

# A Deep Transfer Learning Approach for Accurate Dragon Fruit Ripeness Classification and Visual Explanation using Grad-CAM

Hoang-Tu Vo, Nhon Nguyen Thien, Kheo Chau Mui  
Software Engineering Department  
FPT University, Cantho city, Vietnam

**Abstract**—Dragon fruit, known for its rich antioxidant content and low-calorie attributes, has garnered significant attention as a health-promoting fruit. Its economic value has also surged due to increasing consumer demand and its potential as an export commodity in various regions. The classification of dragon fruit ripeness is a pivotal task in ensuring product quality and minimizing post-harvest losses. This research article presents a comprehensive study on the classification of ripe and unripe dragon fruits (*Hylocereus* spp) using the Densenet201 model through three distinct approaches: as a classifier, feature extractor, and fine-tuner. To explain the outcomes of the image classification model and thereby enhance its performance, optimization, and reliability, this study employs advanced visualization techniques. Specifically, it utilizes Grad-CAM (Gradient-weighted Class Activation Mapping) and Guided Grad-CAM techniques. These techniques offer insights into the model's decision-making process and pinpoint regions of interest within the images. This approach empowers researchers to iteratively validate the model's accuracy and enhance its performance. The utilization of Densenet201 as a classifier, feature extractor, and fine-tuner, coupled with the insights from Grad-Cam and Guided Grad-Cam, presents a holistic approach to enhancing dragon fruit ripeness classification. The findings contribute to the broader discourse on agricultural technology, image analysis, and the optimization of classification models.

**Keywords**—*Dragon fruit classification; ripeness classification; densenet201 model; Grad-CAM visualization; guided grad-CAM; visual interpretation; Explainable AI; XAI; deep learning; pre-trained models; model fine-tuning; transfer learning*

## I. INTRODUCTION

Dragon fruit is a tropical fruit that is widely grown in many countries around the world, including Vietnam [1]. It contains numerous nutrients beneficial to human health, such as vitamin A, vitamin C, and protein [2], [3]. While the ripeness of dragon fruit can be observed visually and harvested manually by humans, the necessity arises for the integration of equipment and robots in the field due to the vast planting area and technological advancements. The automated system is capable of efficiently harvesting significant quantities of dragon fruit within a brief timeframe, leading to time savings in the harvesting process. Therefore, an automatic dragon fruit ripeness grading system is really necessary.

Explaining deep learning models is of paramount importance in today's rapidly evolving technological landscape. Deep learning, with its complex architectures and black-box nature, has demonstrated remarkable capabilities across various domains. However, this complexity often comes at the

cost of interpretability, creating challenges in understanding how and why these models arrive at their decisions. Explaining these models is a critical step to building trust, ensuring fairness, and enabling effective adoption. Interpretability empowers stakeholders to comprehend the factors influencing predictions. By shedding light on the inner workings of deep learning models, explanations help bridge the gap between advanced machine learning techniques and human comprehension, fostering collaboration between data scientists, domain experts, and end-users.

This research paper provides a comprehensive examination of how to classify ripe and unripe dragon fruits (*Hylocereus* spp) using the Densenet201 model. The study explores three distinct approaches to employing the model: as a classifier, a feature extractor, and a fine-tuner. To expound upon the outcomes of the image classification model and, in turn, enhance its performance and reliability, advanced visualization techniques are applied. Specifically, the study makes use of Grad-CAM (Gradient-weighted Class Activation Mapping) and Guided Grad-CAM techniques.

## II. RELATED WORKS

Lately, the domains of image recognition have witnessed extensive utilization of Deep Learning (DL) and Machine Learning (ML) techniques. Including medicine [4], [5], [6], [7]. In self-driving cars [8], [9], [10], [11]. In agriculture [12], [13], [14], [15]. Especially when it comes to automating fruit sorting [16], [17], [18], etc.

Currently, there is a lot of applied research on machine learning, and deep learning on dragon fruit for many different purposes such as: In this study [19], Minh Trieu, N., & Thinh, N. T. present an automated system for classifying dragon fruit, which relies on a convolutional neural network (CNN). This classification system integrates machine learning and image processing through a convolutional neural network model to discern the external characteristics of dragon fruits. The paper [20] presents an automated dragon fruit classification system through a combination of KNN, CNN, and ANN models for identification, feature extraction, and classification. This study [21] devises dragon fruit grading and sorting techniques via machine learning algorithms (CNN, ANN, and SVM). Zhou et al. in this paper [22] presents a novel dragon fruit detection method, utilizing YOLOv7 to locate and classify the dragon fruit and further detect the endpoints of the dragon fruit. Vijayakumar, D. T., & Vinothkanna, M. R. in the paper

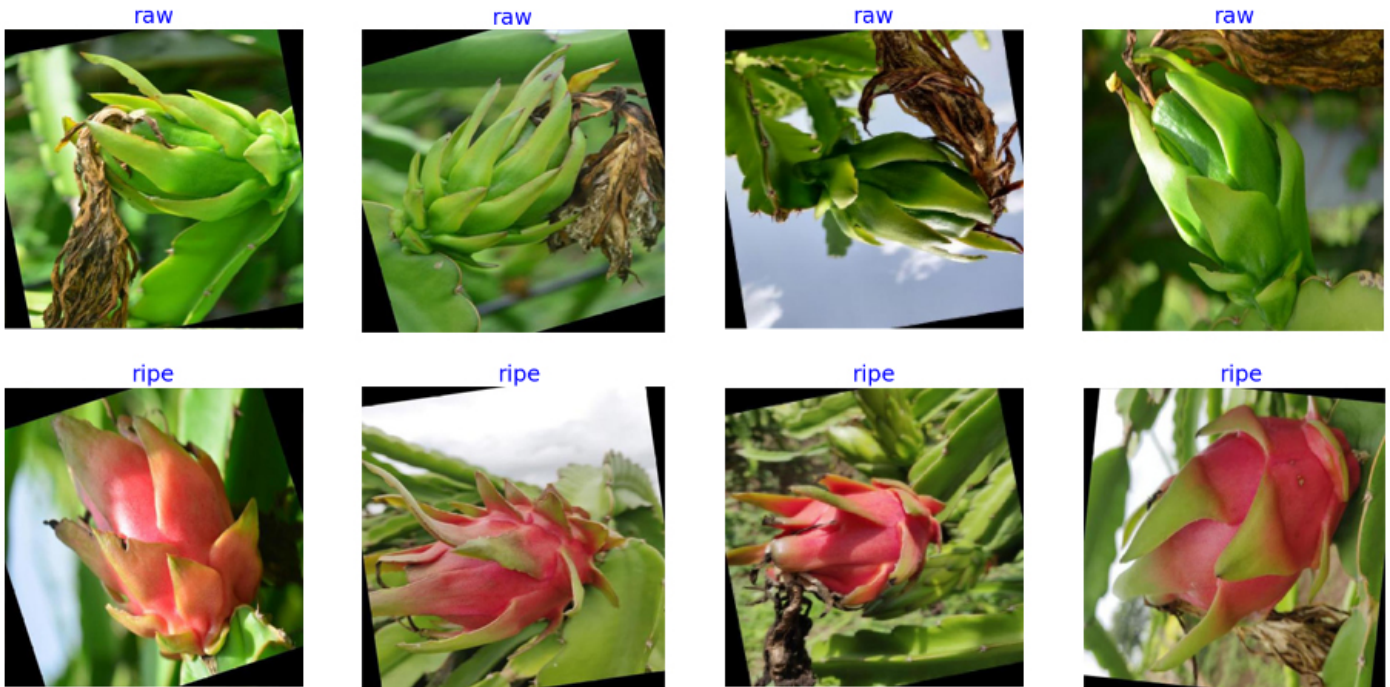


Fig. 1. Some images of dragon fruit dataset.

