Journal of Systems Engineering and Information Technology (JOSEIT)
Volume 4, No. 1, April (2025), Pages 14-19
e-ISSN: 2829-310X | 10.29207/joseit.

Multi-Class CNN Models for Banana Ripeness Classification

Rafaela S. Francisco[1*], Gabriel de S. G. Pedroso[2], Thiago M. Ventura[3]
[123]Institute of Computing - Federal University of Mato Grosso (UFMT) Cuiabá - MT - Brazil

**rafaela.souzaf.rsf.rsf@gmail.com*

Abstract. This study develops and evaluates Convolutional Neural Network (CNN) models for classifying banana maturity stages using images, thereby addressing a significant challenge in the banana supply chain. The banana industry represents a major agricultural sector worldwide, with Brazil exporting 56.2 thousand tons in 2023. Accurate maturity classification is essential for optimizing harvest timing, reducing postharvest losses, and extending shelf life. We utilized a public Brazilian dataset of 1,000 images of Prata Catarina banana categorized into eight ripening stages based on peel coloration standards established by the Brazilian Program for Horticulture Modernization. The images were preprocessed to a standardized 200 × 200-pixel resolution, and we evaluated the effectiveness of the data augmentation techniques, including horizontal flip, vertical flip, rotation, and zoom. Our CNN architecture consisted of five convolutional blocks with a dropout layer prior to flattening. We conducted six experiments to compare three classification scenarios (eight, five, and two ripeness classes) with and without data augmentation. Our findings demonstrate that CNN models can effectively classify banana ripeness, with performance improving significantly as classification granularity decreases. The best-performing model achieved 89.5% accuracy, 87.2% precision, and 89.6% recall when classifying bananas into two categories.

Keywords: *banana ripeness classification; convolutional neural networks; image processing; agricultural automation; data augmentation*

How to cites:
Rafaela S. Francisco, Gabriel de S. G. Pedroso, & Thiago M. Ventura. (2025). Multi-Class CNN Models for Banana Ripeness Classification. Journal of Systems Engineering and Information Technology (JOSEIT), 4(1). https://doi.org/10.29207/joseit.v4i1.6540

Received by the Editor: April 14, 2025
Final Revision: April 25, 2025
Published: 2025-04-30

SDGs contributed:
9 INDUSTRY, INNOVATION AND INFRASTRUCTURE

1. Introduction

The banana industry is one of the most widely farmed and consumed fruits in the world, which is one of the reasons why it is one of the major agricultural commerce sectors worldwide [1]. For this reason, it is vital to have a solid understanding of the most effective production, harvesting, transportation, and storage techniques to ensure the safety of food supplies worldwide. In addition, Damasceno et al. (2019), Brazil exported 56,2 thousand tons of bananas by 2023 [2]. This had a substantial impact on the agricultural industry, which is responsible for 49 percent of Brazil's total exports [3], [4], [5].

Owing to the relevance and effect of the banana supply chain, several initiatives have been developed to address the problems that have been experienced. Consequently, agricultural production operations have become more competitive and efficient. Maturity is one of these issues that must be addressed, and the maturity of the banana is an essential component in the harvesting process [6], [7] because it has a substantial impact on both the quality and market value of the fruit [8]. Farmers are required to determine the stages of maturity that bananas have reached to reduce losses during the post-harvest period and significantly extend the shelf life of the fruit.

The utilization of Artificial Intelligence (AI) models in the production of bananas improves the comprehension of the various phases of ripeness [9], [10], which in turn makes it easier to choose the most appropriate time to harvest the bananas. Because of this, the completed product is guaranteed to be of excellent quality, which not only satisfies the expectations of consumers but also enhances the reputation of the business. Sabilla et al. (2019) investigated three unique phases of ripeness: unripe, ripe, and overripe maturity [11]. The overripe category was maintained by Mazen and Nashat (2019), who also investigated three earlier stages: green, yellowish, and half-ripe [12].

