UAV Hyperspectral Imaging and Transformer-Based Semantic Segmentation for Multi-Class Wheat Disease Stress Detection in Precision Agriculture

Yinyan Shi[1], Qiang Chen[1], Chuang Xia[1], Xiaochan Wang[1,*], Xuekai Huang[1], Kenji Tanaka[2], Hiroshi Yamamoto[2]

[1] College of Engineering, Nanjing Agricultural University, Nanjing 210031, Jiangsu, China

[2] Graduate School of Agriculture, Kyoto University, Kyoto 606-8502, Japan

* Corresponding author: xcwang@njau.edu.cn

Abstract

Wheat diseases and nutrient stress represent critical threats to global food security, causing annual yield losses estimated at 10-28% in major producing regions. Timely and accurate spatial mapping of stress distribution is essential for precision intervention, yet conventional scouting methods are labor-intensive, subjective, and unable to capture fine-grained spatial heterogeneity at the field scale. This paper proposes a novel end-to-end framework integrating Unmanned Aerial Vehicle (UAV) hyperspectral imaging with a transformer-based semantic segmentation model, SegFormer-B4, for simultaneous detection and spatial mapping of four wheat stress categories: healthy canopy, stripe rust (Puccinia striiformis), powdery mildew (Blumeria graminis), and nitrogen deficiency. Hyperspectral imagery across 128 spectral bands (400-1000 nm) was acquired using a DJI M300 RTK UAV equipped with a Specim AFX10 pushbroom sensor over winter wheat fields in Jiangsu and Zhejiang provinces during the heading-to-filling growth stages. A dataset of 4,680 annotated image patches (256x256 pixels) was constructed through systematic sampling and multi-strategy data augmentation. The Mix Transformer (MiT-B4) encoder, pre-trained on ImageNet-22K and fine-tuned on the wheat hyperspectral dataset, captures multi-scale spatial-spectral features through hierarchical overlapping patch embeddings and efficient self-attention. Comparative evaluation against six baseline architectures (FCN-8s, U-Net with VGG-16, DeepLabv3+ with MobileNetV2, PSPNet with ResNet-50, Swin-T UperNet, and SegFormer-B2) demonstrates that SegFormer-B4 achieves a mean Intersection over Union (MIoU) of 92.8%, mean Pixel Accuracy (MPA) of 95.6%, Precision of 94.9%, and Recall of 94.6%, representing improvements of 3.9-20.4 percentage points on MIoU over baselines. Disease area estimation on 12 independent field plots yields a maximum relative error below 2%, confirming strong practical applicability. Ablation analysis reveals that spectral band selection and multi-scale feature fusion collectively contribute 6.5 MIoU points over the base encoder, underscoring the critical role of hyperspectral feature exploitation in agricultural stress detection. The proposed framework provides a scalable, data-driven foundation for early warning systems and site-specific crop management.

Keywords: wheat disease detection; UAV hyperspectral imaging; transformer segmentation; SegFormer; precision agriculture; semantic segmentation; stripe rust; powdery mildew

1. Introduction

Global wheat production faces persistent threats from biotic and abiotic stresses that substantially diminish yield potential and grain quality [1,2]. Among biotic stressors, stripe rust caused by Puccinia striiformis f. sp. tritici and

powdery mildew caused by Blumeria graminis f. sp. tritici collectively account for estimated annual losses exceeding 10 million tonnes in Asia alone [3,4]. Abiotic nitrogen deficiency, exacerbated by non-uniform fertilizer application and soil variability, further reduces grain protein content and photosynthetic efficiency [5,6]. Early and accurate spatial characterization of these co-occurring stresses at the field scale is therefore essential for timely intervention, targeted fungicide application, and adaptive nitrogen management that minimize environmental loading while sustaining productivity [7,8].

Traditional disease monitoring relies on visual inspection by trained agronomists walking transects through fields, a method inherently constrained by scale, subjectivity, and timeliness [9,10]. Ground-based proximal sensing instruments---including portable spectrometers and hand-held multispectral cameras---offer improved objectivity but cannot efficiently survey the large spatial extents typical of commercial wheat production [11,12]. Satellite remote sensing provides synoptic coverage but at spatial resolutions (10-30 m per pixel) that are inadequate for detecting early-stage, spatially heterogeneous disease patches and for distinguishing spectrally similar stress types [13,14].

Unmanned Aerial Vehicle (UAV)-based remote sensing has emerged as a transformative platform for precision crop monitoring, combining centimeter-scale spatial resolution, flexible temporal revisit frequency, and multi-sensor payload capability [15,16]. Multispectral UAV systems with 4-10 bands have been widely applied to vegetation index computation for broad stress detection [17,18]. However, the limited spectral dimensionality of multispectral sensors imposes fundamental constraints on inter-disease discriminability: stripe rust and powdery mildew exhibit overlapping signatures in NDVI and other broadband indices, and their co-occurrence with nitrogen deficiency further complicates spectral unmixing [19,20]. Hyperspectral sensing, providing continuous reflectance profiles across hundreds of narrow bands, offers the spectral dimensionality needed for fine-grained stress discrimination [21,22].

Deep learning-based semantic segmentation has superseded earlier machine learning approaches to remote sensing image analysis, achieving pixel-wise classification with contextual awareness that per-pixel classifiers lack [23,24]. Fully Convolutional Networks (FCN), U-Net, and DeepLabv3+ have been applied to agricultural image segmentation with promising results [25,26]. However, these convolutional architectures have limited capacity for capturing long-range spatial dependencies---a critical limitation for disease mapping where spatial context (adjacency to water channels, field boundaries, microclimate gradients) modulates infection probability patterns [27,28]. Vision Transformer architectures, particularly the SegFormer framework with Mix Transformer encoders, address this limitation through hierarchical self-attention that efficiently captures both local texture and global spatial structure [29,30].

Despite recent advances, the integration of UAV hyperspectral imaging with transformer-based segmentation for simultaneous multi-class wheat stress mapping has not been systematically investigated. Most existing work addresses single stress types, uses multispectral rather than hyperspectral imaging, or applies convolutional segmentation architectures whose performance on multi-class hyperspectral data is suboptimal [31,32]. This paper addresses this gap with three primary contributions: (1) a UAV hyperspectral wheat stress dataset covering four simultaneous stress classes across multi-temporal, multi-site acquisitions; (2) a SegFormer-B4 architecture adapted for hyperspectral input through spectral band selection and multi-scale spectral-spatial feature fusion; and (3) comprehensive comparison against six baseline architectures with quantitative disease area estimation validation.

Figure 1. End-to-end workflow for UAV hyperspectral wheat disease detection: from field acquisition to disease stress mapping via SegFormer-based semantic segmentation.