[23] introduces the utilization of the RESNET 152 deep learning convolutional neural network to identify dragon fruit mellowness, signifying the optimal time for harvest. This study [24] applies color and texture feature extraction techniques, utilizing color moments and gray level co-occurrence matrices (GLCM), to develop a system for recognizing three types of dragon fruit stems through digital image processing, employing Support Vector Machine and k-Nearest Neighbors methods for comparison.

In today's context, with the increasing complexity of deep learning models, there is an even greater need to understand the decisions made by these models. The role of Explainable Artificial Intelligence (XAI) [25], [26] is to provide transparency, clarity, and understanding to the decision-making process of deep learning models and bring the gap between the "black-box" nature of AI and human understanding. XAI has achieved numerous successes across various domains, including: In medicine [27], [28], [29]; In agriculture: [30]; In traffic classification: [31].

The purpose of this research article is to conduct a thorough investigation into the classification of ripe and unripe dragon fruits utilizing the Densenet201 model. The study explores three distinct methodologies: employing the model as a classifier, feature extractor, and fine-tuner. In addition, using Grab-CAM [32], [33], [34] and Guided Grad-Cam (this involves performing an element-wise product between Grad-CAM and Guided Backpropagation [35]) to interpret models for the purpose of evaluating the decision and effectiveness of deep learning models detecting features in a dragon fruit image.

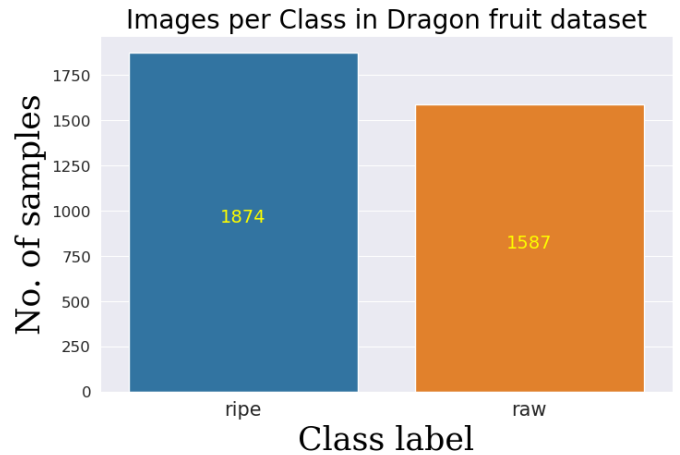


Fig. 2. A dataset distribution.

### III. METHODOLOGY

#### A. Data Collection and Preparation

In this investigation, we utilized a dataset containing 3,461 images, sourced from [36] with two classes: Ripe and Raw. Fig. 1 illustrates sample images of dragon fruit from this dataset, and the dataset distribution is visually depicted in Fig. 2. Prior to model training and evaluation, a preprocessing step is performed on the images, resizing them to 224x224 and applying an image preprocessing function. The data set is partitioned into three distinct sets, comprising a training set, a validation set, and a testing set, with a distribution ratio of 6:2:2.

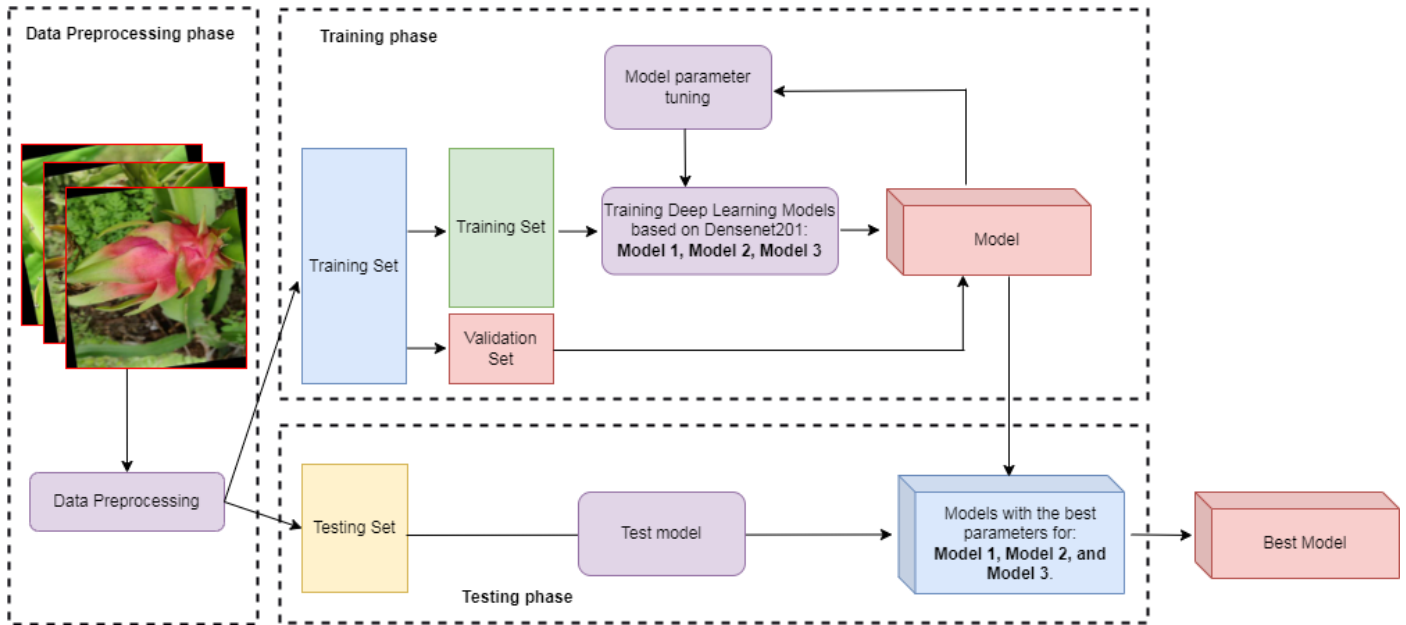


Fig. 3. The training and testing process involves three dragon fruit ripeness models based on densenet201.