Increasing productivity, reducing waste, upgrading agricultural practices, and creating a more sustainable food industry are all possible outcomes that can be achieved via the utilization of artificial intelligence approaches and models for determining the maturity level of bananas. Through the utilization of Artificial Neural Network methods, the purpose of this study is to develop a model capable of determining the ripeness of a banana based on its image. This model will consider practical applications, and will only require the camera of a mobile device.

2. Methods

The photo dataset of Prata catarina banana, supplied by Martins Neto et al. (2023), was employed in this investigation [12]. Photos of bananas sorted into eight different phases of ripeness are included in the repository, which is the first Brazilian dataset to be made available to the public. The collection comprises one one thousand photographs, ranging in resolution from 2248×4000 to 3120×4160, and is categorized into eight distinct stages of ripeness.

The classification criteria for bananas set by the Brazilian Program for the Modernization of Horticulture and Integrated Fruit Production served as the basis for the establishment of these stages [13]. The eight stages were classified according to banana peel coloration. The stages were as follows: completely green, green with slight yellow, mostly green with some yellow, predominantly yellow with some green, yellow with green tips, fully yellow, yellow speckled with brown spots, and yellow with extensive browning. The eight ripening phases that bananas go through are depicted in Figure 1.

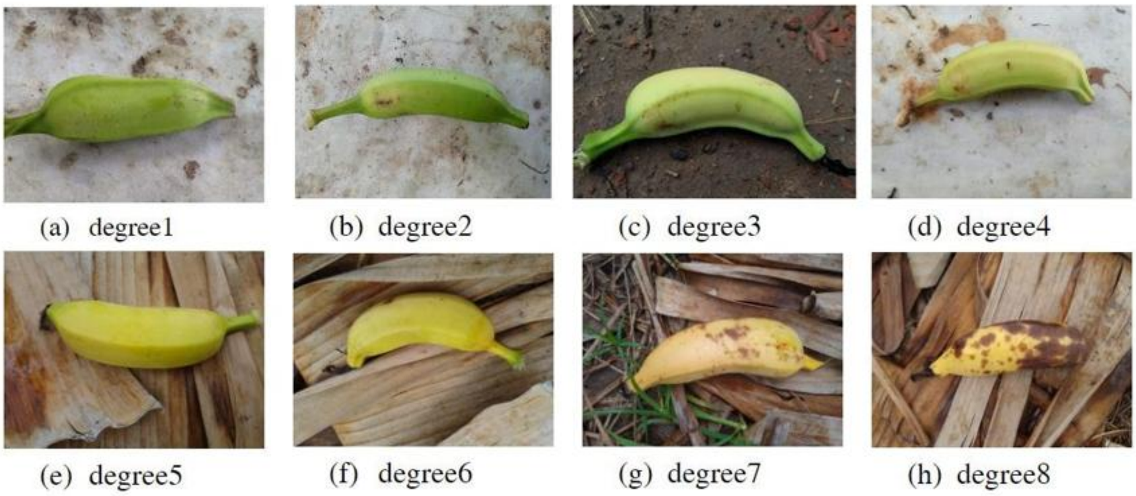

Figure 1. The eight stages of banana maturity, ranging from (a) "entirely green" to (h) "yellow with extensive brown spots."

The preceding section referenced many maturity classifications. To enhance the comparability of the findings, we further evaluated the separation of the stages into five and two classes, alongside the initial division into eight classes. In the assessment, including five categories, bananas were classified as green (grades 1 and 2), green-nearing yellow (grades 3 and 4), yellow (grades 5 and 6), somewhat yellow (grade 7), and heavily speckled (grade 8). In the assessment, which included only two classes, a distinction was made between green bananas (grades 1–4) and yellow bananas (grades 5–8).