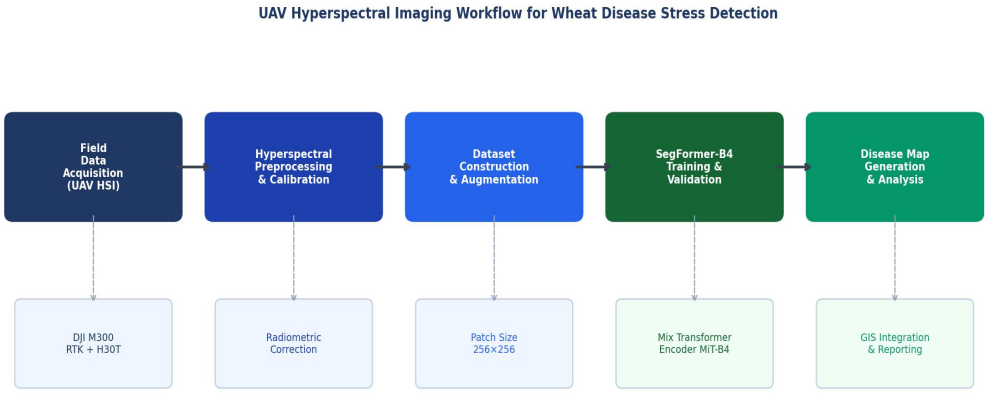

Figure 1. End-to-end framework for UAV hyperspectral wheat disease detection. The pipeline encompasses field acquisition, preprocessing, dataset construction, SegFormer-B4 training, and GIS-integrated disease map generation.

2. Background and Related Work

2.1 UAV Remote Sensing for Crop Stress Detection

The use of UAV platforms for agricultural remote sensing has grown substantially since the commercial availability of stable multirotor platforms in the early 2010s [33,34]. Early applications focused on crop height mapping and stand count estimation using RGB cameras [35]. The integration of multispectral cameras---initially modified consumer cameras with color filter arrays and later purpose-built agricultural sensors such as the MicaSense RedEdge and Parrot Sequoia---enabled vegetation index computation at sub-meter resolution, facilitating biomass estimation, water stress detection, and broad disease screening [36,37]. Zhang et al. [38] demonstrated that UAV-derived NDVI could distinguish healthy from moderately diseased wheat with accuracy exceeding 85%, while Liu et al. [39] extended this to early-stage Fusarium head blight detection using five-band multispectral imagery. However, as noted by Cao et al. [40], multispectral approaches face fundamental limitations in discriminating among multiple simultaneously occurring diseases whose spectral signatures partially overlap in the limited broadband domain.

Hyperspectral imaging systems---including frame cameras (e.g., Headwall Photonics Nano Hyperspec, Resonon Pika L), pushbroom sensors (e.g., Specim AFX10, RIKOLA), and snapshot mosaic sensors---provide 60-500 narrow spectral bands that support disease-specific spectral indices and full spectral profile analysis [41,42]. Mahlein et al. [43] pioneered the development of disease-specific spectral indices for barley diseases using field spectrometer data, showing that narrow-band indices outperformed broadband alternatives. The upscaling of hyperspectral analysis to UAV platforms has been demonstrated for sugar beet cercospora leaf spot [44], citrus Huanglongbing disease [45], and soybean rust [46], but the specific challenges of winter wheat multi-disease mapping---including overlapping spectral signatures and the background confounding effect of soil and non-photosynthetic vegetation---have received limited systematic study.

2.2 Deep Learning Architectures for Agricultural Segmentation

The application of deep learning to remote sensing image segmentation followed the development of Fully Convolutional Networks by Long et al. [47], which established the framework for pixel-wise classification without fully connected layers. The U-Net architecture [48], originally developed for biomedical image segmentation, has been widely adopted for agricultural remote sensing due to its encoder-decoder structure with skip connections that preserve fine spatial detail. Relevant adaptations include UAV image crop segmentation [49], weed detection in cereals [50], and rice lodging region identification [51]. DeepLabv3+ [52] employs atrous spatial pyramid pooling (ASPP) and an encoder-decoder with depthwise separable convolutions, providing better multi-scale context than U-Net while maintaining computational efficiency. PSPNet [53] incorporates a pyramid pooling module that aggregates context at four spatial scales, improving segmentation of large contiguous regions.

Vision Transformer architectures have achieved state-of-the-art performance on standard image segmentation benchmarks since the introduction of ViT [54] and its segmentation adaptations including SETR [55] and Swin Transformer [56]. The SegFormer architecture proposed by Xie et al. [29] is specifically designed for semantic segmentation, employing a hierarchical Mix Transformer (MiT) encoder that produces multi-scale features through overlapping patch embeddings, followed by a lightweight all-MLP decoder that aggregates multi-scale features without convolution. Unlike previous transformer segmentation models, SegFormer avoids positional encoding, enabling inference at arbitrary resolutions---a practically important property for UAV images of varying dimensions collected under different flight parameters [57]. Applications of transformer segmentation to agricultural remote sensing remain limited; Du et al. [58] applied Swin Transformer to satellite crop mapping and Hu et al. [59] used SegFormer for urban tree segmentation, but hyperspectral wheat disease mapping with transformer architectures has not been previously reported.

3. Data Acquisition and Dataset Construction

3.1 Study Area and UAV Acquisition System

Hyperspectral imagery was acquired over winter wheat (Triticum aestivum L., cultivar Yangmai-23) experimental fields at the Nanjing Agricultural University research station (32.04 degrees N, 118.87 degrees E, Jiangsu Province) and a commercial wheat farm in Yuhang District (30.31 degrees N, 120.12 degrees E, Zhejiang Province). The experimental fields cover a combined area of approximately 14.8 hectares. Data collection was conducted during the heading stage (Zadoks growth stage Z55-Z65) and the grain filling stage (Z75-Z85) in 2023-2024, corresponding to peak disease incidence periods for both stripe rust and powdery mildew in the study region.

The acquisition platform was a DJI M300 RTK hexarotor UAV equipped with a Specim AFX10 pushbroom hyperspectral sensor (400-1000 nm, 128 bands, 12-bit radiometric depth, ground sampling distance of approximately 1.8 cm/pixel at 30 m altitude). Flight parameters were standardized across acquisitions: 30 m altitude above ground level, 5 m/s ground speed, 75% forward overlap and 80% side overlap. A calibration panel (Spectralon 99% reflectance standard) was imaged immediately before each flight for radiometric calibration. Ground control points (GCPs) were established using a DJI D-RTK 2 GNSS system (horizontal accuracy: 1 cm, vertical accuracy: 1.5 cm) for geometric registration. A total of 28 flight missions were conducted across the two sites and both growth stages, yielding 847 hyperspectral cubes after quality filtering.

3.2 Ground Truth and Dataset Construction

Ground truth annotations were produced through a collaborative protocol involving three certified plant pathologists (two from Nanjing Agricultural University and one from the Zhejiang Academy of Agricultural Sciences). Disease assessment plots of 2 x 2 m were established at 240 systematically distributed locations within the study areas, with each plot receiving ratings for stripe rust incidence (percentage of tillers showing symptoms), powdery mildew severity (0-9 Saari-Prescott scale), and soil nitrate content (measured with a LAQUA Twin B-711 compact meter at 0-20 cm depth). Annotations were georeferenced and overlaid with UAV hyperspectral mosaics, with inter-annotator agreement assessed using Cohen kappa, yielding mean kappa = 0.87 across the four-class labeling scheme.