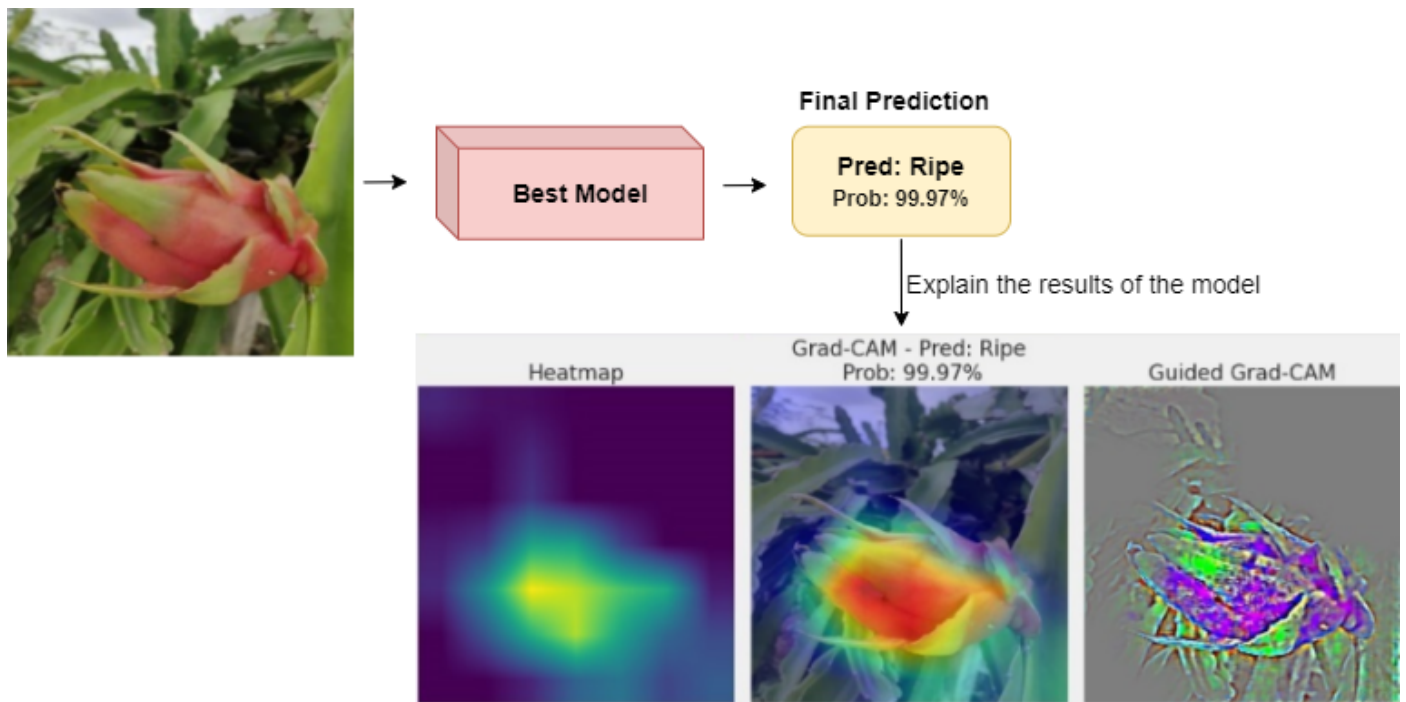


Fig. 4. Classify the dragon fruit ripeness and explain the results of the model.

### B. Transfer Learning for Dragon Fruit Classification

Transfer learning [37], [38], a pivotal concept in the realm of machine learning, is a technique that leverages knowledge gained from one task to improve performance on a different but related task. Rather than starting from scratch, transfer learning enables models to take advantage of patterns and representations learned from a source domain and apply them to a target domain with limited labeled data. This approach has

revolutionized the field by dramatically reducing the need for vast datasets and extensive computational resources, making it feasible to tackle new problems even when data is scarce. Transfer learning's ability to extract and transfer valuable insights across tasks has been instrumental in advancing the efficiency, accuracy, and generalization of machine learning models, thereby accelerating progress across various domains and enabling AI systems to learn more like humans – by

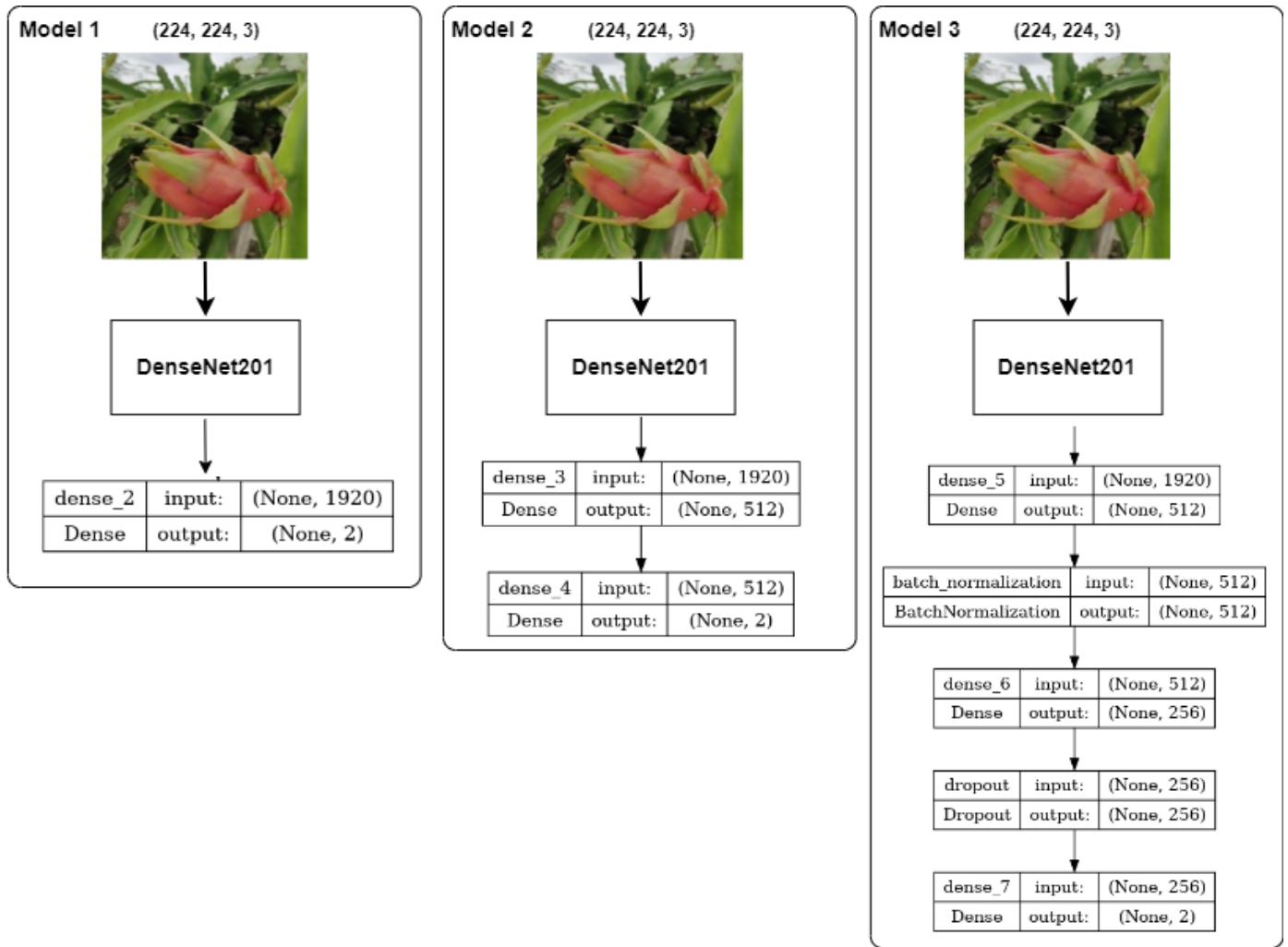


Fig. 5. The architecture of the dragon fruit ripeness classification is constructed using denseNet201.

building upon prior knowledge to solve new challenges.

