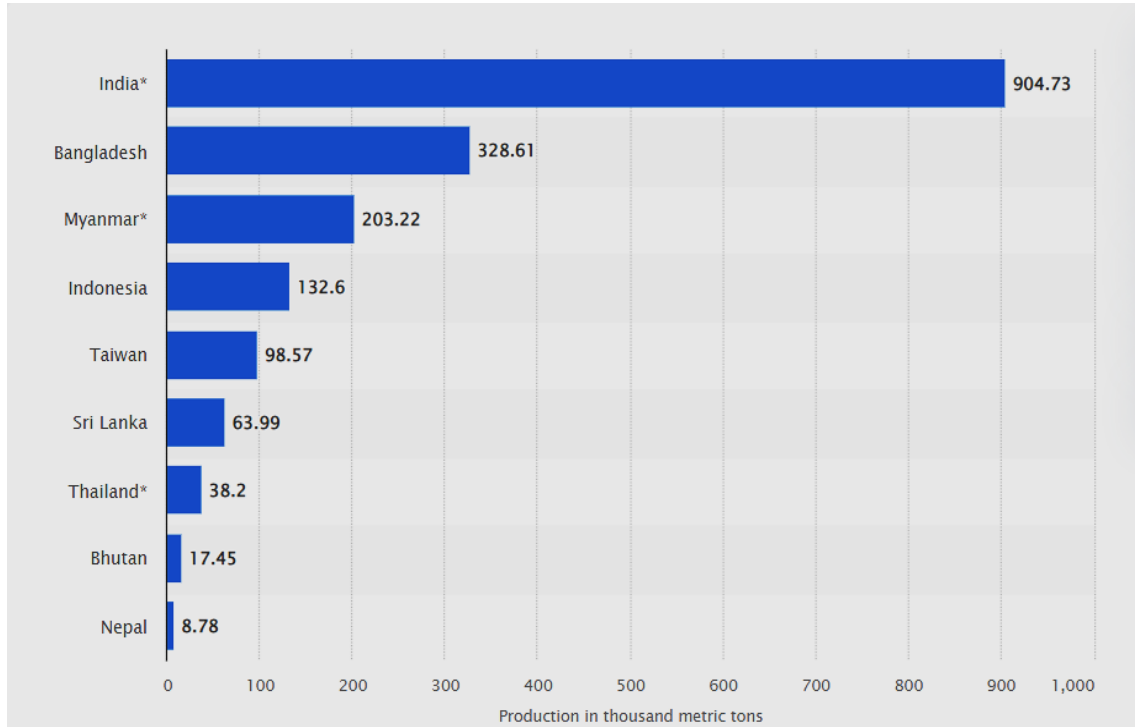


© 2023 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

fabrics, textile dyes and building materials. Hence, due to its high economic significance, Areca nut has become an important cash crop.

As per the latest studies, India tops at the global level, contributing to approximately 904 thousand metric tons in 2020[6]. The top ten Areca nut producing countries over the globe are shown in Figure 1.1.

**Fig 1.1:** Areca nut production in Asia Pacific in 2020 by country (in 1000 metric tons) [6]



The Areca nut kernel is hard from outside with the inner endosperm marbled in dark brown and white [7]. The crucial steps in the areca nut production process are listed below.

Harvesting

1. Drying
2. De-husking
3. Nut segregation based on its quality.

Nut segregation is the most labor-intensive and time-consuming of the production process's aforementioned steps. In Goa (India), Goa Bagayatdar, a cooperative organization, is a leading Areca nut collector. At their collection centres, nuts are classified on the basis of texture, colour and the quality. Here, nuts are segregated in seven different categories (Supari, Safed, Laal, Vench, Kharad, Tukda and Baad) [8]. But, due to the shortage of skilled laborers for the above said work, it is essential to develop a unit of segregation based on its quality. This will not only solve the issue of scarcity of laborers but also will save farmer's time.