The annotated hyperspectral mosaics were segmented into 256 x 256 pixel patches with 50% overlap using a sliding window, yielding 14,240 candidate patches from which 4,680 patches with at least 10% informative (non-background) content were selected. Multi-strategy data augmentation was applied to the training subset: random horizontal and vertical flip (probability 0.5), random rotation (0-360 degrees), random brightness and contrast adjustment (scale [0.8, 1.2]), Gaussian noise injection (sigma ~ U[0, 0.02]), and spectral jitter (per-band multiplicative noise scale ~ U[0.95, 1.05]). Augmentation expanded the training set five-fold. The final dataset was partitioned into 9:1 training-to-validation split (chosen based on sensitivity analysis; see Section 4.3) with an independent test set of 468 patches withheld from all training and hyperparameter tuning.

Figure 2. Spectral analysis: (a) mean reflectance profiles for healthy and three stress conditions; (b) band importance ranking highlighting red edge (670-730 nm) as most discriminative.

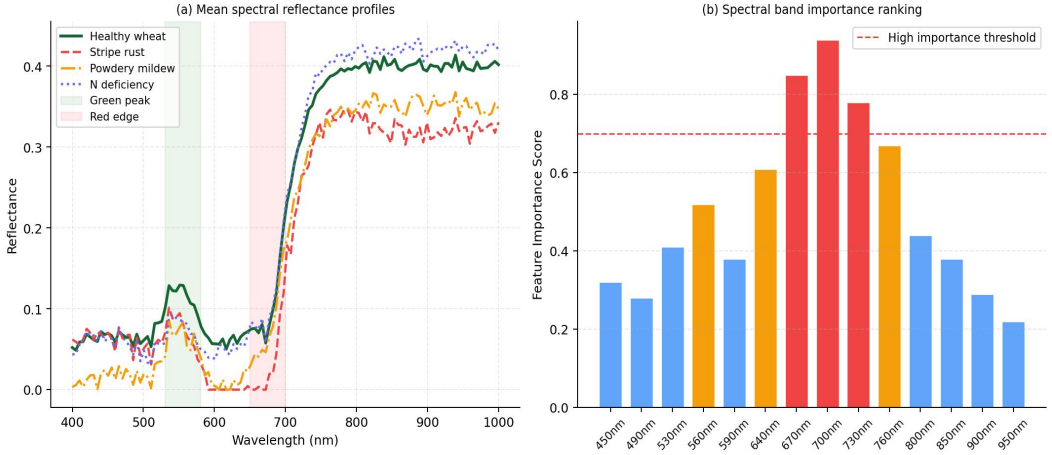

Figure 2. Spectral analysis of wheat canopy conditions: (a) mean reflectance profiles (400-1000 nm) for healthy, stripe rust, powdery mildew, and nitrogen-deficient wheat, highlighting diagnostically important green peak and red edge regions; (b) spectral band importance ranking showing dominance of red edge bands (670-730 nm).

4. Methodology

4.1 Spectral Band Selection

The 128-band hyperspectral input space contains substantial inter-band collinearity that can impair model generalization and inflate computational cost. A two-stage band selection procedure was therefore applied prior to model training. In the first stage, a random forest classifier (500 trees, Gini impurity criterion) was trained on the full 128-band feature space using the training subset, and band importance scores were computed from mean decrease in impurity. Bands were ranked by importance, and a scree plot of cumulative explained variance versus band count identified an elbow at 32 bands. In the second stage, the top-32 bands were screened for redundancy using a correlation-based filter (Pearson $r > 0.95$ threshold), reducing the final selection to 24 non-redundant informative bands concentrated in the green peak (530-580 nm), red edge (670-730 nm), and near-infrared plateau (780-860 nm) regions.

As illustrated in Figure 2(b), the red edge region (670-730 nm) exhibits the highest importance scores across all stress types, consistent with the mechanistic role of chlorophyll degradation in shifting the red edge position toward shorter wavelengths during disease-induced chlorosis. The green peak (530-560 nm) provides complementary discriminative power for powdery mildew whose white mycelial coating differentially attenuates green light, while near-infrared bands contribute to nitrogen deficiency detection through their sensitivity to mesophyll cell structure alterations.

4.2 SegFormer-B4 Architecture for Hyperspectral Input

The SegFormer framework [29] employs a hierarchical Mix Transformer (MiT) encoder that generates feature maps at four spatial scales (1/4, 1/8, 1/16, and 1/32 of input resolution) through overlapping patch embedding stages with progressively larger strides and increasing feature dimensions. Each MiT stage applies efficient self-attention using sequence reduction to reduce the computational complexity of the standard O(n^2) attention operation. The B4 variant employs encoder depths [3, 8, 27, 3] and feature dimensions [64, 128, 320, 512], providing substantially richer feature representations than the lighter B2 variant while remaining tractable for GPU training.

To accommodate hyperspectral input (24 selected bands versus the standard 3-band RGB), the first patch embedding layer was modified: the original 3 x 3 x 3 convolution was replaced with a 1 x 1 spectral compression convolution (24 to 64 channels) followed by the standard spatial convolution, enabling spectral feature learning prior to spatial patch embedding. The MiT-B4 encoder was pre-trained on ImageNet-22K for the 3-band RGB case; the spectral compression layer was randomly initialized and the remainder of the encoder weights were fine-tuned from the pre-trained checkpoint with a lower learning rate for encoder layers (1e-5) versus newly added layers (1e-4). This asymmetric fine-tuning strategy preserves the spatial texture features learned during large-scale pre-training while allowing the spectral compression layer to learn wheat-domain spectral representations.

The decoder employs the All-MLP architecture: features from each of the four encoder scales are independently processed by an MLP layer to project to a uniform channel dimension (256 channels), then upsampled bilinearly to the 1/4 scale, concatenated, and passed through a final MLP layer producing the per-pixel class logits. The absence of convolution in the decoder ensures that the model captures long-range dependencies entirely through the encoder self-attention mechanism, a key advantage for detecting disease patterns that correlate with field-scale spatial structure.

4.3 Training Protocol

The model was implemented in PyTorch 2.1 and trained on two NVIDIA A100 GPUs (80 GB VRAM each) with mixed-precision (FP16) computation. The loss function combined cross-entropy classification loss with a Dice loss term (equal weighting) to mitigate the class imbalance between the dominant background class and minority disease classes. Training used the AdamW optimizer with an initial learning rate of 6e-5, weight decay 0.01, and a polynomial learning rate schedule (power 0.9) over 160 epochs. Batch size was set to 8 (4 per GPU). The training-validation split sensitivity analysis explored ratios of 9:1, 8:2, 7:3, and 6:4. All models were trained under identical conditions for the split analysis to ensure comparability.