### C. Proposed Model

This research paper presents a comprehensive investigation into the categorization of ripe and unripe dragon fruits utilizing the Densenet201 [39] model in three separate ways: as a classifier, feature extractor, and fine-tuner. Model 1 (as a classifier) is built upon the DenseNet201 architecture. The pre-trained layers are frozen to preserve their acquired features. The final softmax layer in the DenseNet201 model has been modified to classify two classes, corresponding to ripe and raw dragon fruit. Model 2 (as a feature extractor) is a modified DenseNet201-based model for the binary classification of ripe and raw dragon fruit. It has two new entirely connected layers, with 512 neurons in the first layer and two neurons in the final. Both the base pre-trained layers and the fully connected layers have been completely retrained. Model 3 (as a feature extractor with fine-tuning) is a base model initialized using the DenseNet201 architecture. A fully connected layer with 512 neurons is added on top of the base model's output. Batch normalization is applied to the output of the previous layer, helping to stabilize and accelerate training. Another fully

connected layer with 256 neurons is added, along with various regularization techniques. A dropout layer is introduced, which helps in preventing overfitting. A final fully connected layer with 2 neurons and softmax activation is added to produce class probabilities. The pre-trained layers are frozen to preserve their acquired feature. Fig. 3 shows The training and testing process involves three Dragon Fruit Ripeness models based on Densenet201. Fig. 4 displays Classify the Dragon Fruit Ripeness and explain the results of the model. The architecture of the Dragon Fruit Ripeness Classification is constructed using DenseNet201 shown in Fig. 5.

### D. Performance Evaluation Measures

In this study, various evaluation metrics, including Accuracy, F1-score, Precision, and Recall, were utilized to assess the effectiveness of the deep learning (DL) models. Accuracy served as a measure of overall performance, while Precision and Recall evaluated the model's ability to correctly predict positive instances. The F1-score provided a balanced perspective by considering both Precision and Recall, enabling informed judgments on the model's effectiveness. Through the

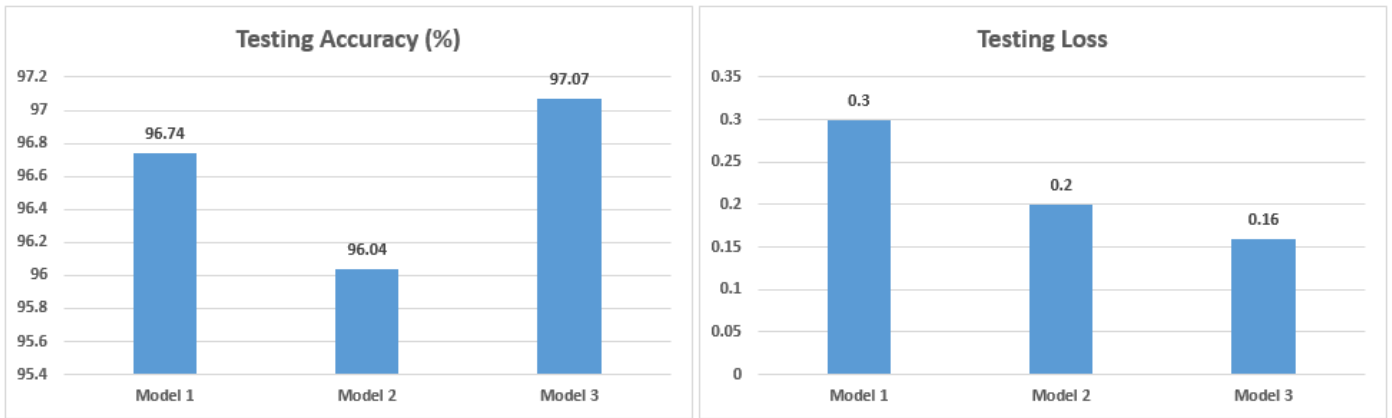


Fig. 6. Confusion matrix of the recommended models. (left) The number of predictions, (right) The percentage.

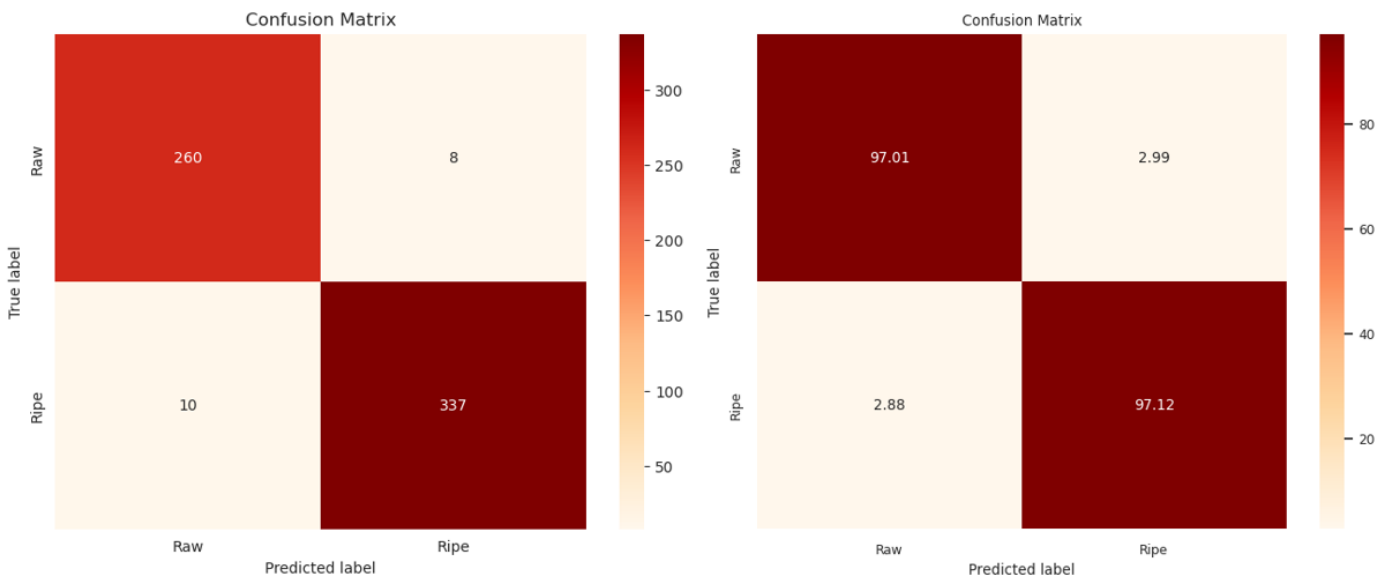


Fig. 7. Confusion matrix of Model 3. (left) The number of predictions, (right) The percentage.

utilization of various evaluation metrics, a thorough comprehension of the model's performance was attained.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

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

$$F_1 - Score = \frac{Precision * Recall}{Precision + Recall} \quad (4)$$

In which, TP represents True Positive, TN signifies True Negative, FP represents False Positive, and FN stands for False Negative.