For preprocessing, a variant of the data was produced by data augmentation (DA), a technique that involves creating synthetic data from existing data, thereby enhancing the overall data volume [14], [15]. Four distinct operations were employed: horizontal flip, vertical flip, rotation, and zoom, with probabilities of 50%, 50%, 70%, and 50%, respectively, for the occurrence of each operation for every image. Subsequent to the procedure, identical photographs were retained for each of the following phases with eight, five, or two classes, resulting in an additional 150 images for training purposes. The initial three processes were selected because they do not modify the banana's structure as they occur naturally. The zoom was adjusted to replicate the various image capture methods.

Convolutional Neural Network (CNN) models have been developed [16], [17]. Python programming language was utilized, particularly the Keras packages for constructing deep learning models [18], [19] and the augmentor for data augmentation [14], [20]. The outcomes of the developed models were evaluated using the metrics of accuracy, precision, and recall, as delineated in Equations 1, 2, and 3, respectively:

$$\text{Accuracy} = \frac{TA}{\mathrm{N}} \qquad (1)$$

$$\text{Precision} = \frac{TP}{\mathrm{TP+FP}} \qquad (2)$$

Recall $= \frac{TP}{\mathrm{TP+FN}}$ (3)

In conclusion, six tests were conducted to assess the use of CNN in the maturity categorization of Catarina silver bananas, considering three maturity level separations and the inclusion or exclusion of data augmentation techniques.

3. Results and Discussion

Accordingly, the entire dataset compiled by Martins Neto et al. (2023) was employed. Two different high resolutions were included in the source photographs. Standardization of the resolution was necessary to guarantee that the models received the same input. Simultaneously, a reduction in resolution was necessary to improve computational efficiency and eliminate characteristics that were not necessary for class categorization.

Based on the results of tests that evaluated a number of different resolutions, the decision was made to use photos with a resolution of 200 × 200 pixels. The needs of the experiment were taken into consideration when implementing data augmentation, which was performed in addition to changing the resolution. Several examples of these four approaches are shown in Figure 2. These techniques include zooming, rotation, horizontal and vertical flipping. To develop and evaluate the models, data were divided into three categories: 70 percent for training, 20 percent for validation, and ten percent for testing. This was simply the training set subjected to the data augmentation procedure.

Figure 2. Illustrations of data augmentation techniques, including horizontal flip, vertical flip, rotation, and zoom.

Interestingly, our analysis revealed that data augmentation techniques, contrary to conventional expectations in deep learning applications, generally reduced the model performance across most experimental configurations. To investigate this counterintuitive finding, we conducted a detailed examination of the augmented images and their impact on classification accuracy. The primary challenge appears to stem from the distinctive visual characteristics of the banana ripening patterns. When subjected to vertical flipping, the natural orientation of bananas is inverted, creating representations rarely encountered in real-world scenarios. Similarly, excessive rotation (applied with 70% probability) often distorted the distinctive curvature and color gradient patterns, which served as critical indicators of ripeness stages.

Further examination of the misclassified instances revealed that augmentation particularly affected the differentiation between adjacent ripeness classes. For example, in the 8-class model, augmented images of stage 4 (predominantly yellow with some green) were frequently misclassified as stage 3 (mostly green with some yellow tips) or stage 5 (yellow with green tips). The color transformations resulting from zoom operations sometimes alter the perceived color distribution across the banana surface, artificially modifying the features crucial for accurate ripeness assessment.

We observed that horizontal flipping was the least detrimental augmentation technique, likely because it preserved the natural orientation of color progression in bananas. Conversely, zoom operations proved to be the most problematic, particularly when distinguishing between Stages 7 and 8, where brown speckling patterns are critical differentiators. The zoomed images sometimes magnify or diminish these speckles, leading to classification errors. Notably, the impact of data augmentation was less pronounced in the binary classification scenario (green versus yellow bananas), where the distinguishing features were sufficiently robust to withstand the transformation. This suggests that data augmentation might still be beneficial for coarse-grained classification tasks but requires careful calibration for finer-grained distinctions in banana ripeness classification.