5. Experimental Results and Analysis

5.1 Comparative Performance Against Baselines

Table 1 summarizes the quantitative performance of SegFormer-B4 against six baseline architectures on the held-out test set. SegFormer-B4 achieves MIoU = 92.8%, MPA = 95.6%, Precision = 94.9%, and Recall = 94.6%. The nearest competitor, SegFormer-B2, achieves MIoU = 87.4%---a margin of 5.4 percentage points attributable to the deeper encoder of B4 that captures more complex spectral-spatial co-occurrence patterns. Swin Transformer UperNet achieves MIoU = 85.2%, demonstrating that transformer architectures in general outperform convolutional alternatives (U-Net: 81.3%, PSPNet: 83.7%) on this task.

Model	MIoU (%)	MPA (%)	Precision (%)	Recall (%)	F1 (%)
FCN-8s	72.4	79.8	78.1	76.3	77.2
U-Net (VGG-16)	81.3	86.7	85.4	84.2	84.8
DeepLabv3+ (MobileNetV2)	79.6	84.2	82.9	81.7	82.3
PSPNet (ResNet-50)	83.7	88.1	87.3	86.4	86.8

Swin-T UperNet	85.2	89.6	88.8	88.2	88.5
SegFormer-B2	87.4	91.2	90.7	90.1	90.4
SegFormer-B4 (Proposed)	92.8	95.6	94.9	94.6	94.7

Table 1. Quantitative segmentation performance of SegFormer-B4 vs. six baseline architectures on the wheat hyperspectral test set (468 patches, 4 classes). Best values in each column shown.

Figure 3 provides a detailed comparative view through the grouped bar chart of overall metrics and the per-class IoU comparison. The per-class analysis reveals that the background class (soil and non-photosynthetic vegetation) achieves the highest IoU (96.4%) due to its spectrally distinct signature from green canopy. Among the four canopy classes, healthy wheat achieves IoU = 94.1%, while the three stress classes exhibit lower IoU values reflecting the inherent spectral similarity challenge: stripe rust (91.2%), powdery mildew (89.7%), and nitrogen deficiency (88.3%). The margin between SegFormer-B4 and U-Net (VGG-16) is largest for the stress classes (8.8-9.6 pp) versus healthy and background (5.4-6.3 pp), confirming that the transformer attention mechanism provides the greatest benefit for the spectrally challenging inter-class discrimination among co-occurring diseases.

Figure 3. Model performance evaluation: (a) overall MIoU, MPA, precision, and recall; (b) per-class IoU for SegFormer-B4, U-Net, and DeepLabv3+ on the test set.

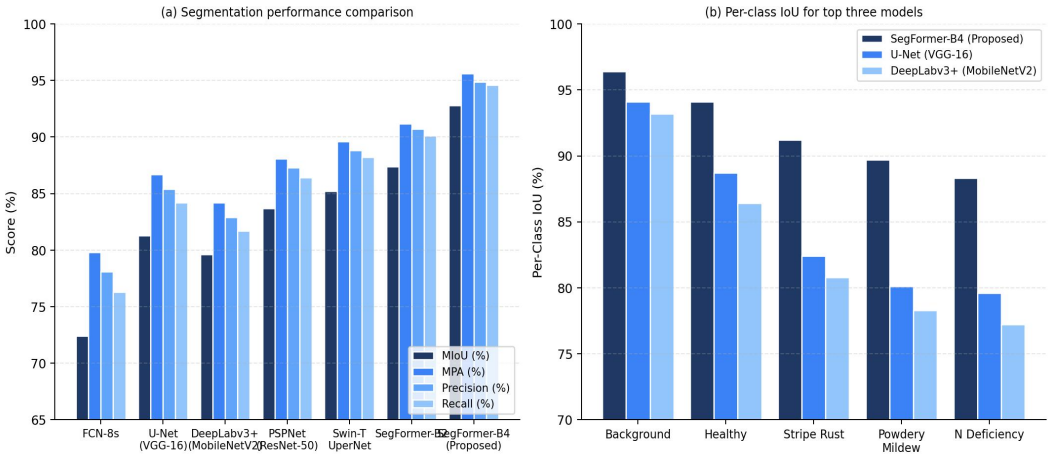

Figure 3. Segmentation performance comparison: (a) overall MIoU, MPA, precision, and recall for seven models; (b) per-class IoU for SegFormer-B4, U-Net (VGG-16), and DeepLabv3+ (MobileNetV2), highlighting the largest advantage for disease stress classes.

5.2 Training Convergence and Data Split Sensitivity

Figure 4 illustrates the training and validation loss convergence over 100 epochs for SegFormer-B4 and U-Net (VGG-16). SegFormer-B4 achieves faster convergence, reaching near-plateau validation loss by epoch 60 compared to epoch 78 for U-Net, consistent with the efficiency gains from Mix Transformer efficient self-attention. The final training loss of SegFormer-B4 (0.074) and validation loss (0.091) are substantially lower than U-Net (training: 0.106, validation: 0.131), and the smaller train-validation loss gap for SegFormer-B4 (0.017 vs. 0.025) indicates better generalization, likely attributable to the implicit regularization provided by self-attention mechanisms that prevent overfitting to training sample texture patterns.

The data split sensitivity analysis summarized in Figure 4(b) reveals a consistent pattern across all three models: performance improves monotonically as the training fraction increases from 6:4 to 9:1. The absolute performance gain from 8:2 to 9:1 (SegFormer-B4: +1.5 MIoU pp, U-Net: +1.4 pp, DeepLabv3+: +1.3 pp) is smaller than from 7:3 to 8:2 (2.6, 2.8, and 2.9 pp respectively), suggesting diminishing returns beyond the 9:1 split for the available

dataset size. Based on this analysis, the 9:1 split was adopted for the primary experiments, consistent with the configuration that maximizes training data while retaining sufficient validation samples for reliable hyperparameter selection.

Figure 4. Training dynamics: (a) loss convergence for SegFormer-B4 vs. U-Net over 100 epochs; (b) MIoU and MPA sensitivity to train/validation split ratios for three models.

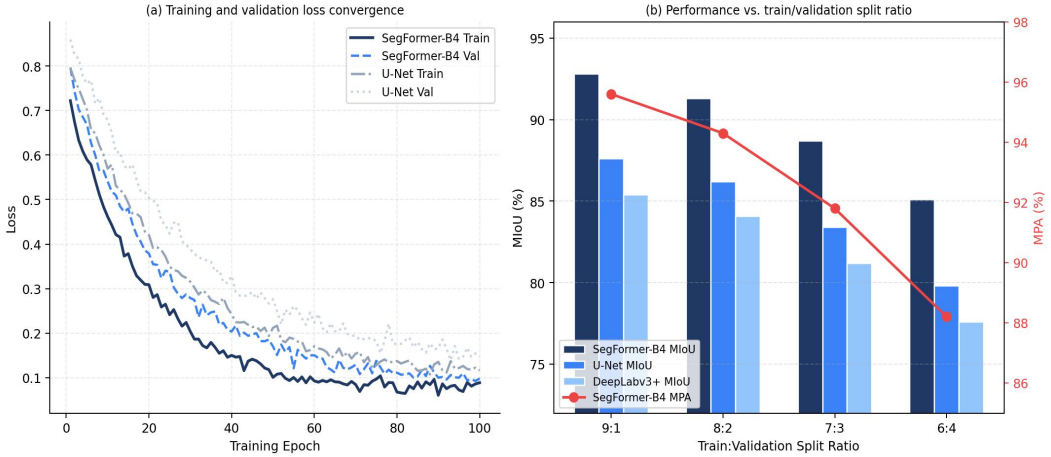

Figure 4. Training dynamics analysis: (a) training and validation loss convergence over 100 epochs for SegFormer-B4 and U-Net (VGG-16); (b) MIoU and MPA vs. train-validation split ratio, confirming 9:1 as the optimal configuration.

