RICE CONDITION CLASSIFICATION FROM UAV RGB IMAGERY USING SLIC SUPERPIXEL SEGMENTATION AND CONVNEXT

NHẬN DẠNG TÌNH TRẠNG CÂY LÚA TỪ ẢNH RGB UAV BẰNG PHÂN ĐOẠN SUPERPIXEL VÀ MẠNG CONVNEXT

Trong Hieu Luu, Quang Hieu Ngo*, Manh Dat Nguyen, Phuc Vinh Nguyen

College of Engineering, Can Tho University, Can Tho, Vietnam

*Corresponding author: nqhieu@ctu.edu.vn

(Received: March 03, 2026; Revised: June 04, 2026; Accepted: July 14, 2026)

DOI: 10.31130/ud-jst.2026.24(7A).131E

Abstract - Rice lodging is a common problem in paddy fields that reduces yield and complicates harvesting. This study presents a UAV-based framework for rice condition classification using RGB imagery. First, the Simple Linear Iterative Clustering (SLIC) algorithm is applied to segment each UAV image into compact superpixel regions, which help preserve major field boundaries and reduce pixel-level noise. Next, a ConvNeXt-Small network is used to classify rice conditions based on the segmented regions. A labeled dataset was manually constructed from UAV RGB images with three classes: green rice, lodged rice, and ripened rice. Data augmentation was further applied to improve sample diversity. Experimental results show that the proposed SLIC–ConvNeXt pipeline can effectively distinguish the three rice conditions, achieving an overall accuracy of 98.07% on the validation set of the current dataset. These results indicate that superpixel-guided deep learning is a promising approach for UAV-based rice condition assessment.

Key words - UAV imagery; rice lodging; Superpixel segmentation; Simple Linear Iterative Clustering (SLIC); ConvNeXt deep learning.

Tóm tắt - Đổ ngã lúa là vấn đề phổ biến, làm giảm năng suất và gây khó khăn cho thu hoạch. Nghiên cứu này trình bày phương pháp sử dụng UAV để phân loại tình trạng lúa từ ảnh RGB. Trước hết, thuật toán Simple Linear Iterative Clustering (SLIC) phân đoạn ảnh UAV thành các vùng superpixel nhỏ gọn, giúp bảo toàn ranh giới chính trên đồng ruộng và giảm nhiễu điểm ảnh. Tiếp theo, mạng ConvNeXt-Small phân loại tình trạng lúa dựa trên các vùng đã được phân đoạn. Một bộ dữ liệu gán nhãn thủ công được xây dựng từ ảnh RGB thu nhận bằng UAV với ba lớp: lúa xanh, lúa đổ ngã và lúa chín. Kỹ thuật tăng cường dữ liệu được áp dụng để nâng cao tính đa dạng của mẫu. Kết quả thực nghiệm cho thấy quy trình SLIC–ConvNeXt phân biệt hiệu quả ba tình trạng lúa, đạt độ chính xác tổng thể 98,07% trên tập xác thực của bộ dữ liệu này. Kết quả này cho thấy học sâu có hướng dẫn bởi superpixel là hướng tiếp cận đầy hứa hẹn cho đánh giá tình trạng lúa từ ảnh UAV.

Từ khóa - Lúa đỗ ngã; không ảnh UAV; phân ngưỡng superpixel; Simple Linear Iterative Clustering (SLIC); mạng học sâu ConvNeXt

1. Introduction

Rice is one of the most important staple crops in the world and plays a crucial role in ensuring global food security. According to global agricultural statistics, rice ranks among the three major food crops together with wheat and maize, and Vietnam is currently one of the leading rice-exporting countries in the world. However, rice production is often affected by several environmental and agronomic factors that can significantly reduce yield and grain quality. Rice lodging, which refers to the bending or falling of rice stems caused by wind, rain, or excessive plant growth, is one of the most common and damaging problems in paddy fields, as it can hinder harvesting operations, increase grain loss, and ultimately reduce overall rice productivity.

Conventional field surveys for rice lodging mainly depend on visual inspection and farmer interviews to estimate lodging severity and its spatial distribution [1]. Recently, lodging detection has shifted toward UAV-based remote sensing, where high-resolution RGB images are captured by unmanned aerial vehicles and then processed using deep learning models to automatically map lodged areas in paddy fields. This approach is attractive because it is cost-effective, can cover large areas, provide fine spatial detail, and allow flexible data collection times depending on field and weather conditions [2]–[4]. When integrated with machine learning, UAV imagery has been widely studied in precision agriculture, including crop monitoring [5], crop type classification [6], yield prediction [7], [8], and plant disease detection [9].

For rice lodging specifically, the study in [10] reported that RGB imagery can outperform multispectral imagery when combined with a U-Net segmentation network, achieving Dice coefficients of 0.9442 and 0.9284 for RGB and multispectral inputs, respectively. In [11], a UAV system equipped with multiple cameras was used to estimate mature rice yield using a Fusion Yield Network (FusYieldNet) together with a five-point sampling strategy; the proposed method also enabled lodging rate estimation and reported yield errors within 3%. The authors of [12] investigated the relationship between rice lodging and spatial distributions of soil-and-plant nitrogen (SAN) using UAS images in a Koshihikari paddy field with one-shot basal fertilization, and they observed a negative correlation between the lodging inclination angle and the estimated nitrogen uptake (SAN). Similarly, study [13] proposed the LodgeNet deep learning network to detect and evaluate rice lodging from

UAV imagery, achieving an accuracy of 97.3% on the experimental model proposed.