## 2. Literature Review

Much work is being done in machine learning and image processing to identify, categorize, and grade agricultural products. S. Siddesha et al., in their study of the texture-based classification of Areca nut, extracted different texture features using Wavelet, Gabor, Gray Level Difference Matrix, Local Binary Pattern (LBP), and Gray Level Co-Occurrence Matrix features. The Nearest Neighbor classifier was used to classify Areca nuts. A classification rate of 91.43% is achieved with Gabor wavelet features [9]. Mallaiiah Suresha et al. have proposed diseased and undiseased classification of Areca nut using texture features of LBP, Haar Wavelets, GLCM, and Gabor. They achieved a 92.00% success rate [10]. T. Liu et al. have tried to achieve automatic classification by extracting the color,

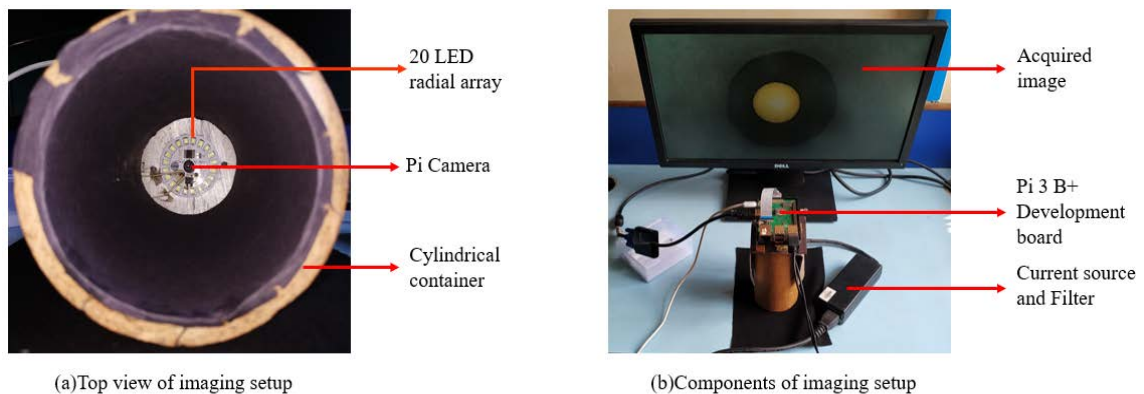
shape, and texture features of de-husked Areca nut [11]. Huang K.Y. used Image processing techniques and Neural Networks for quality detection and classification of areca nuts. Six geometric features, 3 color features, and defects were used for the classification process. This method of classification attained an accuracy of 90.9% [12].

Deep Learning (DL) approaches are increasingly important in machine learning because of their high degrees of abstraction and capacity to automatically identify image patterns [13]. Convolutional Neural Network (CNN) is the most frequently applied deep learning architecture for image processing among the numerous designs employed [14,15,16]. Convolution operations are used by CNN, a kind of Artificial Neural Network (ANN), in a minimum of one of its layers [14].

To the best of our information, very little research has been done on the classification of dehusked Areca nuts using CNN.

### 3. Data Acquisition Setup

This paper deals with the quality classification of Areca nut from the Konkan belt of India, particularly from the state of Goa. Since there are no publicly available database of the Areca nut images, A unique setup was created to create an initial database. The setup consists of a top-mounted camera with a sample table below at a distance of approximately 14cm. Surrounding the camera are radially arranged 20 white LEDs evenly illuminating the sample. A hollow cylinder coated with black paper on its inner sides is placed around the sample and camera to shield the stray light entering the acquisition setup. The black paper prohibits light reflection from the inner walls and creates a glare on the camera lens. The power source for the setup is an AC source of 220V, 50 Hz, which is then converted to a DC constant current source coupled with a high voltage capacitor of 220  $\mu$ F/ 450 V connected in parallel to reduce flicker in the illumination. In this setup, we have used a 5MP lightweight Pi camera module, which communicates with the Raspberry Pi 3 B+ board using the MIPI camera serial interface protocol. At the base of the hollow cylinder, a black cloth is placed over which an Areca nut whose image is to be acquired is kept for the reasons described above. Figure 2.1, shows the data acquisition setup designed for capturing images of Areca nuts.