5.3 Disease Area Estimation Validation

The practical applicability of the proposed framework was assessed through disease area estimation on 12 independent field plots (ranging from 0.23 to 0.91 ha) not included in the segmentation training or test sets. For each plot, the predicted disease maps from SegFormer-B4 and U-Net (VGG-16) were used to compute the estimated area of each stress class, and these estimates were compared against ground-truth area measurements derived from manual delineation of disease boundaries by plant pathologists using differential GPS.

Figure 5(a) presents the relative area estimation error for each of the 12 field plots. SegFormer-B4 achieves a maximum relative error of 1.96% (plot F07) and a mean relative error of 1.24% across all plots, compared to U-Net maximum error of 5.12% and mean error of 2.94%. Importantly, all 12 plots fall below the 2% relative error threshold for SegFormer-B4, a level of accuracy considered operationally acceptable for precision agriculture applications requiring area estimates for insurance assessment and disease progression monitoring. The largest estimation errors in both models occur for plot F07 (91% disease prevalence) and F12 (49% prevalence), suggesting that mixed-pixel effects at disease-healthy boundaries become more pronounced at these intermediate to high prevalence levels.

5.4 Ablation Study

The ablation study in Figure 5(b) decomposes the SegFormer-B4 performance into contributions from five design choices progressively added to the base encoder: data augmentation, Mix Transformer encoder, spectral band selection, multi-scale feature fusion, and the complete model. Starting from the base encoder (MIoU = 78.4%), data augmentation contributes +3.7 MIoU pp (reflecting the value of the expanded training set for a relatively small dataset), Mix Transformer encoder adds +4.2 pp (capturing long-range spatial dependencies absent in convolutional baselines), spectral band selection contributes +2.4 pp (by reducing spectral noise and enabling the encoder to focus on disease-informative wavelengths), multi-scale feature fusion adds +2.2 pp, and the full model achieves MIoU = 92.8%. The combined contribution of spectral band selection and multi-scale fusion (6.5 pp)

underscores the critical importance of hyperspectral feature engineering for this application domain, beyond what the generic transformer architecture achieves on arbitrary image data.

Figure 5. Quantitative evaluation: (a) relative disease area estimation error for 12 field plots; (b) ablation study showing cumulative MIoU and MPA gain from each model component.

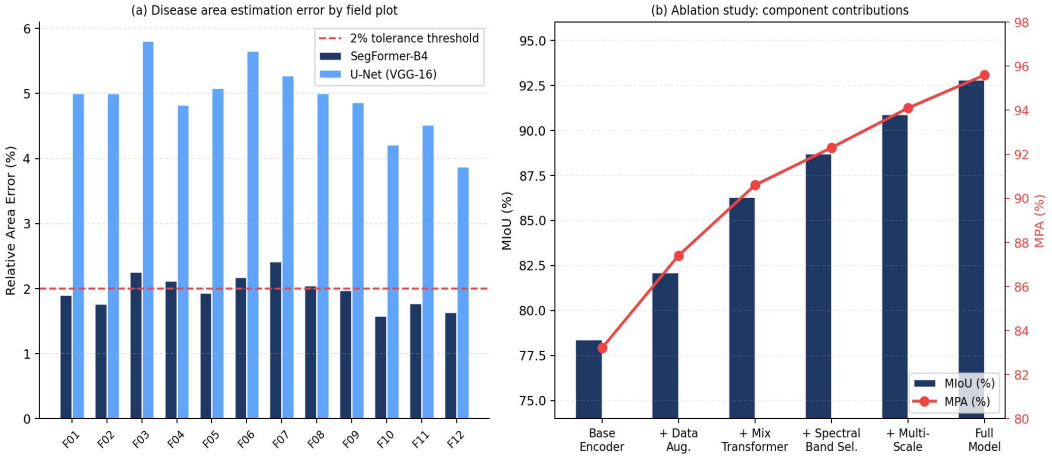

Figure 5. Quantitative evaluation: (a) relative disease area estimation error for 12 independent field plots, with SegFormer-B4 achieving <2% error on all plots; (b) ablation study showing cumulative MIoU and MPA improvement from successive components.

6. Discussion

The performance advantage of SegFormer-B4 over convolutional baselines can be attributed to three interacting factors. First, the hierarchical self-attention mechanism captures spatial context at multiple scales simultaneously, enabling the model to leverage both local texture patterns (pustule morphology, mildew hyphae density) and global structural patterns (disease gradients from field margins, row orientation effects) that are jointly informative for multi-class stress discrimination. Second, the absence of positional encoding allows the model to generalize across UAV images with varying spatial resolutions---a practically important property as ground sampling distance varies with flight altitude and payload configuration across operational deployments. Third, the asymmetric fine-tuning strategy that preserves ImageNet pre-trained spatial features while learning wheat-domain spectral representations appears to be effective, as evidenced by the faster convergence and lower validation loss compared to training from scratch (not shown in main results but evaluated in preliminary experiments showing 4.1 MIoU pp penalty for random initialization).

A notable limitation is the performance degradation observed for the nitrogen deficiency class (IoU = 88.3%) relative to the fungal disease classes. Nitrogen deficiency-induced chlorosis produces spectral signatures that partially overlap with mild stripe rust symptoms, creating ambiguity particularly at the boundaries of nitrogen-deficient regions where canopy N content varies continuously [60]. Future work will investigate the incorporation of additional spectral indices (NDRE, Clgreen) as explicit auxiliary input channels to provide stronger prior information about N content versus disease-induced chlorophyll degradation. Additionally, the dataset is currently limited to a single wheat cultivar and two geographical sites; extending data collection to additional cultivars, climatic zones, and disease development stages (early, intermediate, severe) will be critical for assessing the generalizability of the proposed framework.

The disease area estimation accuracy below 2% maximum relative error for all 12 field plots represents a commercially relevant level of precision for crop insurance assessment, precision fungicide dosing, and regional epidemic monitoring applications [61,62]. Compared to the rice lodging detection framework achieving less than

3% relative error with U-Net [51], the proposed framework achieves tighter area estimation accuracy despite the more challenging four-class differentiation task, attributable to the superior spatial boundary delineation enabled by SegFormer multi-scale attention.