This study presents a UAV-based framework for rice condition classification from RGB imagery by integrating Simple Linear Iterative Clustering (SLIC) superpixel segmentation with a ConvNeXt-based deep learning classifier. The purpose is to examine whether superpixel-based region construction can provide a more stable representation for distinguishing three rice conditions, namely lodged rice, green rice, and ripened rice. By grouping pixels into compact local regions while preserving important field boundaries, the proposed framework offers a practical and application-oriented approach for UAV-based rice field analysis.

2. Methods

2.1. Device and data acquisitions

Data acquisition was conducted on March 13, 2022, in Long My Town, Hau Giang Province (Long My Town, Can Tho City since July 2025). The data was collected at approximately 9:00 AM using a DJI Phantom 4 Pro. During this time, weather conditions were favorable with light winds and no clouds, making the environment highly suitable for UAV data collection.

The UAV was equipped with a 20 MP RGB camera to capture high-quality aerial imagery. Following the recommendations of the FieldAgent software, the flight path was set at an altitude of 20 meters with a velocity of 4 m/s. A total of 40 aerial images were collected. Field observations indicate that the rice has reached maturity and is ready for harvest; however, widespread lodging is present, which significantly impacts both the final yield and the harvesting methods (Figure 1).

Figure 1. *Data acquisition area*

2.2. Superpixel segmentation using SLIC

2.2.1. Image pre-processing

Let $I \in R^{H\times W\times 3}$ denote an RGB image size, which is first loaded and converted to CIELAB color space to ensure consistent channel ordering. To reduce the influence of high-frequency noise and compression artifacts, a mild Gaussian smoothing is applied within the SLIC algorithm [14] using the parameter σ and then normalized before performing superpixel segmentation. For high-resolution inputs, the images were resized such that

$$\max(H, W) \le 1600 \quad (1)$$

We use SLIC to partition the image into approximately K superpixels. SLIC performs clustering in a 5D space formed by color and spatial coordinates. For each pixel p, the feature vector performs as

$$f(p) = [L(p), a(b), b(p), x(p), y(p)] \quad (2)$$

Where (L, a, b) are color values in CIELAB color space and (x, y) are pixel coordinates. SLIC iteratively assigns pixels to cluster centers by minimizing a distance that balances color similarity and spatial proximity

$$D = \sqrt{d_c^2 + \left(\frac{m}{S}\right)^2 d_s^2} \quad (3)$$

where, d_c is the color distance, d_s is the spatial distance, S is the approximate grid interval between superpixel centers, and m is a compactness parameter.

SLIC is applied to generate approximately K superpixels. The compactness parameter controls the trade-off between following image boundaries and producing regular regions, and σ determines the amount of Gaussian smoothing used to reduce noise. In this paper, parameters are set at $K = 400, m = 20, \sigma = 2$. Output is a label map with

$$S \epsilon \{1, 2, \ldots, K'\}^{H\times W} \quad (4)$$

Where $S(x, y)$ is the superpixel index assigned to pixel (x, y) and K' is the actual number regions produced.

2.2.2. Region construction and feature extraction

Each superpixel r corresponds to the set of the pixels is conducted as

$$\Omega_r = \{(x, y) \mid S(x, y) = r\} \quad (5)$$

and the mean-color map $\overline{I_r}$ where each region is filled by its average RGB value is calculated as

$$\overline{I_r} = \frac{1}{|\Omega_r|} \sum_{(x,y)\in\Omega_r} I(x, y) \quad (6)$$

In addition, to support texture analysis, the RGB image is converted to a grayscale intensity image I_{gray} using a luminance transform

$$I_{gray} = 0.299R + 0.587G + 0.114B \quad (7)$$

Because texture can appear at different scales, we use simple and stable descriptors that are less sensitive to illumination changes:

(1) Intensity statistics

The mean intensity μ_r which represents the average brightness of the region, and the standard deviation σ_r, which measures intensity variation within the region, are computed as follows:

$$\mu_r = \sqrt{\frac{1}{|\Omega_r|} \sum_{(x,y)\in\Omega_r} I_{gray}(x, y)} \quad (8)$$

$$\sigma_r = \sqrt{\frac{1}{|\Omega_r|} \sum_{(x,y)\in\Omega_r} \left(I_{gray}(x, y) - \mu_r\right)^2} \quad (9)$$

In addition, three intensities are calculated including $P_{10,r}$, $P_{50,r}$, and $P_{90,r}$. Each percentile $P_{q,r}$ is defined as the intensity value such that $q\%$ of pixels in Ω_r have intensity less than or equal to:

$$P_{q,r} = \inf\left\{t \middle| \frac{1}{|\Omega_r|}\sum_{(x,y)\in\Omega_r}(I_{gray}(x,y) \le t) \ge \frac{q}{100}\right\} \quad q \in \{10, 50, 90\} \tag{10}$$