## IV. RESULTS

### A. Environmental Settings

The experimental results were obtained by conducting the experiments on the Kaggle platform. The system used for the experiments had 13GB of RAM and a GPU P100 with 16GB of memory. The models were trained for a total of 35 epochs, and a batch size of 32 was used during the training process.

### B. Evaluation Overall

The results presented in Fig. 6 reveal that the majority of the models exhibit remarkably high accuracy and low loss. Notably, Model 3 stands out with the lowest test loss of 0.1606 and the highest accuracy of 97.07% among the three models. It demonstrates excellence on the test dataset. Additionally, both Model 1 and Model 2 perform exceptionally well, achieving test accuracies of 96.75% and 96.04%, respectively. However, they do have significantly higher test losses compared to Model 3.

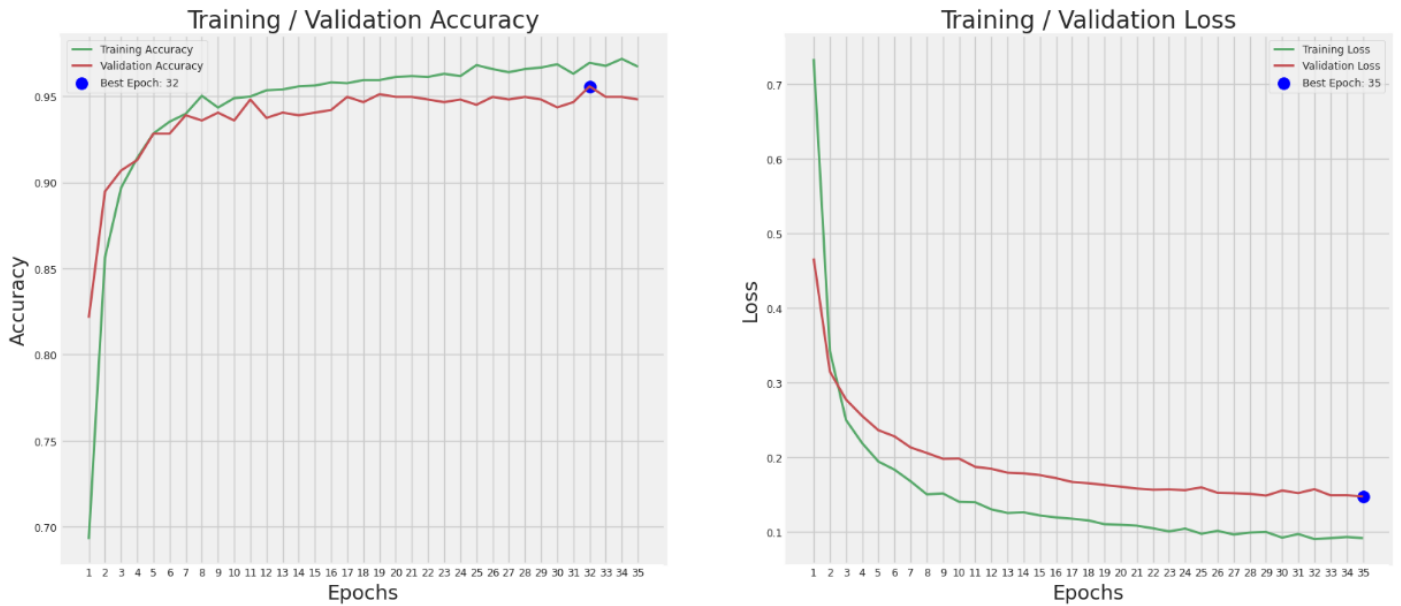


Fig. 8. Loss and accuracy plots of the model 3.

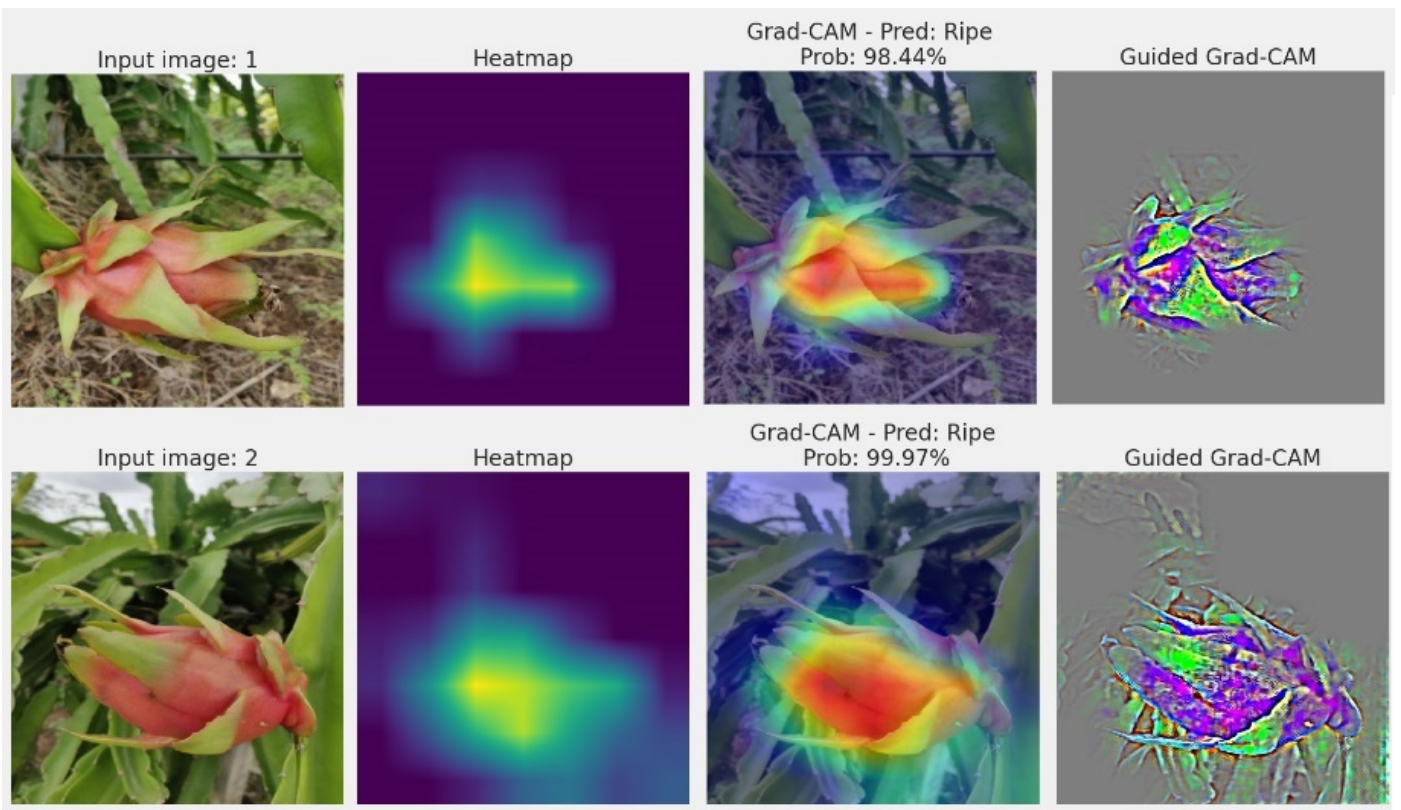


Fig. 9. Examples of the ripe feature explained using Grad-CAM from model 3.