**Figure 2.1:** Data acquisition setup

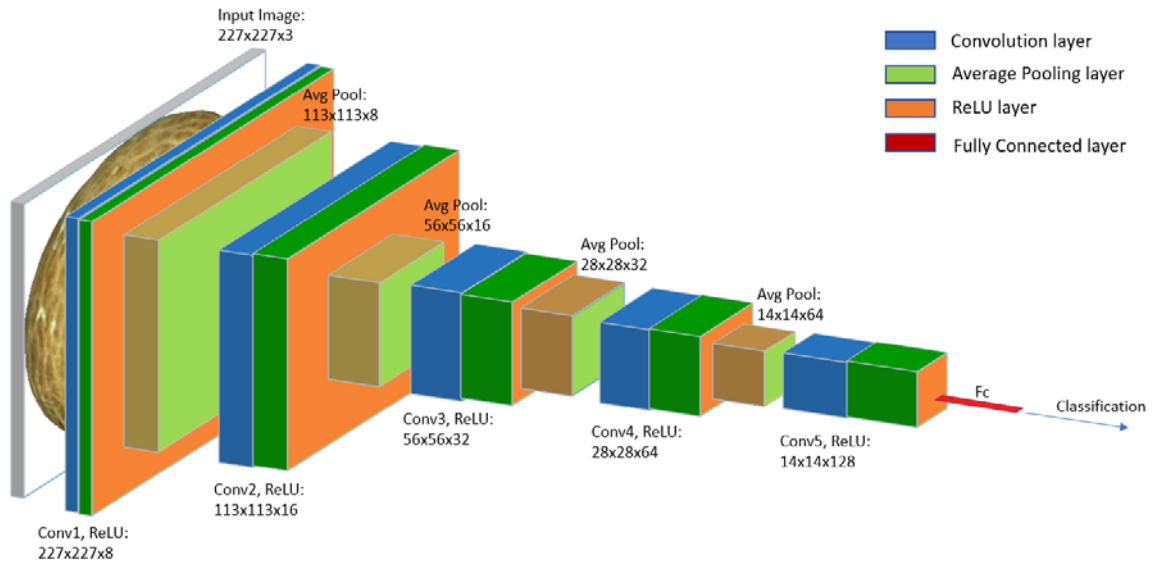
©Journal of Applied and Natural Science 15(4) ----2023

## 4. Methodology

### 4.1. Convolutional Neural Network

Convolutional Neural Networks (CNN) have been extensively studied in recent literature [17,18]. CNN is a class of deep learning algorithms that is incredibly efficient in classifying data by recognizing patterns in an image. A CNN is a feed-forward network consisting of basic building blocks like a convolutional layer, pooling layer, and activation layer, which are stacked with varying permutations and combinations. This varying arrangement of convolutional layer, pooling layer,

and activation layer together form the feature extraction segment of a CNN [19]. Within the classification segment, the extracted features are fed into the fully connected layer and the classification layer [20]. The details of the various layers used in our custom CNN model are detailed in Figure 3.1.



**Figure 3.1:** The custom CNN model

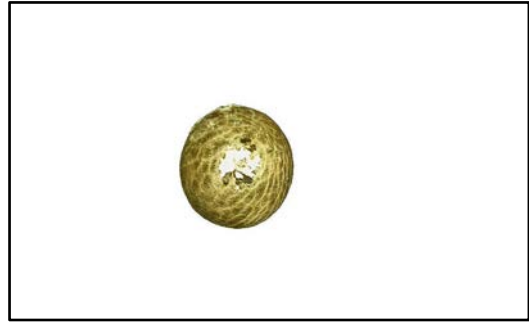
It should be emphasized that all of the photos used in this study show the Areca nut from the top. This is because, the very shape of Areca nut, which normally stabilises with its flat surface at the bottom. Also, we wanted to study the accuracy of segregation based on the top view to design an algorithm that will take reduced time for classification and thus increase the speed of segregation. In this experiment, we have performed different image processing operations as detailed below to get a better understanding of which operations will yield the best outcomes with the CNN network.

1. Areca Nut image has only been segmented and not cropped to a Region of Interest (ROI) closest to its edges. This database is labeled as NoCrop\_NoContrast.
2. The Areca Nut image has been cropped to ROI closest to its edge. This database is labelled as Crop\_NoContrast.
3. Areca Nut image has only been segmented and not cropped to a ROI closest to its edges and has been contrast-enhanced using Contrast -limited Adaptive Histogram Equalization (CLAHE). This database is labelled as NoCrop\_Contrast.
4. Areca Nut image has been cropped to ROI closest to its edge and has been contrast-enhanced using CLAHE. This database is labelled as Crop\_Contrast.

Thus, in this experiment, we are working with four distinct databases. The images of each database have been illustrated in Figure 4.1 below.