(2) Gradient-based texture

A gradient magnitude image M is computed from the grayscale image I_{gray} to describe local edge strength. Spatial derivatives $\partial_x I_{gray}$, $\partial_y I_{gray}$ are estimated by using Sobel filters and the gradient magnitude at pixel (x, y) is calculated as:

$$M(x,y) = \sqrt{\left(\partial_x I_{gray}(x,y)\right)^2 + \left(\partial_y I_{gray}(x,y)\right)^2} \tag{11}$$

For each superpixel region Ω_r, two summary measures are computed from $M(x, y)$. The mean gradient magnitude $\overline{M_r}$ and the standard deviation of gradient $\sigma(M)_r$ are computed as:

$$\overline{M_r} = \frac{1}{|\Omega_r|}\sum_{(x,y)\in\Omega_r} M(x,y) \tag{12}$$

$$\sigma(M)_r = \sqrt{\frac{1}{|\Omega_r|}\sum_{(x,y)\in\Omega_r}(M(x,y) - \overline{M_r})^2}$$

These measures are used to reflect local roughness and edge density within each region.

(3) Local binary pattern (LBP)

Local Binary Pattern (LBP) features are computed to describe micro-texture by comparing each pixel with its neighbors. For each region Ω_r, an LBP histogram $h_r \in R^B$ is calculated and normalized, where B denotes the number of histogram bins. The normalized histogram is used as a compact texture descriptor for the region.

Finally, all region-level features are concatenated into one vector:

$$x_r = [\mu_r, \sigma_r, P_{10,r}, P_{50,r}, P_{90,r}, \overline{M_r}, \sigma(M)_r, h_r] \tag{13}$$

The final dataset is formed as a set of region-level samples $\{\mathbf{x}_r\}_{r=1}^{K'}$, where $K'$is the number of superpixels. These feature vectors can be used for classification in ConvNet in next step of this research.

2.3. Dataset generation and ConvNeXt-Based Classification

A total of 40 UAV RGB images were collected in this study. Among them, 12 images were selected to construct the labeled patch dataset for model development. These images were chosen to contain representative rice-canopy surface conditions only, and regions containing infrastructure or irrelevant background objects were excluded. Three classes were defined: Lodged rice, Green rice, and Ripened rice ($C = 3$). Image patches $X_i \in \mathbb{R}^{H_p \times W_p \times 3}$were selected and labeled by visual inspection, forming the dataset

$$\mathcal{D} = \{(X_i, y_i)\}_{i=1}^{N}, y_i \in \{1, \dots, C\} \tag{14}$$

To increase sample diversity, data augmentation was applied using Gaussian blurring and rotation at 90°, 180°, and 270°. The dataset was randomly split into training/validation sets with a 90%/10% ratio. Before training, each patch was resized to a fixed input size at 224×224 pixels and normalized.

For deep learning classification, ConvNeXt-Small, proposed by Liu [15], was adopted with ImageNet-pretrained weights. The backbone is composed of a stem followed by four stages of ConvNeXt blocks and channel dimensions (96, 192, 384, 768), with downsampling between stages (Figure 2). After the last stage, the final feature map is aggregated by global average pooling and passed to a linear layer to produce logits $\mathbf{z}_i \in \mathbb{R}^C$:

$$\mathbf{z}_i = f_\theta(X_i) \tag{15}$$

Class probabilities are obtained using softmax,

$$\hat{p}_{i,c} = \frac{e^{z_{i,c}}}{\sum_{k=1}^{C} e^{z_{i,k}}}, c = 1, \dots, C \tag{16}$$

and the predicted label is selected as

$$\hat{y}_i = \arg\max_c \hat{p}_{i,c} \tag{17}$$

Training is performed using the multi-class cross-entropy loss. Let $y_{i,c} \in \{0,1\}$denote the one-hot label for class c. The loss is defined as

$$\mathcal{L} = -\frac{1}{N}\sum_{i=1}^{N}\sum_{c=1}^{C} y_{i,c}\log(\hat{p}_{i,c}) \tag{18}$$

To evaluate the model on unseen data, validation was performed using UAV images outside the 12 images used for dataset construction. Therefore, the validation images were completely independent from the images used for sample generation and model training. This design reduces the risk of information leakage and provides a more realistic assessment of classification performance on unseen field images.

3. Results

3.1. Image segementation using SLIC

The SLIC results in Figure 3 indicate that the UAV rice-field image can be partitioned into meaningful local regions while preserving major field structures. In the original images, clear visual transitions are observed between neighboring plots, including dark plot borders and narrow green strips (e.g., vegetation or a ditch), together with gradual changes in canopy color and texture across the field. After applying SLIC, the scene is decomposed into many compact superpixels. In visually homogeneous areas of the rice canopy, superpixels are formed with relatively regular shapes and similar sizes, which suggests that local texture within each region is consistent and that region-based statistics can be computed reliably. Near strong boundaries, the superpixel contours become more irregular and are observed to follow real edges in the scene. Plot borders and green strips are separated from surrounding rice regions, and boundary locations are traced by multiple superpixels, which reduces mixing between different surface types. The zoomed views further show that smaller regions are created around edge-dense or highly textured areas, while larger regions are maintained in smoother canopy areas. Overall, these results support the use of SLIC as a region proposal step, because spatial complexity is reduced by grouping similar pixels while important structures are still retained for subsequent feature extraction and classification. In this study, SLIC was performed with $K = 400$, compactness $m = 20$, and smoothing $\sigma = 2$ to produce the superpixel maps.

