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
[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. 1. Some images of dragon fruit dataset.](image1.png)

![Fig. 2. A dataset distribution.](image2.png)

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

\[
\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}
\]  \hspace{1cm} (1)

\[
\text{Precision} = \frac{TP}{TP + FP}
\]  \hspace{1cm} (2)

\[
\text{Recall} = \frac{TP}{TP + FN}
\]  \hspace{1cm} (3)

\[
F_1 - \text{Score} = \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}
\]  \hspace{1cm} (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.
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.
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>
<thead>
<tr>
<th>Class</th>
<th>Precision</th>
<th>Recall</th>
<th>F1-Score</th>
<th>Support</th>
</tr>
</thead>
<tbody>
<tr>
<td>Raw</td>
<td>0.96</td>
<td>0.97</td>
<td>0.97</td>
<td>268</td>
</tr>
<tr>
<td>Ripe</td>
<td>0.98</td>
<td>0.97</td>
<td>0.97</td>
<td>347</td>
</tr>
</tbody>
</table>

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. Specifi-
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