(a) Uncropped and Segmented Areca nut image.



(b) Uncropped and Segmented Areca nut image with contrast enhancement using CLAHE



(c) Cropped Areca Nut image.



(d) Cropped Areca Nut image with contrast enhancement using CLAHE.

**Figure 4.1:** Images of databases of Areca nut used for the classification.

## 4.2 Contrast-limited Adaptive Histogram Equalization (CLAHE)

CLAHE is an algorithm used to enhance the contrast between unprocessed images. It performs histogram equalizations on non-overlapping sections of a given image and is called tiles. The surrounding tiles are then blended using bilinear interpolation to prevent introducing false borders [21].

We have also tried to find the outcome of data augmentation with each database on the final classification accuracy and the standard deviation. Therefore, with each database, we aimed to determine the classification accuracy and standard deviation with CNN, using data augmentation (with and without data augmentation).

## 4.3 Data Augmentation

Data augmentation is a technique in CNN, and is normally applicable, when the training samples are limited. Thus, we can produce more training examples for a network by leveraging existing images. This is accomplished by applying image processing techniques such as scaling, rotation about an axis, translation, and reflection about an axis. This results in a significantly bigger training sample size from the existing data [22].

To evaluate our CNN, we use 10-fold cross-validation. In 10-fold cross-validation, the database is split into 10 distinct folds, of which 9 folds will be used in training, and the 10th fold will be used for

testing. This means that each sample used for testing is now comprised of one in the training set, and one from the training set is used for testing. Thus, the procedure is repeated 10 times, with every iteration having a new fold from one of the 10 folds for testing [23].

## 5. Results and Analysis

The current section presents the classification accuracy for all four databases. Specifically, in the current section, we train a CNN with 10-fold cross-validation with and without data augmentation for each database.

**Table 5.1:** Classification accuracy and standard deviation for uncropped images with no contrast enhancement.

	Model	Trial 1	Trial2	Trial3	Trial4	Trial5	Trial6	Trial7	Trial8	Trial9	Trial10	Average	Std. Deviation
No Augmentation	CNN	92.86%	80.95%	92.86%	80.95%	85.71%	80.95%	80.95%	78.57%	80.95%	83.33%	83.81%	5.12%
Augmentation	CNN	71.43%	64.29%	57.14%	61.90%	71.43%	73.81%	54.76%	52.38%	66.67%	61.90%	63.57%	7.36%

Table 5.1 gives the result of uncropped images with no contrast enhancement. It is observed here that no augmentation gives better results as compared to augmentation with 5.12% of standard deviation.

**Table 5.2:** Classification accuracy and standard deviation for cropped images with no contrast enhancement.

	Model	Trial 1	Trial2	Trial3	Trial4	Trial5	Trial6	Trial7	Trial8	Trial9	Trial10	Average	Std. Deviation
No Augmentation	CNN	83.33%	66.67%	80.95%	80.95%	78.57%	78.57%	80.95%	69.05%	83.33%	83.33%	78.57%	5.94%
Augmentation	CNN	64.29%	73.81%	66.67%	71.43%	73.81%	66.67%	83.33%	71.43%	76.19%	69.05%	71.67%	5.55%

Table 5.2 gives the result of cropped images with no contrast enhancement. The standard deviation here for no augmentation has increased compared to the augmentation for uncropped images, and the results are the opposite of those given in Table 5.1. However, both methods have a standard deviation higher than 5%.

**Table 5.3:** Classification accuracy and standard deviation for uncropped images with contrast enhancement.

	Model	Trial 1	Trial2	Trial3	Trial4	Trial5	Trial6	Trial7	Trial8	Trial9	Trial10	Average	Std. Deviation
No Augmentation	CNN	92.86%	83.33%	78.57%	78.57%	83.33%	80.95%	90.48%	78.57%	76.19%	80.95%	82.38%	5.41%
Augmentation	CNN	88.10%	61.90%	57.14%	73.81%	66.67%	61.90%	42.86%	73.81%	71.43%	52.38%	65.00%	12.75%