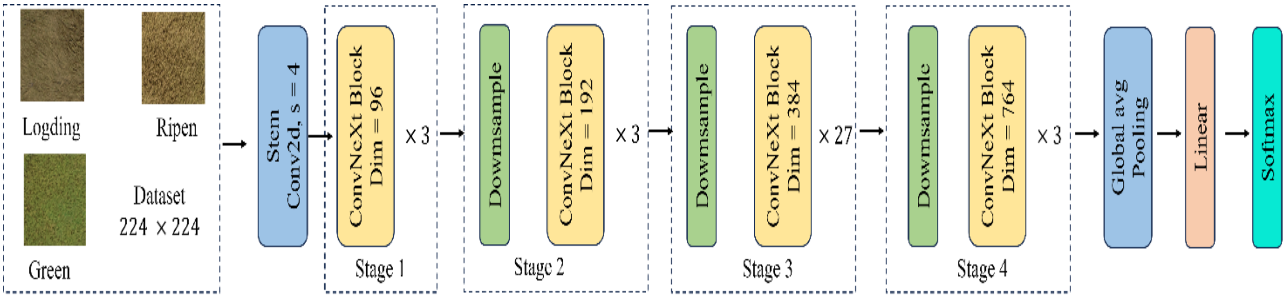

Figure 2. *The ConvNeXt-Small architecture*

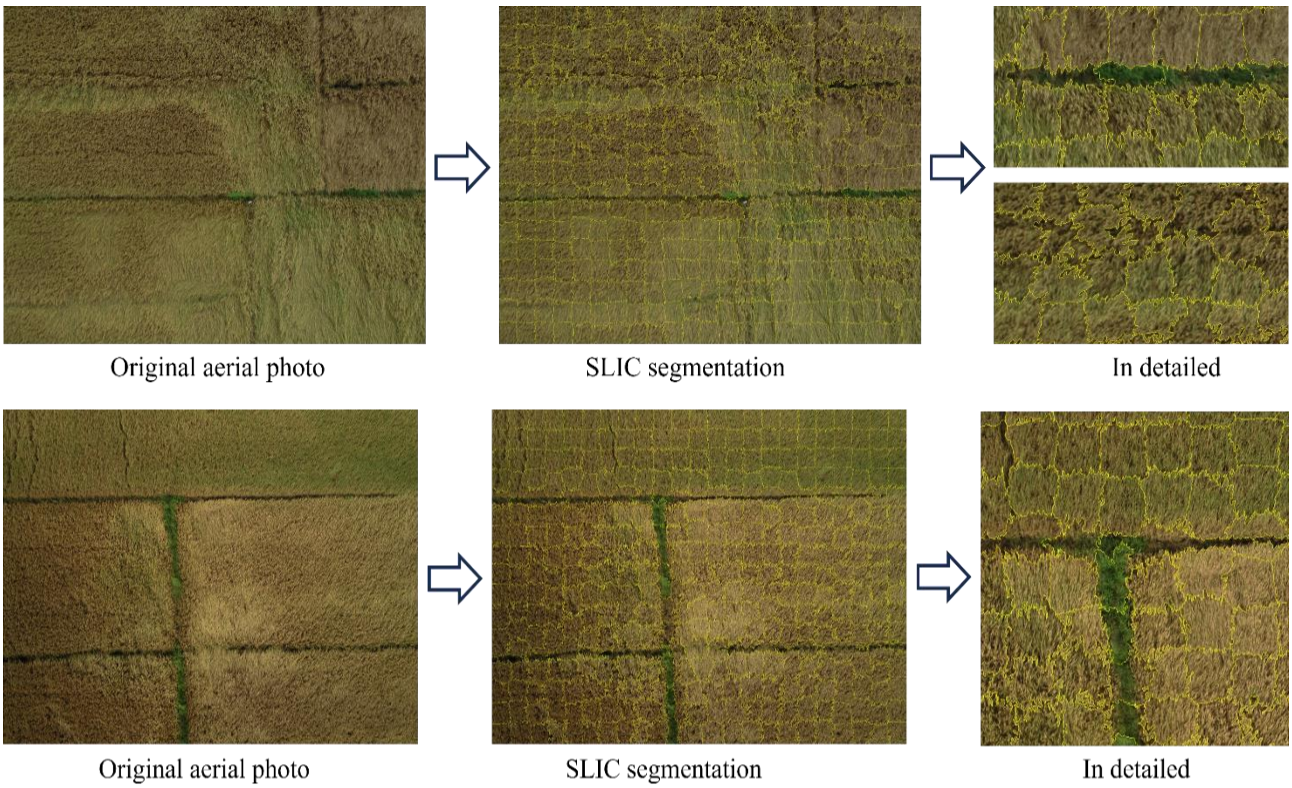

Figure 3. *The SLIC segmentation*

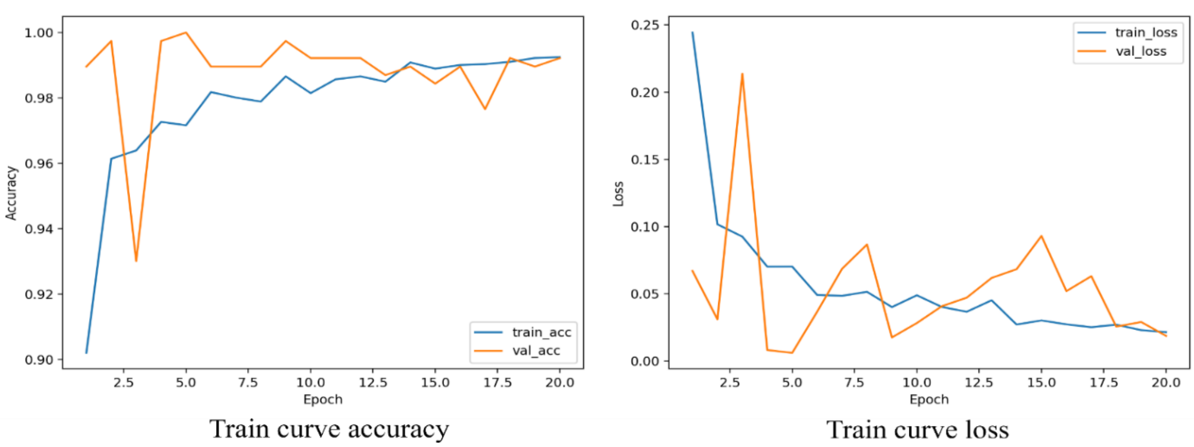

Figure 4. *The ConvNeXt-based model performance*

3.2. ConvNeXt-Based performance and classification of rice field conditions

The model performance is shown in Figure 4. The training curves in Figure 4 show that the model was trained for 20 epochs and reached a stable performance early. Training accuracy increased rapidly from about 0.90 at the first epoch to above 0.96 within the first few epochs, and then slowly improved to around 0.99 toward the end of training. In the same period, the training loss decreased sharply (from roughly 0.24 to below 0.10 in the first epochs) and continued to decline gradually, reaching a small value (about 0.02) at the final epochs. This behavior indicates that most learning occurred in the early stage, while later epochs provided only minor refinement.

Validation accuracy stayed very high (mostly around 0.99–1.00) but showed a noticeable drop at approximately epoch 3, and the validation loss also exhibited occasional spikes. Such fluctuations are consistent with a small validation set (only 10% of the data), where a few difficult samples can strongly affect the validation metrics. Importantly, no clear overfitting trend is observed because training and validation accuracy remain close and both losses stay low. Overall, the model can be considered converged around epochs 