The accuracy of Model 3 reaching 97.07% on the test set is outstanding and represents an exceptional level of performance. It means that the model can correctly classify almost all instances in the test data. Such a high level of accuracy is a strong indicator of the model's ability to generalize well and make precise predictions on previously unseen data. Fig.

7 illustrates the confusion matrix for Model 3.

Furthermore, Fig. 8 presents an overview of the performance metrics, including loss and accuracy, which were assessed throughout both the training and validation stages of Model 3.

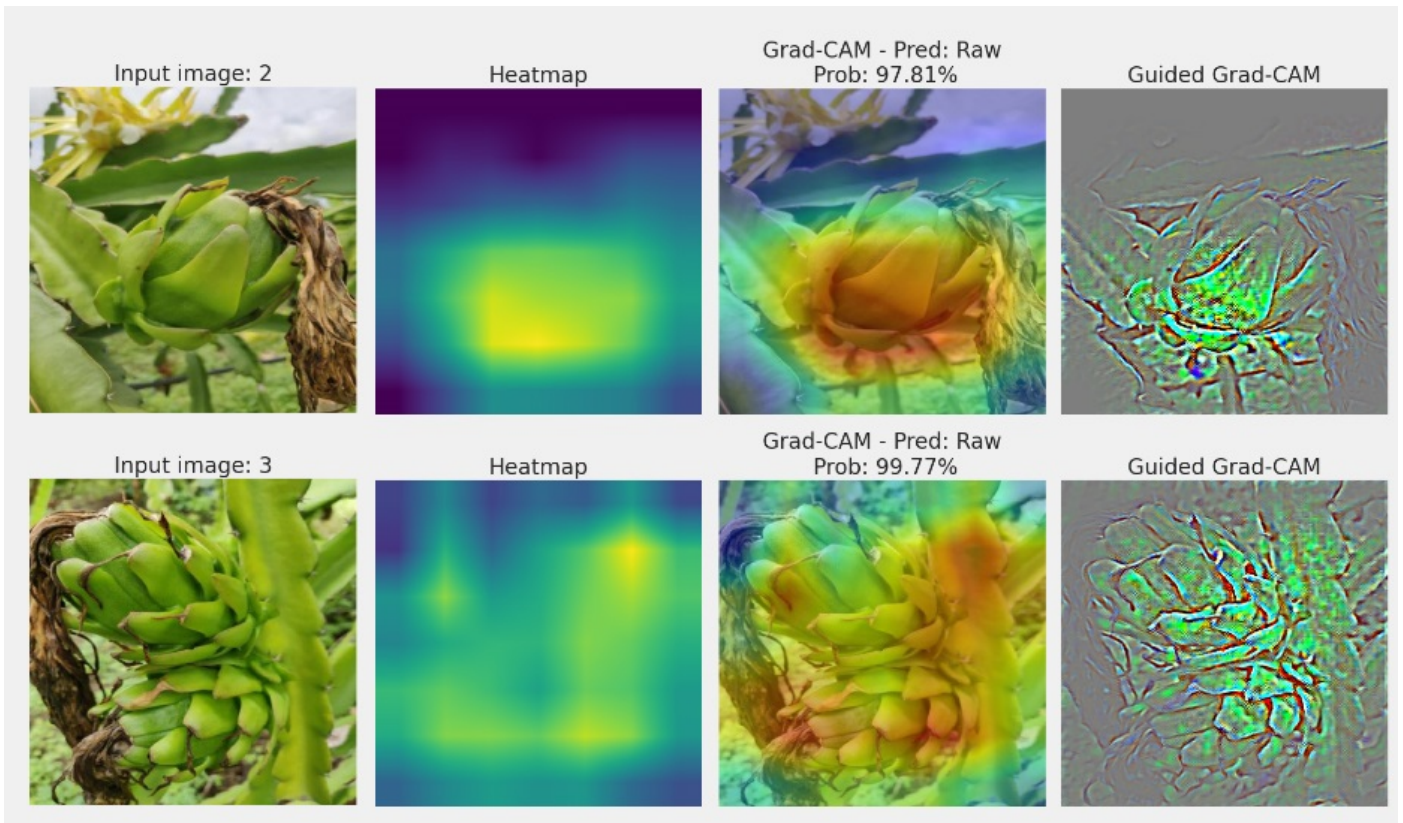


Fig. 10. Examples of the raw feature explained using Grad-CAM from model 3.

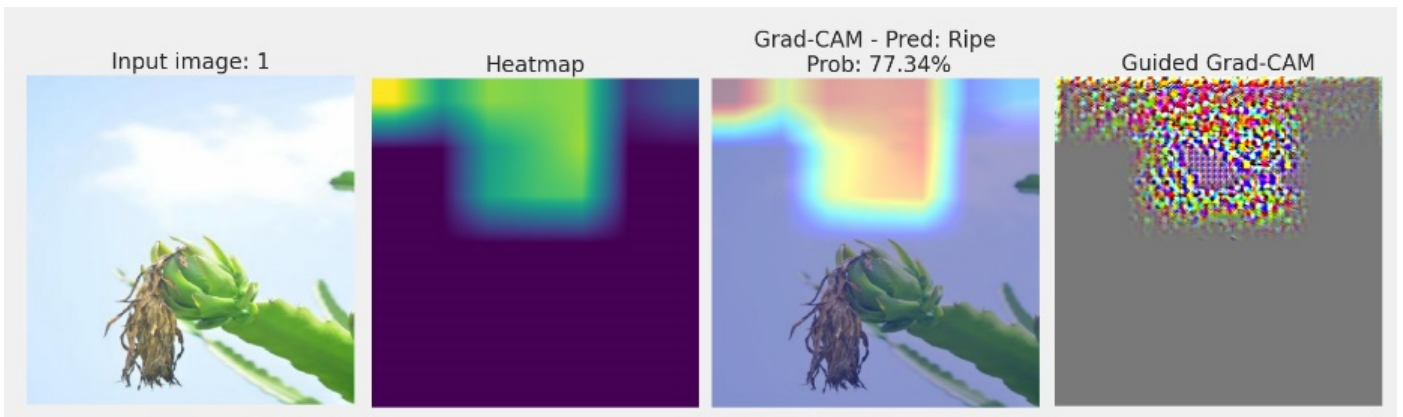


Fig. 11. Examples of ripe feature misclassification are explained using Grad-CAM from model 3.