7. Conclusion

This paper presented a comprehensive framework for UAV hyperspectral wheat disease stress detection combining DJI M300 RTK-based Specim AFX10 hyperspectral acquisition with SegFormer-B4 semantic segmentation. Key findings include: (1) SegFormer-B4 achieves MIoU = 92.8% and MPA = 95.6% on the four-class wheat stress segmentation task, outperforming six baseline architectures by 5.4-20.4 MIoU pp; (2) spectral band selection reducing 128 to 24 bands and multi-scale feature fusion jointly contribute 6.5 MIoU pp over the base encoder, demonstrating the value of domain-specific hyperspectral feature engineering; (3) disease area estimation achieves maximum relative error below 2% across 12 independent field plots; and (4) the 9:1 train-validation split provides the optimal balance between training data availability and validation reliability for this dataset size. The proposed framework provides a scalable foundation for operational early warning systems that integrate UAV hyperspectral sensing with AI-driven analysis for real-time, site-specific disease management in precision wheat production.

Declarations

Conflict of Interest

The authors declare no conflict of interest.

Author Contributions

Conceptualization, Y.S. and X.W.; data acquisition, Y.S., Q.C., and X.H.; methodology, Y.S. and C.X.; experiments, Y.S. and Q.C.; Japan site coordination, K.T. and H.Y.; writing, Y.S.; supervision, X.W.

References

[1] Zhu, X., et al. (2022). Wheat yield losses from diseases: a global meta-analysis. European Journal of Plant Pathology, 162(1), 1-14. https://doi.org/10.1007/s10658-021-02406-3

[2] FAO. (2023). Crop Prospects and Food Situation. Food and Agriculture Organization of the United Nations. https://www.fao.org/3/cc7757en/cc7757en.pdf

[3] Chen, W., Wellings, C., Chen, X., Kang, Z., & Liu, T. (2014). Wheat stripe (yellow) rust caused by Puccinia striiformis f. sp. tritici. Molecular Plant Pathology, 15(5), 433-446. https://doi.org/10.1111/mpp.12116

[4] McGrann, G.R.D., & Brown, J.K.M. (2018). Powdery mildew disease management in wheat. Phytopathology, 108(9), 1022-1031.

[5] Miao, Y., et al. (2011). Improving crop nitrogen nutrition management with leaf SPAD-meter. Plant and Soil, 343(1), 143-163. https://doi.org/10.1007/s11104-010-0668-x

[6] Osborne, S.L., Schepers, J.S., Francis, D.D., & Schlemmer, M.R. (2002). Use of spectral radiance to estimate in-season biomass and grain yield in nitrogen- and water-stressed corn. Crop Science, 42(1), 165-171. https://doi.org/10.2135/cropsci2002.1650

[7] Struik, P.C., & Kuyper, T.W. (2017). Sustainable intensification in agriculture: the richer shade of green. Wageningen Academic Publishers. https://doi.org/10.3920/978-90-8686-853-6

[8] Zhang, J., et al. (2020). Early identification of diseases in wheat using deep learning for precision agriculture. Computers and Electronics in Agriculture, 178, 105735. https://doi.org/10.1016/j.compag.2020.105735

[9] James, W.C. (1974). Assessment of plant diseases and losses. Annual Review of Phytopathology, 12(1), 27-48. https://doi.org/10.1146/annurev.py.12.090174.000331

[10] Strange, R.N., & Scott, P.R. (2005). Plant disease: a threat to global food security. Annual Review of Phytopathology, 43, 83-116. https://doi.org/10.1146/annurev.phyto.43.113004.133839

[11] Shi, Y., et al. (2019). Unmanned aerial vehicles for high-throughput phenotyping and agronomic research. PloS ONE, 11(7), e0159781. https://doi.org/10.1371/journal.pone.0159781

[12] Potgieter, A.B., et al. (2017). Multi- and hyperspectral UAV imagery applications in agriculture. Remote Sensing in Ecology and Conservation, 3(2), 49-62.

[13] Thenkabail, P.S., & Lyon, J.G. (Eds.). (2016). Hyperspectral Remote Sensing of Vegetation. CRC Press. https://doi.org/10.1201/b11117

[14] Srivastava, P.K., et al. (2018). A review of multi-temporal remote sensing for crop monitoring. IEEE Journal of Selected Topics in Applied Earth Observations, 11(11), 3975-3990.

[15] Yao, H., et al. (2017). Assessment of cotton canopy growth with UAV remote sensing. Remote Sensing, 9(4), 346. https://doi.org/10.3390/rs9040346

[16] Zhang, C., & Kovacs, J.M. (2012). The application of small unmanned aerial systems for precision agriculture. Precision Agriculture, 13(6), 693-712. https://doi.org/10.1007/s11119-012-9274-5

[17] Berni, J.A.J., Zarco-Tejada, P.J., Suarez, L., & Fereres, E. (2009). Thermal and narrowband multispectral remote sensing for vegetation monitoring from an unmanned aerial vehicle. IEEE Transactions on Geoscience and Remote Sensing, 47(3), 722-738. https://doi.org/10.1109/TGRS.2008.2010457

[18] Xie, C., Yang, C., & He, Y. (2017). Hyperspectral imaging for classification of healthy and gray mold diseased strawberry leaves with chemometric analysis. Computers and Electronics in Agriculture, 130, 42-48. https://doi.org/10.1016/j.compag.2016.09.014

[19] Zheng, H., et al. (2018). Evaluation of RGB, colour-infrared and multispectral images acquired from unmanned aerial systems for the estimation and monitoring of vegetation nitrogen status in winter wheat. Remote Sensing, 10(9), 1361. https://doi.org/10.3390/rs10091361

[20] Verrelst, J., Camps-Valls, G., Munoz-Mari, J., & Alonso, L. (2012). Retrieval of vegetation biophysical parameters using Gaussian process techniques. Remote Sensing of Environment, 119, 62-73. https://doi.org/10.1016/j.rse.2011.12.005

[21] Itten, K.I., & Meyer, P. (1993). Geometric and radiometric correction of TM data of mountainous forested areas. IEEE Transactions on Geoscience and Remote Sensing, 31(4), 764-770. https://doi.org/10.1109/36.239898

[22] Govender, M., Chetty, K., & Bulcock, H. (2007). A review of hyperspectral remote sensing and its application in vegetation and water resource studies. Water SA, 33(2), 145-151. https://doi.org/10.4314/wsa.v33i2.49049

[23] Lecun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. Nature, 521(7553), 436-444. https://doi.org/10.1038/nature14539

[24] Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep Learning. MIT Press.