8–12, after which the curves become relatively stable and additional epochs yield limited improvement. The class-wise classification results are summarized in Table 1. All three rice classes achieved F1-scores above 0.97, indicating consistently high classification performance. Green_rice obtained the highest precision of 1.0000, whereas Ripened_rice achieved the highest recall of 1.0000.

Table 1. *Precision, Recall, F1-score, and class accuracy for each rice class*

Class	Precision	Recall	F1-score	Class Acc
Green_rice	1.0000	0.9650	0.9822	0.9862
Lodge_rice	0.9727	0.9817	0.9772	0.9862
Ripened_rice	0.9652	1.0000	0.9823	0.9890

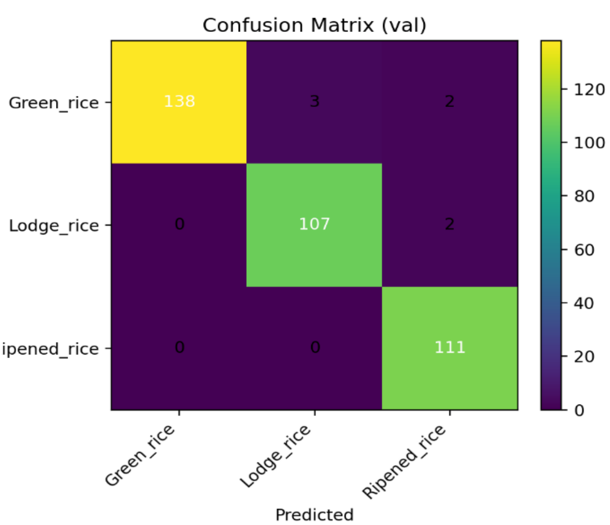

Figure 5. *The confusion matrix*

Figure 5 presents the confusion matrix on the validation set of 363 samples. The results show strong diagonal dominance, indicating that most samples are correctly classified across the three classes. A total of 356 samples are predicted correctly, yielding an overall accuracy of 98.07%. Minor misclassifications are mainly observed between Green_rice, Lodge_rice, and Ripened_rice in a small number of visually similar cases. Among the three classes, Ripened_rice achieves the most consistent performance, with no false negatives observed in the current validation set. Overall, the confusion matrix suggests that the proposed model can effectively distinguish the three rice conditions, while only limited confusion remains in a few boundary cases. Representative validation predictions are provided in Figure 6.