In the course of this research, a number of tests were conducted to determine the layers that constitute the CNN model. The tests included models that contained three to five convolutional blocks (Convolutional+ReLU and Max Pooling), with variations in the output dimension that either matched or differed from the input. The purpose of these tests was to evaluate the ability to capture broader regions as opposed to the allocation to specific points that were influenced by the movement of the convolution window across the images. As a result of the alterations made to achieve a satisfactory conclusion, the same model was utilized across all experiments. The only modification that was made was to the final classification layer to accommodate the varying number of classes that were present in each trial (as described in the methodology section).

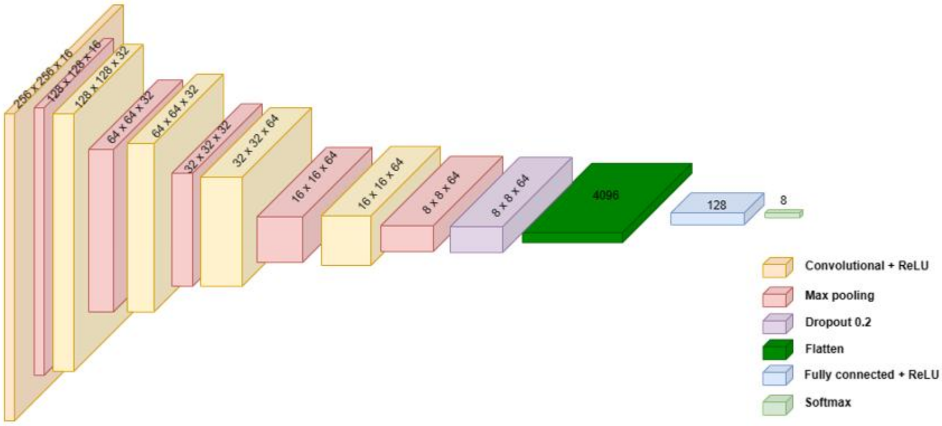

Figure 3. The architecture of the developed CNN model is illustrated using an experiment involving eight classes as a case study.

Five convolutional blocks were established, incorporating a dropout layer that reduced weights by 20% prior to flattening, culminating in a fully connected layer utilizing the ReLU activation function, designed for eight output classes for each experiment. Figure 3 shows an example of the construction of the model. All models were trained across 10 seasons with a batch size of 32 samples, employing the Adam optimizer (Kingma and Ba, 2014) for parameter optimization, and utilizing a sparse categorical cross-entropy loss function.

Many models have been constructed and assessed using the criteria of accuracy, precision, and recall. The outcomes derived from both the training and testing datasets are shown in Tables 1 (training) and 2 (testing).

Table 1. Results of models used to categorize trials using 8, 5, and 2 classes, training data, and data augmentation (percentages show results)

Number of Classes	Accuracy		Precision		Recall	
	Without DA	**With AD**	**Without DA**	**With AD**	**Without DA**	**With AD**
8 classes	52,3	46,0	51,0	49,0	40,0	43,4
5 classes	65,3	69,2	64,3	72,4	61,0	63,0
2 classes	88,6	78,6	87,9	79,4	87,7	81,7

Table 2. Results of models used to categorize trials using 8, 5, and 2 classes together with test data, utilizing or not data augmentation (percentages show results)

Number of Classes	Accuracy		Precision		Recall	
	Without DA	**With AD**	**Without DA**	**With AD**	**Without DA**	**With AD**
8 classes	45,3	34,7	30,3	23,3	36,7	31,1
5 classes	57,9	60,0	59,3	48,6	56,5	52,3
2 classes	89,5	76,8	87,2	77,0	89,6	81,3

The outcomes for the eight classes during training, with and without data augmentation, yielded accuracies of 52.3% and 46%, respectively. Precision was recorded at 51% and 49%, whereas recall was superior with data augmentation at 43.4%, in contrast to 40% without it. For the test data, the accuracy was 45.3% without data augmentation and 34.7% with it; the precision was 30.3% and 23.3%, while the recall was 36.7% and 31.1%, respectively. The optimal performance for this experiment was 45.3%, considering the eight distinct stages of development, and utilizing only the original data for training.