The classification report Table I provides a detailed analysis of the evaluation metrics for each grape disease class. The classification report is a summary that assesses a model's classification performance. It includes precision, recall, and F1-score metrics for each class label. The report helps to evaluate the model's accuracy in correctly classifying instances for each class, with higher scores indicating better performance. In this case, the model achieved perfect scores for all classes, demonstrating excellent accuracy in its classification task.

TABLE I. THE MODEL RESULTS OF THE CLASSIFICATION REPORT

Class	Precision	Recall	F1-Score	Support
Raw	0.96	0.97	0.97	268
Ripe	0.98	0.97	0.97	347

### C. Visualizing the Interpretation of Model Predictions Using Grad-CAM

To better understand the significant regions in the images that the model focuses on for making predictions, the research team employed Grad-CAM on corner images. Specif-

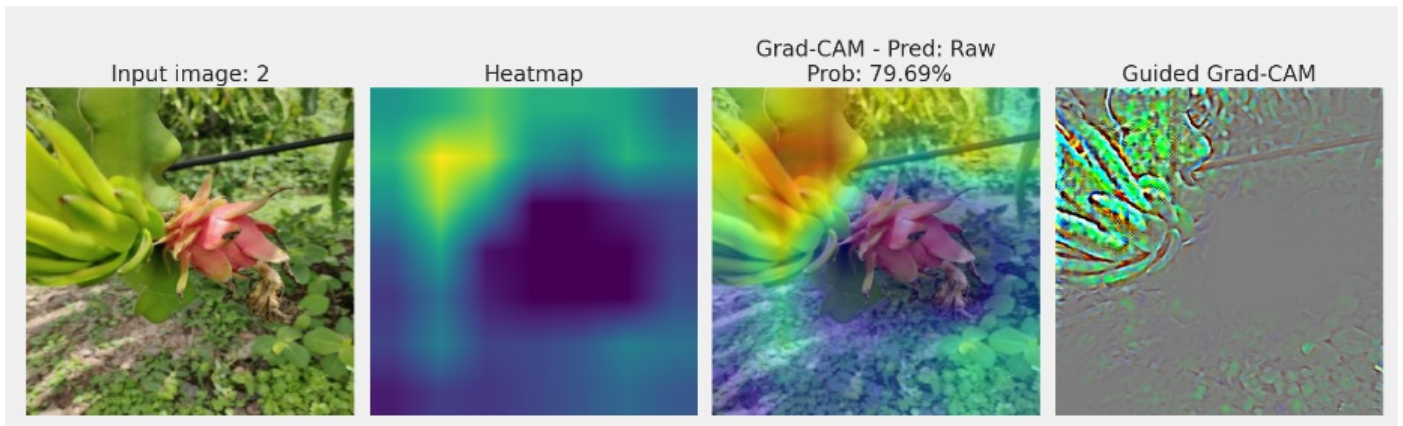


Fig. 12. Examples of raw feature misclassification are explained using Grad-CAM from model 3.

ically, when applying Grad-CAM to an image, it generates a “heatmap” that highlights important positions in the image. This heatmap indicates the areas of the image that the model is paying attention to while making predictions. High values on the heatmap usually correspond to important regions relevant to the classification decision. In addition, the research team also used Guided Grad-CAM to better understand the parts of the image that the model is interested in to make predictions.

Based on the results from Fig. 9 and Fig. 10, it is evident that the accurate classification by the model relies on specific regions unique to each type. The heatmap also highlights key areas in the image that the model focuses on while making predictions that align with its own predictions.

Grad-CAM and Guided Grad-CAM are also useful in comparing misclassified results to understand why the model may have “misinterpreted” certain images. For example, Fig. 11 and 12 explain why the model is misclassified.

## V. CONCLUSION

This study conducts an extensive investigation into the classification of ripe and unripe dragon fruit, employing the Densenet201 model across three distinct approaches: as a classifier, feature extractor, and fine-tuner. All three proposed models yield exceptionally impressive outcomes. Particularly noteworthy, Model 3 (functioning as a feature extractor with fine-tuning) stands out with the highest accuracy, achieving 97.07% among the three models. Furthermore, both Model 1 and Model 2 showcase exceptional performance. To delve deeper into understanding the significant areas within images that the model emphasizes for prediction, the research team applied Grad-CAM to corner images. Additionally, the team employed Guided Grad-CAM to enhance comprehension of image regions that capture the model’s attention for prediction purposes. Both Grad-CAM and Guided Grad-CAM prove to be invaluable tools in the comparative analysis of misclassified results, providing insights into the potential reasons for the model’s “misinterpretation” of specific images. The study team will use HiResCAM [40] in the future for model explanation. HiResCAM serves the same functions as Grad-CAM but with the added benefit of highlighting only the regions used by the model.

## REFERENCES

- [1] Edmundo M Mercado-Silva. Pitaya—hylocereus undatus (haw). *Exotic fruits*, pages 339–349, 2018.
- [2] Md Farid Hossain, Sharker Md Numan, and Shaheen Akhtar. Cultivation, nutritional value and health benefits of dragon fruit (hylocereus spp.): A review. *International Journal of Horticultural Science & Technology*, 8(3), 2021.
- [3] Truc-Linh Le, Nga Huynh, Pablo Quintela-Alonso, et al. Dragon fruit: A review of health benefits and nutrients and its sustainable development under climate changes in vietnam. *Czech Journal of Food Sciences*, 39(2):71–94, 2021.
- [4] Fei Wang, Lawrence Peter Casalino, and Dhruv Khullar. Deep learning in medicine—promise, progress, and challenges. *JAMA internal medicine*, 179(3):293–294, 2019.
- [5] Francesco Piccialli, Vittorio Di Somma, Fabio Giampaolo, Salvatore Cuomo, and Giancarlo Fortino. A survey on deep learning in medicine: Why, how and when? *Information Fusion*, 66:111–137, 2021.
- [6] Travers Ching, Daniel S Himmelstein, Brett K Beaulieu-Jones, Alexandr A Kalinin, Brian T Do, Gregory P Way, Enrico Ferrero, Paul-Michael Agapow, Michael Zietz, Michael M Hoffman, et al. Opportunities and obstacles for deep learning in biology and medicine. *Journal of The Royal Society Interface*, 15(141):20170387, 2018.
- [7] Muhammad Imran Razzak, Saeda Naz, and Ahmad Zaib. Deep learning for medical image processing: Overview, challenges and the future. *Classification in BioApps: Automation of Decision Making*, pages 323–350, 2018.
- [8] Qing Rao and Jelena Frtunijk. Deep learning for self-driving cars: Chances and challenges. In *Proceedings of the 1st international workshop on software engineering for AI in autonomous systems*, pages 35–38, 2018.
- [9] Jianjun Ni, Yinan Chen, Yan Chen, Jinxiu Zhu, Deena Ali, and Weidong Cao. A survey on theories and applications for self-driving cars based on deep learning methods. *Applied Sciences*, 10(8):2749, 2020.
- [10] Truong-Dong Do, Minh-Thien Duong, Quoc-Vu Dang, and My-Ha Le. Real-time self-driving car navigation using deep neural network. In *2018 4th International Conference on Green Technology and Sustainable Development (GTSD)*, pages 7–12. IEEE, 2018.
- [11] Zhenchao Ouyang, Jianwei Niu, Yu Liu, and Mohsen Guizani. Deep cnn-based real-time traffic light detector for self-driving vehicles. *IEEE transactions on Mobile Computing*, 19(2):300–313, 2019.
- [12] Andreas Kamilaris and Francesc X Prenafeta-Boldú. Deep learning in agriculture: A survey. *Computers and electronics in agriculture*, 147:70–90, 2018.
- [13] Nanyang Zhu, Xu Liu, Ziqian Liu, Kai Hu, Yingkuan Wang, Jinglu Tan, Min Huang, Qibing Zhu, Xunsheng Ji, Yongnian Jiang, et al. Deep learning for smart agriculture: Concepts, tools, applications, and opportunities. *International Journal of Agricultural and Biological Engineering*, 11(4):32–44, 2018.