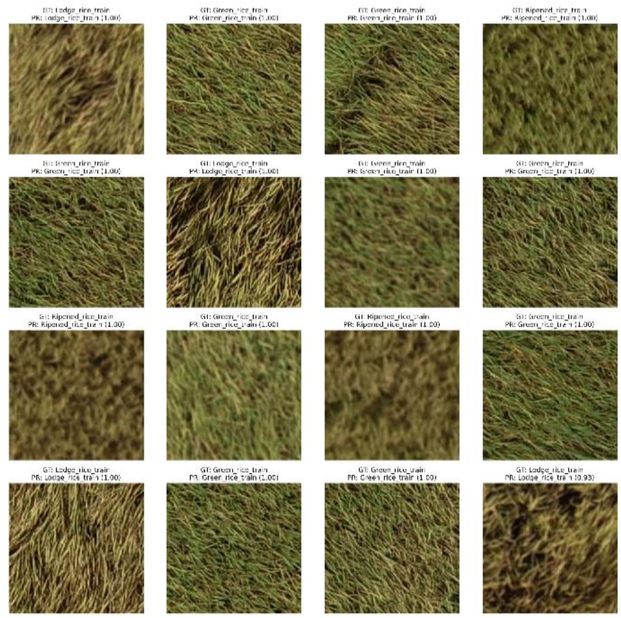

Figure 6. *Validation sample prediction*

Figure 7 presents the rice condition classification results from UAV RGB imagery using the proposed pipeline. The SLIC boundary maps show that the scene is divided into compact superpixel regions that generally follow major field structures, such as plot borders and vegetation strips, which provides a more stable region representation for the classification stage. Based on these regions, the ConvNeXt network demonstrates the ability to classify rice conditions; however, some local misclassifications are still observed because the rice canopy structure on the field surface is spatially complex and highly variable. The results also indicate that green rice is often detected as relatively large and consistent areas, suggesting that this class has clearer visual cues in RGB images. In contrast, ripened rice is more difficult to identify because its color and texture can change across the field and may overlap with lodging patterns under different illumination and canopy conditions. Therefore, the ripened class tends to appear more fragmented and is occasionally mixed with lodged rice, especially near transition zones, implying that finer labels and more diverse training samples are needed to produce a more stable ripened-rice map.