The outcomes for the five classes exhibited distinct behaviors compared with the model trained with eight classes. As anticipated, the overall performance improved owing to the reduced number of classes. The accuracies were 65.3% and 69.2% for training with and without data augmentation, respectively, and the test accuracies were 57.9% and 60%, respectively. The precision was 64.3% for training without data augmentation and 72.4% for training with data augmentation, accompanied by recall values of 61% and 63%, respectively. In the test, the precision was 59.3% and 48.6%, while the recall was 56.5% and 52.3%, respectively, with and without data augmentation.

Ultimately, with only two classes, the model demonstrated enhanced generalization capability, achieving accuracies of 88.6% and 78.6% with and without data augmentation, respectively, during training. The accuracy was 89.5% without data augmentation and 76.8% with data augmentation using test data. The precision was 87.9% and 79.4% during training, whereas the recall was 87.7% and 81.7% with and without data augmentation, respectively. In the test, the precision was 87.2% and 77%, while the recall was 89.6% and 81.3% for the scenarios with and without data augmentation, respectively.

4. Conclusions

The objective of this study was to develop a CNN model to determine the maturity of Prata catarina bananas. The development of multiple models resulted in an accuracy with test data ranging from 45% to 89.5%. The ideal precision and recall were determined to be 87.2% and 89.6%, respectively, with the optimal outcomes contingent on the number of classes requiring detection. An increase in the number of classes leads to a greater degree of resemblance across instances, which in turn complicates the procedure. Future research should concentrate on several promising methodological enhancements to improve classification performance. The efficacy of transfer learning methodologies employing pretrained networks, such as ResNet-50, MobileNetV3, or EfficientNet, should be assessed. These architectures exhibit superior feature extraction capabilities in analogous visual classification tasks while requiring a reduced number of training samples. Attention mechanisms such as squeeze-and-excitation blocks or transformer-based approaches have the potential to be integrated into the model to assist in focusing on the most discriminative regions of banana images. This integration could lead to enhanced fine-grained classification between visually similar ripening stages.

References

[1] M. P. S. Jiménez and I. Tobes, "Upcycling the Banana Industry in Ecuador: A Methodology to Estimate Biowaste and Catalogue of Bioproducts," *Int J Adv Sci Eng Inf Technol*, vol. 13, no. 5, 2023, doi: 10.18517/ijaseit.13.5.19031.

[2] C. L. Damasceno *et al.*, "Postharvest biocontrol of anthracnose in bananas by endophytic and soil rhizosphere bacteria associated with sisal (Agave sisalana) in Brazil," *Biological Control*, vol. 137, 2019, doi: 10.1016/j.biocontrol.2019.104016.

[3] D. H. Reinhardt, J. A. Dos Santos-Serejo, and J. Da, "Panorama of the banana industry in latin america and the caribbean islands, with a special focus on brazil," *Acta Hortic*, vol. 986, 2013, doi: 10.17660/actahortic.2013.986.1.

[4] G. M. Napoleão, P. R. R. Jesus, and S. Leonel, "Cultivar diversification of banana production in Brazil," *Agronomy Science and Biotechnology*, vol. 7, 2021, doi: 10.33158/asb.r127.v7.2021.

[5] É. I. A. de Souza, H. dos S. Pereira, and A. E. D. de Sousa, "Evaluation of losses in bananas produced in the Rio Preto da Eva, Amazonas, Brazil," *Revista em Agronegocio e Meio Ambiente*, vol. 14, no. 3, 2021, doi: 10.17765/2176-9168.2021V14N3E8269.

[6] D. Surya Prabha and J. Satheesh Kumar, "Assessment of banana fruit maturity by image processing technique," *J Food Sci Technol*, vol. 52, no. 3, 2015, doi: 10.1007/s13197-013-1188-3.