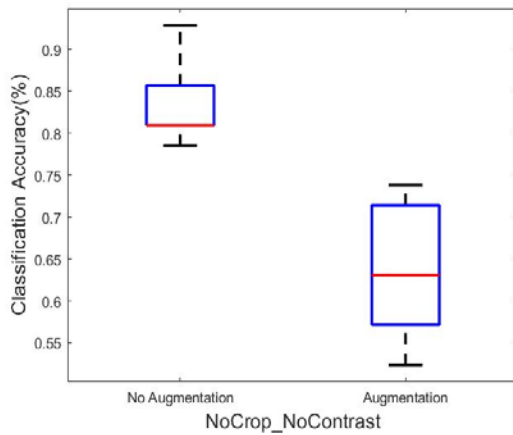
Table 5.3 shows the results of uncropped images with contrast enhancement. The results indicate that these methods do not improve significantly over the earlier two methods, whose results are listed in Table 5.1 and Table 5.2. The augmentation process gives the worst result, with a standard deviation of more than 10%.



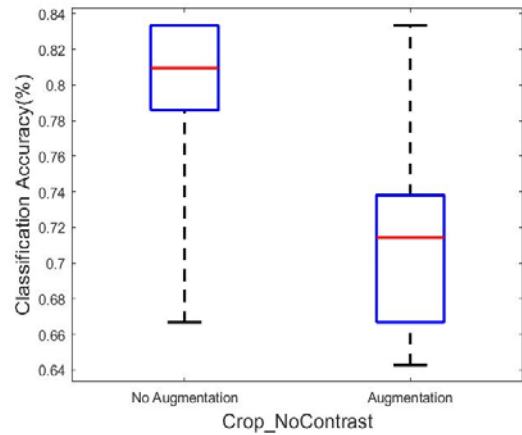
**Table 5.4:** Classification accuracy and standard deviation for cropped images with contrast enhancement.

	Model	Trial 1	Trial2	Trial3	Trial4	Trial5	Trial6	Trial7	Trial8	Trial9	Trial10	Average	Std. Deviation
No Augmentation	CNN	76.19%	80.95%	73.81%	83.33%	83.33%	73.81%	80.95%	83.33%	73.81%	80.95%	<b>79.05%</b>	<b>4.17%</b>
Augmentation	CNN	71.43%	73.81%	52.38%	59.52%	71.43%	59.52%	69.05%	66.67%	73.81%	64.29%	66.19%	7.17%

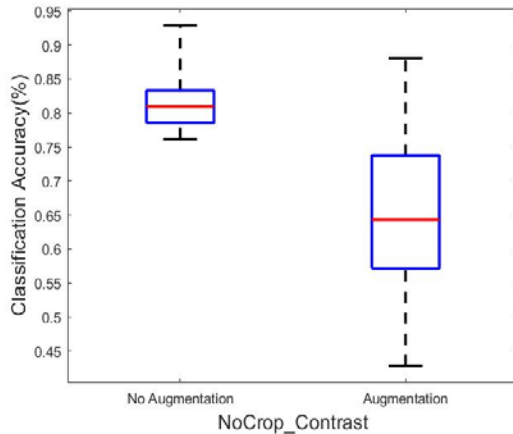
The Table 5.4 gives the result of cropped images with contrast enhancement. Here, it may be seen that no augmentation with our custom CNN model gives a standard deviation close to 4%, which defends our claim that the top view can alone be used for the segregation process. It may be noted that the cropped image with no augmentation worked quite well, but it did not fare so well when the augmentation process was utilized on the samples. All the above results and analysis have been shown in the boxplot in Figure 5.1.



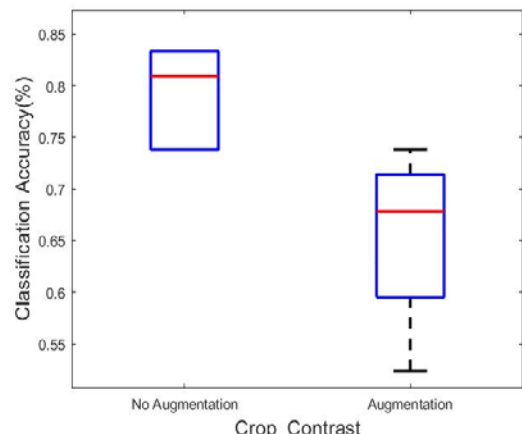
(a) Classification for uncropped images with no contrast enhancement.



(b) Classification for cropped images with no contrast enhancement.



(c) Classification for uncropped images



(d) Classification for cropped images with contrast enhancement.