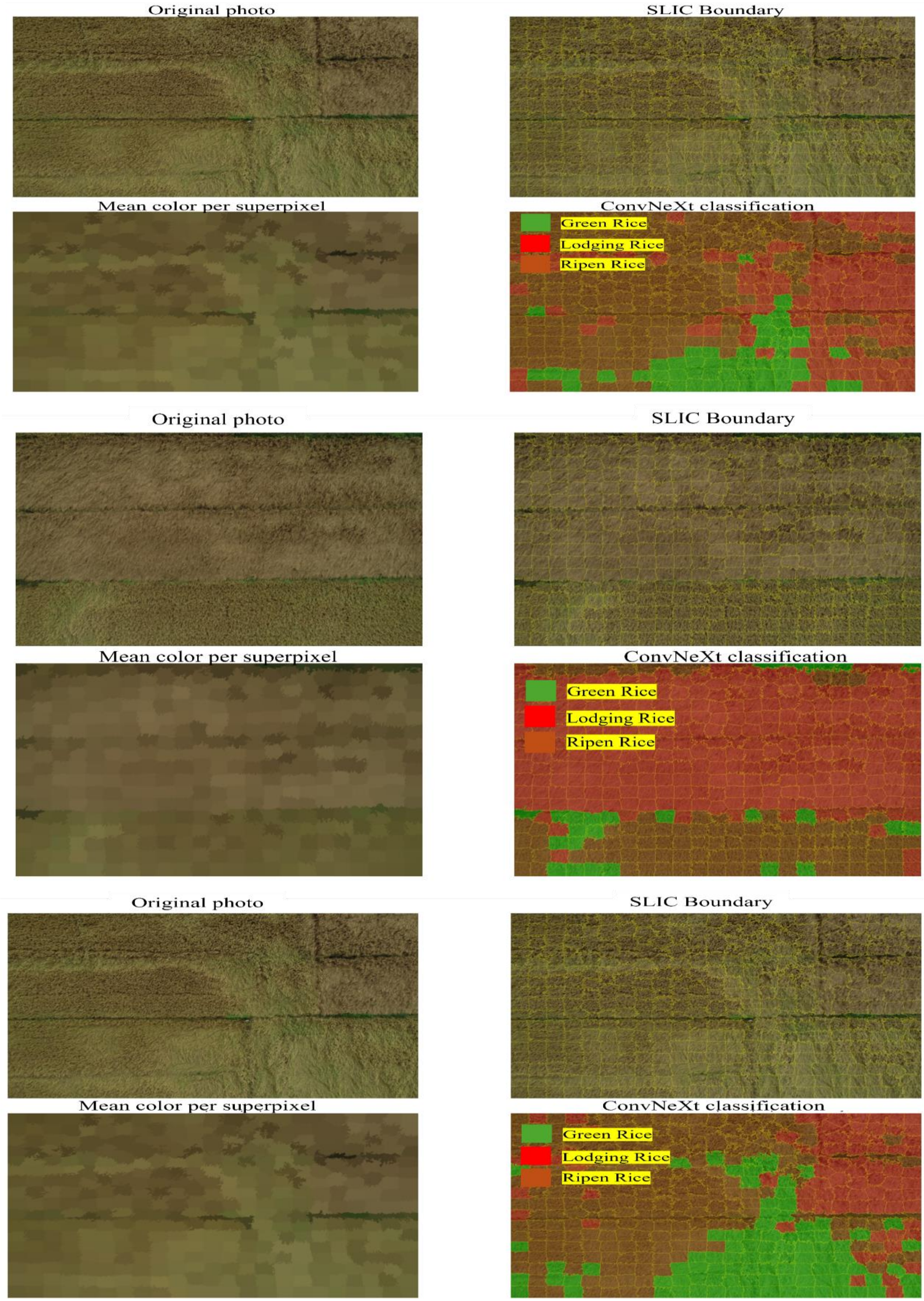

Figure 7. *Visual results of the proposed SLIC–ConvNeXt pipeline on UAV RGB rice-field images: (a) Original UAV image, (b) SLIC superpixel boundaries, (c) mean color representation per superpixel, and (d) ConvNeXt-based classification map*

3.3. Discussion

In various state-of art, lodging detection has been widely investigated using remote sensing data because lodging changes crop canopy structure, reflectance, and texture. The authors of [13] proposed the MI-UConvNeXt model for rice lodging detection based on semantic segmentation and intelligent optimization. Their model achieved a pixel accuracy (PA) of 95.59%, a mean pixel accuracy (mPA) of 95.62%, and a mean intersection over union (mIoU) of 91.91%. These results show that an optimized deep learning architecture can improve segmentation accuracy and reduce computational complexity. In another rice lodging study, the authors of [16] used UAV multispectral imagery together with spectral features, vegetation indices, colour features, and texture features for plot-level lodging assessment. Their Random Forest classifier achieved an overall accuracy of 96.1% and a Kappa coefficient of 0.92. The estimated lodged area also showed strong agreement with visual interpretation, with an R^2 value of 0.97. Chauhan et al. [17] showed that UAV multispectral imagery could be used to assess wheat lodging severity. In their study, red-edge and NIR bands were useful for distinguishing non-lodged and lodged wheat, and the classification achieved an overall accuracy of 90%. At a larger spatial scale, the authors of [18] used multi-temporal Sentinel-1 SAR images for maize lodging monitoring. They found that VH time-series features were sensitive to lodging severity. Their TwDTW-based method achieved an accuracy of 76% and a Kappa coefficient of 0.70, showing the usefulness of temporal SAR features for regional lodging assessment.

The main result of the present study shows that the proposed SLIC–ConvNeXt pipeline can classify rice-field conditions from UAV RGB images into three groups: green rice, lodged rice, and ripened rice. In this method, SLIC superpixel segmentation is applied before ConvNeXt classification to reduce pixel-level noise, preserve important field boundaries, and create stable image regions for classification. The proposed model achieved an overall validation accuracy of 98.07%. This result suggests that low-cost RGB UAV imagery can be used as a practical tool for rapid rice-condition assessment in small paddy fields. A brief comparison between previous lodging detection studies and the present study is shown in Table 2.

Table 2. *Comparison of recent studies on crop lodging detection and the present study*

Study	Method used	Main finding
Zhang et al. [13]	UAV remote sensing images with MI-UConvNeXt semantic segmentation and intelligent optimization	The proposed model achieved high rice lodging segmentation performance, with PA = 95.59%, mPA = 95.62%, and mIoU = 91.91%. The results showed that optimized deep learning can improve segmentation accuracy and computational efficiency.
Kumar et al. [16]	UAV multispectral imagery with spectral features, vegetation indices, colour features, texture features, and Random Forest classifier	The method detected rice lodging at the plot level with an overall accuracy of 96.1% and a Kappa coefficient of 0.92. The estimated lodged area agreed well with visual interpretation, with $R^2 = 0.97$.
Chauhan et al. [17]	High-resolution UAV multispectral imagery and object-based classification	Red-edge and NIR bands were useful for distinguishing non-lodged and lodged wheat. The method achieved an overall accuracy of 90% for lodging severity assessment.
Qu et al. [18]	Multi-temporal Sentinel-1 SAR images with VH time-series features and TwDTW classification	VH time-series features were sensitive to maize lodging severity. The TwDTW-based method achieved an accuracy of 76% and a Kappa coefficient of 0.70 for county-scale lodging monitoring.
Present study	UAV RGB imagery with SLIC superpixel segmentation and ConvNeXt-Small classification	The proposed SLIC–ConvNeXt pipeline classified rice-field conditions into green rice, lodged rice, and ripened rice. The model achieved an overall validation accuracy of 98.07%, showing the potential of low-cost RGB UAV imagery for rapid rice-condition assessment in small paddy fields.

4. Conclusion

This study presented a two-stage framework for rice condition classification from UAV RGB imagery. In the first stage, SLIC superpixel segmentation was applied to divide each image into compact local regions while preserving major field boundaries. In the second stage, a ConvNeXt-Small network was used to classify the segmented regions into three classes: green rice, lodged rice, and ripened rice. Experimental results on the constructed dataset showed that the proposed pipeline achieved high classification performance, with an overall accuracy of 98.07% on the validation set.

