From plots to region: Machine learning-based UAV-satellite integration for mapping fractional coverage of peanut southern blight

by Li, S., Liu, J., Zhang, Y., Feng, S., Liu, L., Yue, J., Liu, Y., Shu, M., Cheein, F.A. and Guo, W.

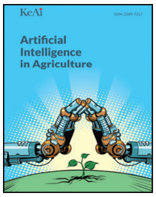
Li, S., Liu, J., Zhang, Y., Feng, S., Liu, L., Yue, J., Liu, Y., Shu, M., Cheein, F.A. and Guo, W. (2026) 'From plots to region: Machine learning-based UAV-satellite integration for mapping fractional coverage of peanut southern blight', *Artificial Intelligence in Agriculture*, 16(2), pp. 974-997.

Artificial Intelligence in Agriculture 16 (2026) 974–997

Contents lists available at ScienceDirect

Artificial Intelligence in Agriculture

journal homepage: http://www.keaipublishing.com/en/journals/artificial-intelligence-in-agriculture/



From plots to region: Machine learning-based UAV-satellite integration for mapping fractional coverage of peanut southern blight

Shangzhou Li [a,1], Juan Liu [b,1], Yapeng Zhang [a], Sifan Feng [a], Lu Liu [a], Jibo Yue [a], Yang Liu [a], Meiyan Shu [a], Fernando Auat Cheein [c], Wei Guo [a,*]

[a] College of Information and Management Science, Henan Agricultural University, Zhengzhou 450046, China
[b] Institute of Peanut Research, Henan Academy of Agricultural Sciences, Zhengzhou 450002, China
[c] Department of Engineering, Harper Adams University, Newport TF10 8NB, UK

ARTICLE INFO

Article history:
Received 4 November 2025
Received in revised form 13 April 2026
Accepted 14 April 2026
Available online 15 April 2026

Keywords:
Peanut southern blight
Fractional coverage
UAV-satellite integration
Machine learning
Precision agriculture

ABSTRACT

Peanut southern blight (PSB), a soil-borne fungal disease, threatens peanut production. Because PSB originates at the root collar, canopy symptoms develop gradually and subtly, limiting satellite monitoring. UAV observations can capture these symptoms but are spatially restricted. To address this, we developed a machine learning-based UAV-satellite integration framework using fractional coverage (FCover) as a severity and scale-bridging metric for regional-scale PSB monitoring. First, Sentinel-2 composites from Google Earth Engine were used to extract temporal features and construct NDVI time series, from which a novel Autumn Crop Index was derived and applied within the Dynamic World crop mask to delineate autumn crop extent. Within this extent, stable temporal features were identified via fully nested 10-fold cross-validation (CV), combining Random Forest (RF) importance with redundancy filtering (Separability Index and Pearson correlation). An RF classifier trained on these features with a grid-optimized hyperparameter set, chosen based on validation performance across all folds, achieved a mean overall accuracy of 0.98 and was subsequently applied for peanut mapping. Next, UAV-derived spectral, texture, and structural features were selected via ReliefF, correlation analysis, and recursive elimination to train RF classifiers. The resulting classification maps were aggregated to Sentinel pixels to generate weighted FCover samples. Finally, stable, temporally matched Sentinel-2-derived spectral and texture features were identified via fully nested 5-fold CV using weighted correlation. These features were then used to model FCover with weighted Bayesian-optimized RF, Extra Trees, Gradient Boosting, and eXtreme Gradient Boosting (XGB). XGB achieved the highest CV performance (overall $R^2 = 0.718$, RMSE $= 0.123$) and, using its optimal hyperparameters, was trained with FCover weights to generate a regional FCover map strongly agreeing with field-surveyed incidence ($R^2 = 0.89$). This study demonstrates, for the first time, that UAV-satellite integration enables effective PSB monitoring, providing a scalable approach for precision disease management in peanut-producing regions.

1. Introduction

Peanut (*Arachis hypogaea* L.) is an annual legume forming symbioses with rhizobia, enabling biological nitrogen fixation and enhancing soil fertility (Akram et al., 2018). Widely cultivated across tropical and subtropical regions, it contributes about 50 million tons to global annual production.[2] As an oilseed, food, and feed crop, peanut underpins food security, supports agro-industrial development, and sustains farmers' livelihoods, contributing to the global economy (Biswal et al., 2024; Maestri, 2024; Variath and Janila, 2017). Despite its economic and agronomic value, peanut cultivation faces increasing biotic stresses, among which peanut southern blight (PSB), caused by *Sclerotium rolfsii* (= *Athelia rolfsii*), is one of the most destructive (Chen et al., 2018b; Liu et al., 2024a). First reported in the southeastern United States in 1914, PSB has since been detected in peanut production regions worldwide, typically causing yield losses of 10–40%, and up to 80% in severe outbreaks (Aycock, 1966; Xie et al., 2014).