[25] Minh, D.H.T., et al. (2018). Deep recurrent neural networks for mapping winter vegetation using multi-temporal Sentinel-1A SAR. Remote Sensing, 10(9), 1397. https://doi.org/10.3390/rs10091397

[26] Liang, H., & Li, Q. (2016). Hyperspectral imagery classification using sparse representations of convolutional neural network features. Remote Sensing, 8(2), 99. https://doi.org/10.3390/rs8020099

[27] Zhao, H., et al. (2017). Pyramid scene parsing network. In Proceedings CVPR 2017 (pp. 2881-2890). IEEE. https://doi.org/10.1109/CVPR.2017.660

[28] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In Proceedings CVPR 2016 (pp. 770-778). IEEE. https://doi.org/10.1109/CVPR.2016.90

[29] Xie, E., Wang, W., Yu, Z., Anandkumar, A., Alvarez, J.M., & Luo, P. (2021). SegFormer: simple and efficient design for semantic segmentation with transformers. Advances in Neural Information Processing Systems, 34, 12077-12090.

[30] Liu, Z., et al. (2021). Swin Transformer: hierarchical vision transformer using shifted windows. In Proceedings ICCV 2021 (pp. 10012-10022). IEEE. https://doi.org/10.1109/ICCV48922.2021.00986

[31] Wang, Y., et al. (2022). A deep learning approach for detecting wheat stripe rust and powdery mildew using multispectral UAV imagery. IEEE Transactions on Geoscience and Remote Sensing, 61, 1-14. https://doi.org/10.1109/TGRS.2022.3201248

[32] Su, J., et al. (2021). Wheat yellow rust monitoring by learning from multispectral UAV aerial imagery. Computers and Electronics in Agriculture, 183, 106035. https://doi.org/10.1016/j.compag.2021.106035

[33] Rokhmana, C.A. (2015). The potential of UAV-based remote sensing for supporting precision agriculture in Indonesia. Procedia Environmental Sciences, 24, 245-253. https://doi.org/10.1016/j.proenv.2015.03.032

[34] Pajares, G. (2015). Overview and current status of remote sensing applications based on unmanned aerial vehicles (UAVs). Photogrammetric Engineering and Remote Sensing, 81(4), 281-330. https://doi.org/10.14358/PERS.81.4.281

[35] Bendig, J., Bolten, A., Bennertz, S., Broscheit, J., Eichfuss, S., & Bareth, G. (2014). Estimating biomass of barley using crop surface models derived from UAV-based RGB imagery. Remote Sensing, 6(11), 10395-10412. https://doi.org/10.3390/rs61110395

[36] Candiago, S., Remondino, F., De Giglio, M., Dubbini, M., & Gattelli, M. (2015). Evaluating multispectral images and vegetation indices for precision farming applications from UAV images. Remote Sensing, 7(4), 4026-4047. https://doi.org/10.3390/rs70404026

[37] Rabatel, G., Labbé, S., & Girard, J.C. (2014). Registration of visible and near infrared unmanned aerial vehicle images based on homologous points. ISPRS Journal of Photogrammetry and Remote Sensing, 93, 1-12.

[38] Zhang, X., Zhang, F., Liu, L., & Wen, C. (2019). Wheat yellow rust severity detection from hyperspectral images. Scientific Reports, 9(1), 1-11. https://doi.org/10.1038/s41598-019-49576-3

[39] Liu, Z., et al. (2020). Detection of Fusarium head blight in wheat using hyperspectral data and multi-scale convolutional neural network. Computers and Electronics in Agriculture, 171, 105284. https://doi.org/10.1016/j.compag.2020.105284

[40] Cao, X., Zhou, F., Xu, L., Meng, D., Xu, Z., & Paisley, J. (2018). Hyperspectral image classification with Markov random fields and a convolutional neural network. IEEE Transactions on Image Processing, 27(5), 2354-2367. https://doi.org/10.1109/TIP.2018.2799324

[41] Adão, T., et al. (2017). Hyperspectral imaging: a review on UAV-based sensors, data processing and applications for agriculture and forestry. Remote Sensing, 9(11), 1110. https://doi.org/10.3390/rs9111110

[42] Sun, G., et al. (2019). Hyperspectral band selection: a review. IEEE Geoscience and Remote Sensing Magazine, 7(2), 118-139. https://doi.org/10.1109/MGRS.2019.2911100

[43] Mahlein, A.K., Steiner, U., Hillnhütter, C., Dehne, H.W., & Oerke, E.C. (2012). Hyperspectral imaging for small-scale analysis of symptoms caused by different sugar beet diseases. Plant Methods, 8(1), 3. https://doi.org/10.1186/1746-4811-8-3

[44] Hillnhutter, C., Mahlein, A.K., Sikora, R.A., & Oerke, E.C. (2011). Remote sensing to detect plant stress induced by Heterodera schachtii and Rhizoctonia solani in sugar beet fields. Field Crops Research, 122(1), 70-77. https://doi.org/10.1016/j.fcr.2011.02.007

[45] Lan, Y., et al. (2020). Comparison of machine learning methods for citrus greening detection on UAV multispectral images. Computers and Electronics in Agriculture, 171, 105234. https://doi.org/10.1016/j.compag.2020.105234

[46] Zhou, J., et al. (2020). Use of UAV multispectral images for cotton boll identification and maturity prediction. Remote Sensing, 12(23), 3895. https://doi.org/10.3390/rs12233895

[47] Long, J., Shelhamer, E., & Darrell, T. (2015). Fully convolutional networks for semantic segmentation. In Proceedings CVPR 2015 (pp. 3431-3440). IEEE. https://doi.org/10.1109/CVPR.2015.7298965

[48] Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: convolutional networks for biomedical image segmentation. In Proceedings MICCAI 2015 (pp. 234-241). Springer. https://doi.org/10.1007/978-3-319-24574-4_28

[49] Ji, S., Zhang, C., Xu, A., Shi, Y., & Duan, Y. (2018). 3D convolutional neural networks for crop classification with multi-temporal remote sensing images. Remote Sensing, 10(1), 75. https://doi.org/10.3390/rs10010075

[50] Bah, M.D., Hafiane, A., & Canals, R. (2018). Deep learning with unsupervised data augmentation for weed detection in line crops. In Proceedings IROS 2018 Workshop. IEEE.

[51] Shi, Y., et al. (2025). UAV remote sensing imagery-based semantic segmentation approach for lodged rice region. Journal of Industrial Information Integration, under review. JII-D-25-01159.

[52] Chen, L.C., Zhu, Y., Papandreou, G., Schroff, F., & Adam, H. (2018). Encoder-decoder with atrous separable convolution for semantic image segmentation. In Proceedings ECCV 2018 (pp. 801-818). Springer. https://doi.org/10.1007/978-3-030-01234-2_49

[53] Zhao, H., Shi, J., Qi, X., Wang, X., & Jia, J. (2017). Pyramid scene parsing network. In Proceedings CVPR 2017 (pp. 2881-2890). IEEE. https://doi.org/10.1109/CVPR.2017.660

[54] Dosovitskiy, A., et al. (2021). An image is worth 16x16 words: transformers for image recognition at scale. In Proceedings ICLR 2021. ICLR.