- [14] Luís Santos, Filipe N Santos, Paulo Moura Oliveira, and Pranjali Shinde. Deep learning applications in agriculture: A short review. In *Robot 2019: Fourth Iberian Robotics Conference: Advances in Robotics, Volume 1*, pages 139–151. Springer, 2020.
- [15] Muhammad Hammad Saleem, Johan Potgieter, and Khalid Mahmood Arif. Automation in agriculture by machine and deep learning techniques: A review of recent developments. *Precision Agriculture*, 22:2053–2091, 2021.
- [16] Chandra Sekhar Nandi, Bipan Tudu, and Chiranjib Koley. An automated machine vision based system for fruit sorting and grading. In *2012 Sixth International Conference on Sensing Technology (ICST)*, pages 195–200. IEEE, 2012.
- [17] Rashmi Pandey, Sapan Naik, and Roma Marfatia. Image processing and machine learning for automated fruit grading system: A technical review. *International Journal of Computer Applications*, 81(16):29–39, 2013.
- [18] CS Nandi, B Tudu, and C Koley. Machine vision based techniques for automatic mango fruit sorting and grading based on maturity level and size. *Sensing technology: current status and future trends II*, pages 27–46, 2014.
- [19] Nguyen Minh Trieu and Nguyen Truong Thinh. Quality classification of dragon fruits based on external performance using a convolutional neural network. *Applied Sciences*, 11(22):10558, 2021.
- [20] Nguyen Minh Trieu and Nguyen Truong Thinh. A study of combining knn and ann for classifying dragon fruits automatically. *Journal of Image and Graphics*, 10(1):28–35, 2022.
- [21] Pallavi U Patil, Sudhir B Lande, Vinay J Nagalkar, Sonal B Nikam, and GC Wakchaure. Grading and sorting technique of dragon fruits using machine learning algorithms. *Journal of Agriculture and Food Research*, 4:100118, 2021.
- [22] Jialiang Zhou, Yueyue Zhang, and Jinpeng Wang. A dragon fruit picking detection method based on yolov7 and psp-ellipse. *Sensors*, 23(8):3803, 2023.
- [23] Dr T Vijayakumar and Mr R Vinothkanna. Mellowness detection of dragon fruit using deep learning strategy. *Journal of Innovative Image Processing*, 2(1):35–43, 2020.
- [24] Lutfi Hakim, Sepyan Purnama Kristanto, Dianni Yusuf, Mohammad Nur Shodiq, and Wahyu Ade Setiawan. Disease detection of dragon fruit stem based on the combined features of color and texture. *INTENSIF: Jurnal Ilmiah Penelitian dan Penerapan Teknologi Sistem Informasi*, 5(2):161–175, 2021.
- [25] David Gunning. Explainable artificial intelligence (xai). *Defense advanced research projects agency (DARPA), nd Web*, 2(2):1, 2017.
- [26] Wojciech Samek, Thomas Wiegand, and Klaus-Robert Müller. Explainable artificial intelligence: Understanding, visualizing and interpreting deep learning models. *arXiv preprint arXiv:1708.08296*, 2017.
- [27] Amitojdeep Singh, Sourya Sengupta, and Vasudevan Lakshmi-narayanan. Explainable deep learning models in medical image analysis. *Journal of imaging*, 6(6):52, 2020.
- [28] Kathryn Schutte, Olivier Moindrot, Paul Hérent, Jean-Baptiste Schiratti, and Simon Jégou. Using stylegan for visual interpretability of deep learning models on medical images. *arXiv preprint arXiv:2101.07563*, 2021.
- [29] Loveleen Gaur, Mohan Bhandari, Tanvi Razdan, Saurav Mallik, and Zhongming Zhao. Explanation-driven deep learning model for prediction of brain tumour status using mri image data. *Frontiers in genetics*, 13:448, 2022.
- [30] Luyi-Da Quach, Khang Nguyen Quoc, Anh Nguyen Quynh, Nghe Nguyen Thai, and Tri Gia Nguyen. Using gradient-weighted class activation mapping to explain deep learning models on agricultural dataset. *IEEE Access*, 2023.
- [31] Seyoung Ahn, Jeehyeong Kim, Soo Young Park, and Sunghyun Cho. Explaining deep learning-based traffic classification using a genetic algorithm. *IEEE Access*, 9:4738–4751, 2020.
- [32] Ramprasaath R Selvaraju, Michael Cogswell, Abhishek Das, Ramakrishna Vedantam, Devi Parikh, and Dhruv Batra. Grad-cam: Visual explanations from deep networks via gradient-based localization. In *Proceedings of the IEEE international conference on computer vision*, pages 618–626, 2017.
- [33] Ramprasaath R Selvaraju, Abhishek Das, Ramakrishna Vedantam, Michael Cogswell, Devi Parikh, and Dhruv Batra. Grad-cam: Why did you say that? *arXiv preprint arXiv:1611.07450*, 2016.
- [34] Harsh Panwar, PK Gupta, Mohammad Khubeb Siddiqui, Ruben Morales-Menendez, Prakhar Bhardwaj, and Vaishnavi Singh. A deep learning and grad-cam based color visualization approach for fast detection of covid-19 cases using chest x-ray and ct-scan images. *Chaos, Solitons & Fractals*, 140:110190, 2020.
- [35] Jost Tobias Springenberg, Alexey Dosovitskiy, Thomas Brox, and Martin Riedmiller. Striving for simplicity: The all convolutional net. *arXiv preprint arXiv:1412.6806*, 2014.
- [36] first. rp dataset. <https://universe.roboflow.com/first-zr3cq/rp-rombw>, apr 2023. visited on 2023-07-31.
- [37] Lisa Torrey and Jude Shavlik. Transfer learning. In *Handbook of research on machine learning applications and trends: algorithms, methods, and techniques*, pages 242–264. IGI global, 2010.
- [38] Sinno Jialin Pan and Qiang Yang. A survey on transfer learning. *IEEE Transactions on knowledge and data engineering*, 22(10):1345–1359, 2009.
- [39] Gao Huang, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q Weinberger. Densely connected convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4700–4708, 2017.
- [40] Rachel Lea Draelos and Lawrence Carin. Use hirescam instead of grad-cam for faithful explanations of convolutional neural networks. *arXiv preprint arXiv:2011.08891*, 2020.