[7] M. Amin, M. Hossain, M. Rahim, and M. Uddin, "Determination of optimum maturity stage of banana," *Bangladesh Journal of Agricultural Research*, vol. 40, no. 2, 2015, doi: 10.3329/bjar.v40i2.24557.

[8] L. Ma *et al.*, "Prediction of banana maturity based on the sweetness and color values of different segments during ripening," *Curr Res Food Sci*, vol. 5, 2022, doi: 10.1016/j.crfs.2022.08.024.

[9] J. Enríquez and M. Macas, "Artificial Intelligence-Based Banana Ripeness Detection," in *Communications in Computer and Information Science*, 2023. doi: 10.1007/978-3-031-24985-3_15.

[10] E. Almeyda and W. Ipanaqué, "RECENT DEVELOPMENTS OF ARTIFICIAL INTELLIGENCE FOR BANANA: APPLICATION AREAS, LEARNING ALGORITHMS, AND FUTURE CHALLENGES," *Engenharia Agricola*, vol. 42, no. SpecialIssue, 2022, doi: 10.1590/1809-4430-Eng.Agric.v42nepe20210144/2022.

[11] I. A. Sabilla, C. S. Wahyuni, C. Fatichah, and D. Herumurti, "Determining banana types and ripeness from image using machine learning methods," in *Proceeding - 2019 International Conference of Artificial Intelligence and Information Technology, ICAIIT 2019*, 2019. doi: 10.1109/ICAIIT.2019.8834490.

[12] F. M. A. Mazen and A. A. Nashat, "Ripeness Classification of Bananas Using an Artificial Neural Network," *Arab J Sci Eng*, vol. 44, no. 8, 2019, doi: 10.1007/s13369-018-03695-5.

[13] G. Morais *et al.*, "Classification Criteria for Cavendish Bananas and a Framework to Indicate the Correct Destination," in *The 1st International Online Conference on Agriculture—Advances in Agricultural Science and Technology*, Basel Switzerland: MDPI, Jul. 2022, p. 65. doi: 10.3390/IOCAG2022-12313.

[14] C. Shorten and T. M. Khoshgoftaar, "A survey on Image Data Augmentation for Deep Learning," *J Big Data*, vol. 6, no. 1, 2019, doi: 10.1186/s40537-019-0197-0.

[15] C. Shorten, T. M. Khoshgoftaar, and B. Furht, "Text Data Augmentation for Deep Learning," *J Big Data*, vol. 8, no. 1, 2021, doi: 10.1186/s40537-021-00492-0.

[16] R. Amalia and F. Panjaitan, “Mask Detection Using Convolutional Neural Network Algorithm,” *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 6, no. 4, pp. 639–647, 2022, doi: 10.29207/resti.v6i4.4276.
[17] O. D. Annesa, C. Kartiko, and A. Prasetiadi, “Identification of Reptile Species Using Convolutional Neural Networks (CNN),” *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 4, no. 5, pp. 899–906, 2020, doi: 10.29207/resti.v4i5.2282.
[18] V. Chang, R. O. Eniola, L. Golightly, and Q. A. Xu, “An Exploration into Human–Computer Interaction: Hand Gesture Recognition Management in a Challenging Environment,” *SN Comput Sci*, vol. 4, no. 5, 2023, doi: 10.1007/s42979-023-01751-y.
[19] J. Hung *et al.*, “Keras R-CNN: Library for cell detection in biological images using deep neural networks,” *BMC Bioinformatics*, vol. 21, no. 1, 2020, doi: 10.1186/s12859-020-03635-x.
[20] A. Muhandisin and Y. Azhar, “Malaria Blood Cell Image Classification using Transfer Learning with Fine-Tune ResNet50 and Data Augmentation,” *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 6, no. 5, pp. 891–897, 2022, doi: 10.29207/resti.v6i5.4322.