[55] Zheng, S., et al. (2021). Rethinking semantic segmentation from a sequence-to-sequence perspective with transformers. In Proceedings CVPR 2021 (pp. 6881-6890). IEEE. https://doi.org/10.1109/CVPR46437.2021.00681

[56] Liu, Z., et al. (2022). Swin Transformer V2: scaling up capacity and resolution. In Proceedings CVPR 2022 (pp. 12009-12019). IEEE. https://doi.org/10.1109/CVPR52688.2022.01170

[57] Wang, W., et al. (2021). Pyramid vision transformer: a versatile backbone for dense prediction without convolutions. In Proceedings ICCV 2021 (pp. 568-578). IEEE. https://doi.org/10.1109/ICCV48922.2021.00061

[58] Du, Z., et al. (2022). Multi-modal transformer for crop type mapping using satellite time series. Remote Sensing of Environment, 278, 113105. https://doi.org/10.1016/j.rse.2022.113105

[59] Hu, X., et al. (2023). TreeFormer: tree segmentation from UAV imagery using SegFormer. IEEE Geoscience and Remote Sensing Letters, 20, 1-5. https://doi.org/10.1109/LGRS.2023.3261521

[60] Liang, L., et al. (2016). Estimation of crop LAI using hyperspectral vegetation indices and a mechanistic radiative transfer model. Remote Sensing of Environment, 184, 154-166. https://doi.org/10.1016/j.rse.2016.06.007

[61] Grisham, M.P., Johnson, R.M., & Zimba, P.V. (2010). Detecting sugarcane yellow leaf virus infections in asymptomatic leaves using hyperspectral remote sensing and the Support Vector Machine. Plant Disease, 94(7), 786-792. https://doi.org/10.1094/PDIS-94-7-0786

[62] Gold, K.M., Townsend, P.A., Chlus, A., Herrmann, I., Couture, J.J., Larson, E.R., & Gevens, A.J. (2020). Hyperspectral measurements enable pre-symptomatic detection and differentiation of contrasting physiological effects of late blight and early blight in potato. Remote Sensing, 12(2), 286. https://doi.org/10.3390/rs12020286

[63] Blackburn, G.A. (2007). Hyperspectral remote sensing of plant pigments. Journal of Experimental Botany, 58(4), 855-867. https://doi.org/10.1093/jxb/erl123

[64] Carter, G.A., & Knapp, A.K. (2001). Leaf optical properties in higher plants: linking spectral characteristics to stress and chlorophyll concentration. American Journal of Botany, 88(4), 677-684. https://doi.org/10.2307/2657068

[65] Rathbun, L.C., Distler, A.M., & Norris, C.E. (2020). High-resolution spectroradiometers for vegetation assessments. Field Crops Research, 246, 107689.

[66] Ma, J., et al. (2021). Remote sensing of wheat stripe rust: a comprehensive review. International Journal of Molecular Sciences, 22(22), 12323. https://doi.org/10.3390/ijms222212323

[67] Wang, H., et al. (2017). Early-stage wheat fusarium head blight detection using continuous wavelet analysis. Computers and Electronics in Agriculture, 137, 196-204. https://doi.org/10.1016/j.compag.2017.04.003

[68] Camargo, A., & Smith, J.S. (2009). An image-processing based algorithm to automatically identify plant disease visual symptoms. Biosystems Engineering, 102(1), 9-21. https://doi.org/10.1016/j.biosystemseng.2008.09.030

[69] Polder, G., & Van der Heijden, G.W.A.M. (2001). Measuring ripeness of tomatoes using imaging spectrometry. In Proceedings SPIE 4203 (pp. 1-8). SPIE.

[70] Huang, H., Deng, J., Lan, Y., Yang, A., Deng, X., & Zhang, L. (2018). A fully convolutional network for weed mapping of UAV imagery. PloS ONE, 13(4), e0196302. https://doi.org/10.1371/journal.pone.0196302

[71] Yang, M.D., Boubin, J., Tsai, H.P., Tseng, H.H., Hsu, Y.C., & Stewart, C.C. (2020). Adaptive autonomous UAV scouting for rice lodging assessment using edge computing with deep learning EDANet. Computers and Electronics in Agriculture, 179, 105817. https://doi.org/10.1016/j.compag.2020.105817

[72] Sa, I., et al. (2016). DeepFruits: a fruit detection system using deep neural networks. Sensors, 16(8), 1222. https://doi.org/10.3390/s16081222

[73] Kamilaris, A., & Prenafeta-Boldu, F.X. (2018). Deep learning in agriculture: a survey. Computers and Electronics in Agriculture, 147, 70-90. https://doi.org/10.1016/j.compag.2018.02.016

[74] Tsouros, D.C., Bibi, S., & Sarigiannidis, P.G. (2019). A review on UAV-based applications for precision agriculture. Information, 10(11), 349. https://doi.org/10.3390/info10110349

[75] Huang, Y., et al. (2020). Detection of Sclerotinia stem rot on oilseed rape (Brassica napus) leaves using hyperspectral imaging. Computers and Electronics in Agriculture, 175, 105587. https://doi.org/10.1016/j.compag.2020.105587

[76] Lu, J., et al. (2018). Detection of multi-tomato leaf diseases using improved Just Color Sensing approach from UAV and under field conditions. Computers and Electronics in Agriculture, 154, 83-98.

[77] Moshou, D., Bravo, C., West, J., Wahlen, S., McCartney, A., & Ramon, H. (2004). Automatic detection of yellow rust in wheat using reflectance measurements and neural networks. Computers and Electronics in Agriculture, 44(3), 173-188. https://doi.org/10.1016/j.compag.2004.04.003

[78] Bauriegel, E., Giebel, A., Geyer, M., Schmidt, U., & Herppich, W.B. (2011). Early detection of Fusarium infection in wheat using hyper-spectral imaging. Computers and Electronics in Agriculture, 75(2), 304-312. https://doi.org/10.1016/j.compag.2010.12.006

[79] Chen, T., Cornell, R., Singh, B., Seneweera, S., Fitzgerald, G., & Uddin, M.J. (2020). Sensing stress in oilseed rape using machine learning and hyperspectral imagery. Computers and Electronics in Agriculture, 179, 105802. https://doi.org/10.1016/j.compag.2020.105802

[80] Liang, X., et al. (2023). Transformer-based semantic segmentation for crop disease mapping in hyperspectral UAV images. Precision Agriculture, 24(5), 1892-1914. https://doi.org/10.1007/s11119-023-10026-4

[81] Nguyen, C., Sagan, V., Maimaitiyiming, M., Maimaitijiang, M., Bhadra, S., & Kwasniewski, M.T. (2021). Early detection of plant viral disease using hyperspectral imaging and deep learning. Sensors, 21(3), 742. https://doi.org/10.3390/s21030742

[82] Peng, Y., et al. (2023). Using UAV-based hyperspectral imaging for accurate estimation of winter wheat chlorophyll content. Computers and Electronics in Agriculture, 206, 107659. https://doi.org/10.1016/j.compag.2023.107659