**Figure 5.1:** Boxplot of the results and analysis of the image classification

Here, Figure 5.1, as above, shows the Boxplot of the results and analysis of the image classification.

From the boxplot (a), it may be seen that, in the case of no augmentation, the accuracy is close to 80% for most trials. Whereas, in augmentation, it widely varies with the least going almost close to 50%, which is not desirable

The boxplot (b) also has accuracy for both models (for cropped and no contrast enhancement) varying widely from 66% to 84%, therefore casting doubt on the process of classification. The same is true in augmented images, wherein the accuracy varies from 64% to 84%.

As discussed in boxplot (b), boxplot (c) (for uncropped and contrast enhancement) also has a similar behavior wherein the accuracy varies widely over 75% to 95% (for no augmentation) and 45% to 90% (for augmentation). Thus, signifying that they are not consistent.

In boxplot (d), the accuracy for non-augmented images is centered around 82% with a small deviation from 74% to 84% for our custom CNN model. Thus, suggesting this method is more reliable for classification of Areca nuts. However, the same is not true in case of augmentation. The accuracy in the case of augmentation varies from 50% to 75%.

## 6. Conclusion

In the above article, we have carried out four diverse classification methods based on 10-fold cross-validation training of a custom CNN model using contrast enhancement and data augmentation of the images. The results indicate that using the custom CNN model, the classification method using no augmentation and contrast enhancement for cropped images, has yielded the best outcomes with a standard deviation of less than 5%. The standard deviation of less than 5% is a significant number for agriculturalists for the segregation of Areca nuts, considering we have used only the top view. Single-image segregation can greatly increase the speed of segregation; thus, the payments to the farmers can be given on the spot, and the loss of revenue due to human fatigue can be reduced. The above experiments suggest that the algorithm can be implemented and a machine can be manufactured to segregate Areca nuts automatically. This article provides concept validation for the manufacturing of automated Areca segregation units.

## Acknowledgement

Authors acknowledge the help extended by the skilled segregators for classification and officials of Goa Bagayatdar for providing many samples of Areca nut used for this work.

## Statements and Declarations

### Data Availability

The data that support the findings of this study are available from the corresponding author, upon reasonable request.

### Funding and/or Conflicts of Interests/Competing Interests

**Funding:** The authors declare that they have not availed any funding from any agency for the above-carried work.

**Conflict of Interest:** The authors declare that they have no conflict of interest.

## References

- [1] Areca Nut - an overview | ScienceDirect Topics. (2024). <https://www.sciencedirect.com/topics/neuroscience/areca-nut>
- [2] Origin | Arecanut. (2024). <https://arecanut.org/arecanut-1/origin/>
- [3] Areca nut, *Wikipedia*. (2024). [https://en.wikipedia.org/w/index.php?title=Areca\\_nut&oldid=1247804910](https://en.wikipedia.org/w/index.php?title=Areca_nut&oldid=1247804910)