Despite these promising results, several limitations remain. The dataset is still limited in size and was collected under a specific field condition, which may restrict the generalizability of the current model. In addition, the present study focuses on three broad rice-condition classes and does not yet address lodging severity levels or field-scale quantitative estimation. Future work will therefore concentrate on expanding the dataset, improving annotation detail, and developing a full-field mapping framework for more practical lodging assessment.

Funding: "This research is funded by Vietnam National Foundation for Science and Technology Development (NAFOSTED) under grant number 53/2025/TN".

REFERENCES

[1] H. P. Government, "*Chapter 4. Research Method of Plant Growth and Harvest,"* in Principle of Guidance for Rice, Wheat and Soybean. Kobe, Japan: Hyogo Prefecture, 2017, pp. 129–140.

[2] H. Xiang and L. Tian, "Development of a low-cost agricultural remote sensing system based on an autonomous unmanned aerial vehicle (UAV)," *Biosyst. Eng*., vol. 108, pp. 174–190, 2011.Tài liệu tham khảo 2.

[3] M. Weiss, F. Jacob, and G. Duveiller, "Remote sensing for agricultural applications: A meta-review," *Remote Sens. Environ.*, vol. 236, Art. no. 111402, 2020.

[4] H. Yao, R. Qin, and X. Chen, "Unmanned aerial vehicle for remote sensing applications-A review," *Remote Sens.*, vol. 11, no. 12, Art. no. 1443, 2019.

[5] T. B. Shahi, C.-Y. Xu, A. Neupane, and W. Guo, "Machine learning methods for precision agriculture with UAV imagery: A review," *Electron. Res. Arch.*, vol. 30, pp. 4277–4317, 2022.

[6] H. Tao *et al.*, "Estimation of crop growth parameters using UAV-based hyperspectral remote sensing data," *Sensors*, vol. 20, no. 5, Art. no. 1296, 2020.

[7] T. Lu, L. Wan, S. Qi, and M. Gao, "Land cover classification of UAV remote sensing based on transformer–CNN hybrid architecture," *Sensors*, vol. 23, no. 12, Art. no. 5288, 2023.

[8] M. S. Mia *et al.*, "Multimodal deep learning for rice yield prediction using UAV-based multispectral imagery and weather data," *Remote Sens.*, vol. 15, Art. no. 2511, 2023.

[9] P. Ahmadi, S. Mansor, B. Farjad, and E. Ghaderpour, "Unmanned aerial vehicle (UAV)-based remote sensing for early-stage detection of Ganoderma," *Remote Sens.*, vol. 14, Art. no. 1239, 2022.

[10] Z. Su, Y. Wang, Q. Xu, R. Gao, and Q. Kong, "LodgeNet: Improved rice lodging recognition using semantic segmentation of UAV high-resolution remote sensing images," *Comput. Electron. Agric.*, vol. 196, Art. no. 106873, 2022, doi: 10.1016/j.compag.2022.106873.

[11] Y. Chen, X. Yang, X. Zhou, Y. Deng, D. Kong, Z. Liu, and X. Zhu, "Estimating the yield of mature rice with UAV-based multimodal images fusion and five-points sampling," *IEEE Access*, vol. 14, pp. 25115–25129, 2026, doi: 10.1109/ACCESS.2026.3663405.

[12] N. K. Sato, T. Tsuji, Y. Iijima, N. Sekiya, and K. Watanabe, "Predicting rice lodging risk from the distribution of available nitrogen in soil using UAS images in a paddy field," *Sensors*, vol. 23, no. 14, Art. no. 6466, 2023, doi: 10.3390/s23146466.

[13] P. Zhang, S. Zhang, J. Wang, and X. Sun, "Identifying rice lodging based on semantic segmentation architecture optimization with UAV remote sensing imaging," *Computers and Electronics in Agriculture*, vol. 227, pt. 1, Art. no. 109570, 2024, doi: 10.1016/j.compag.2024.109570.

[14] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Süsstrunk, "SLIC superpixels compared to state-of-the-art superpixel methods," *IEEE Trans. Pattern Anal. Mach. Intell*., vol. 34, no. 11, pp. 2274–2282, 2012, doi: 10.1109/TPAMI.2012.120.

[15] Z. Liu, H. Mao, C.-Y. Wu, C. Feichtenhofer, T. Darrell, and S. Xie, "A ConvNet for the 2020s," *arXiv preprint arXiv:2201.03545*, 2022. [Online]. Available: https://arxiv.org/abs/2201.03545

[16] M. Kumar, B. K. Bhattacharya, M. R. Pandya, and B. K. Handique, "Machine learning based plot level rice lodging assessment using multi-spectral UAV remote sensing," *Computers and Electronics in Agriculture,* vol. 219, Art. no. 108754, 2024, doi: 10.1016/j.compag.2024.108754.

[17] S. Chauhan *et al.*, "Wheat lodging assessment using multispectral UAV data," *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. XLII-2/W13, pp. 235–240, 2019, doi: 10.5194/isprs-archives-XLII-2-W13-235-2019.

[18] X. Qu, J. Zhou, X. Gu, Y. Wang, Q. Sun, and Y. Pan, "Monitoring maize lodging severity based on multi-temporal Sentinel-1 images using Time-weighted Dynamic Time Warping," *Computers and Electronics in Agriculture*, vol. 215, Art. no. 108365, 2023, doi: 10.1016/j.compag.2023.108365.