- [4] A. Kumar *et al.*, Assessment of areca nut use, practice and dependency among people in Guwahati, Assam: a cross-sectional study, *ecancer*, vol. 15, (2021), doi: 10.3332/ecancer.2021.1198.
- [5] M. S. Amudhan, Begum V Hazeena, and H. K. B. , A REVIEW ON PHYTOCHEMICAL AND PHARMACOLOGICAL POTENTIAL OF ARECA CATECHU L. SEED, *IJPSR*, vol. 3, no. 11, pp. 4151–4157,(2012),  
[https://www.researchgate.net/publication/264710991\\_A\\_review\\_on\\_phytochemical\\_and\\_pharmacological\\_potential\\_of\\_Areca\\_catechu\\_L\\_Seed](https://www.researchgate.net/publication/264710991_A_review_on_phytochemical_and_pharmacological_potential_of_Areca_catechu_L_Seed)
- [6] APAC: areca nut production by country 2022, Statista. (2024).  
<https://www.statista.com/statistics/657902/asia-pacific-areca-nut-production-by-country/>
- [7] V. Raghavan and H. K. Baruah, Arecanut: India’s popular masticatory –history, chemistry and utilization, *Econ Bot*, vol. 12, no. 4, 315–345, Oct. 1958, doi: 10.1007/BF02860022.
- [8] Goa Bagayatdar Bazar – One-stop-shop for all. (2024). <https://goabagayatdar.com/>
- [9] S. Siddesha, S. K. Niranjana, and V. N. Manjunath Aradhya, Texture based classification of arecanut, in *2015 International Conference on Applied and Theoretical Computing and Communication Technology (iCATccT)*, (2015), 688–692. doi: 10.1109/ICATCCCT.2015.7456971.
- [10] S. Mallaiah, Ajit Danti, and N. S. K , Classification of Diseased Arecanut based on Texture Features , *International Journal of Computer Applications*, vol. NCRAIT 3, 1–6, (2014), [Online]. Available: /proceedings/ncrait/number3/15152-1419/
- [11] T. Liu, J. Xie, Y. He, M. Xu, and C. Qin, An automatic classification method for betel nut based on computer vision, *International Conference on Robotics and Biomimetics (ROBIO)*, 1264–1267, (2009), doi: 10.1109/ROBIO.2009.5420823.
- [12] K.-Y. Huang, Detection and classification of areca nuts with machine vision, *Computers & Mathematics with Applications*, vol. 64, no. 5, 739–746, (2012), doi: 10.1016/j.camwa.2011.11.041.
- [13] J. Naranjo-Torres, M. Mora, R. Hernández-García, R. J. Barrientos, C. Fredes, and A. Valenzuela, A Review of Convolutional Neural Network Applied to Fruit Image Processing, *Applied Sciences*, vol. 10, no. 10, 3443, (2020), doi: 10.3390/app10103443.
- [14] I. Goodfellow, Y. Bengio, and A. Courville, *Deep learning*. in Adaptive computation and machine learning. Cambridge, Massachusetts: The MIT Press, (2016).
- [15] Y. LeCun, Y. Bengio, and G. Hinton, Deep learning, *Nature*, vol. 521, no. 7553, 436–444, (2015), doi: 10.1038/nature14539.
- [16] M. D. Zeiler and R. Fergus, Visualizing and Understanding Convolutional Networks, in *Computer Vision – ECCV 2014*, vol. 8689, D. Fleet, T. Pajdla, B. Schiele, and T. Tuytelaars, Eds., Cham: Springer International Publishing, (2014), 818–833. doi: 10.1007/978-3-319-10590-1\_53.
- [17] W. Jia, Y. Tian, R. Luo, Z. Zhang, J. Lian, and Y. Zheng, Detection and segmentation of overlapped fruits based on optimized mask R-CNN application in apple harvesting robot, *Computers and Electronics in Agriculture*, vol. 172, 105380, (2020), doi: 10.1016/j.compag.2020.105380.
- [18] X. Mai, H. Zhang, X. Jia, and M. Q.-H. Meng, Faster R-CNN With Classifier Fusion for Automatic Detection of Small Fruits, *IEEE Trans. Automat. Sci. Eng.*, 1–15, (2020), doi: 10.1109/TASE.2020.2964289.
- [19] M. Khoshdeli, R. Cong, and B. Parvin, Detection of nuclei in H&E stained sections using convolutional neural networks, in *2017 IEEE EMBS International Conference on Biomedical & Health Informatics (BHI)*, Orland, FL, USA: IEEE, (2017), 105–108. doi: 10.1109/BHI.2017.7897216.
- [20] A. Krizhevsky, I. Sutskever, and G. E. Hinton, ImageNet classification with deep convolutional neural networks, *Commun. ACM*, vol. 60, no. 6, 84–90, (2017), doi: 10.1145/3065386.
- [21] S. Aboshosha, O. Zahran, M. I. Dessouky, and F. E. Abd El-Samie, Resolution and quality enhancement of images using interpolation and contrast limited adaptive histogram equalization, *Multimed Tools Appl*, vol. 78, no. 13, 18751–18786, (2019), doi: 10.1007/s11042-018-7022-1.

- [22] Z. Zhong, L. Zheng, G. Kang, S. Li, and Y. Yang, Random Erasing Data Augmentation, *AAAI*, vol. 34, no. 07, 13001–13008, (2020), doi: 10.1609/aaai.v34i07.7000.
- [23] T.-T. Wong and P.-Y. Yeh, Reliable Accuracy Estimates from  $k$  -Fold Cross Validation, *IEEE Trans. Knowl. Data Eng.*, vol. 32, no. 8, 1586–1594, (2020), doi: 10.1109/TKDE.2019.2912815.