Symptoms of PSB generally appear between 60 and 90 days after cultivation, initially manifested as fluffy, spreading white mycelia on the soil surface and on the lateral branches or main stem of the peanut plant (Sanjel et al., 2024; Sconyers et al., 2005). As the disease progresses, leaves may exhibit a water-soaked appearance, followed by

---

* Corresponding author.
E-mail address: guowei@henau.edu.cn (W. Guo).

[1] These authors contributed equally to this work.

[2] https://web.archive.org/web/20250531190716/https://ipad.fas.usda.gov/cropexplorer/cropview/commodityView.aspx?cropid=2221000&sel_year=2025&rankby=Production (accessed June 25, 2025).

https://doi.org/10.1016/j.aiia.2026.04.008
2589-7217/

Fig. 1. Symptoms of PSB, including field manifestations, white mycelial growth on the basal stem at the soil surface, and infected pods.

chlorosis and wilting; in severe cases, the entire plant may die (Woodward et al., 2013). Infected pods often show brown spots and are covered with mycelia (see Fig. 1). Furthermore, if infection occurs at an early growth stage, pod formation may be inhibited (Branch and Csinos, 1987). PSB exhibits an aggregated or sporadic spatial distribution pattern in the field that can expand from localized outbreaks to regional epidemics when favorable environmental and climatic conditions prevail over larger areas (Punja, 1985; Rideout et al., 2008). In this context, regional-scale monitoring of PSB enhances understanding of epidemic extent and severity, supporting targeted interventions and long-term control strategies.

The rapid development of satellite remote sensing has enabled large-scale crop disease monitoring, addressing the limited spatial coverage of traditional manual scouting (Yuan et al., 2025). Extensive research has leveraged imagery from multiple platforms (e.g., Ziyuan-3, RapidEye, SPOT-6, Sentinel-2, Landsat-8, and MODIS) for crop disease monitoring by integrating spectral feature extraction, statistical analysis, and machine learning (ML) approaches, particularly targeting wheat rust and powdery mildew, thereby forming the conventional regional-scale disease monitoring frameworks (Chen et al., 2018a; Du et al., 2019; Ma et al., 2018, 2019, 2024a; Yuan et al., 2013, 2014, 2017; Zhang et al., 2014; Zheng et al., 2018). Beyond wheat, other crop diseases have also attracted growing research interest. For instance, Santoso et al. (2011) employed vegetation indices (VIs) derived from QuickBird imagery with threshold-based segmentation to map basal stem rot in oil palm, achieving accuracies of 59–67%. Similarly, Song et al. (2017) applied ISODATA unsupervised classification on Sentinel-2 A imagery to detect cotton root rot, reaching an overall accuracy (OA) of 91.2%, while Guo et al. (2022) utilized Random Forest (RF) on PlanetScope data to detect areca yellow leaf disease, obtaining an OA of 88.24%. Although both publicly accessible and restricted satellite datasets have been utilized, the former (e.g., Sentinel-2) remains the primary source. In recent years, Google Earth Engine (GEE) has provided a practical solution for efficient access and processing of these datasets (Nakalembe et al., 2021).

The platform[3] hosts petabytes of publicly accessible remote sensing datasets and derived products spanning over five decades, sourced from Sentinel, Landsat, MODIS, and other major satellite missions, and continuously updated with near real-time (NRT) satellite observations (Tamiminia et al., 2020). Users can access, process, and visualize these data interactively via GEE's web-based JavaScript code editor or Python API, with all operations powered by built-in functions that run in parallel on Google's cloud infrastructure (Amani et al., 2020; Velastegui-Montoya et al., 2023). Although launched in 2010, the platform only gained widespread adoption in scientific research after 2015, thanks to major improvements in its interface, data structure, and programming environment (Gorelick et al., 2017; Zhao et al., 2021). Since then, agriculture, particularly large-scale crop mapping, has emerged as one of its key application domains (Di et al., 2022; Dong et al., 2016; Peng et al., 2024; Wang et al., 2020; Wang et al., 2023b; Xun et al., 2021), underscoring the platform's value in supporting agricultural monitoring. Nevertheless, the limitations in spatial and temporal resolution (i.e., ground sampling distance and revisit interval) of the data it provides remain a critical bottleneck for achieving timely and precise field-level monitoring (Zhang et al., 2020). Moreover, high procurement costs and processing demands make sustained use of higher-resolution satellite data in precision agriculture both impractical and unnecessary (Ye et al., 2020). By comparison, unmanned aerial vehicles (UAVs) equipped with sensors (e.g., red-green-blue (RGB), multispectral, and hyperspectral) enable flexible, on-demand deployment for high-frequency monitoring and provide centimeter-level resolution imagery at a relatively low acquisition cost, making them a valuable complement to satellite remote sensing (Sishodia et al., 2020; Su et al., 2018; Zhang et al., 2021).

Building on these advantages, UAV remote sensing has been widely applied for field-level crop disease monitoring, with RGB and multispectral data being the two most commonly used sources (Kouadio et al., 2023; Neupane and Baysal-Gurel, 2021; Zhao et al., 2023). The former captures red, green, and blue light intensities, facilitating the generation of digital surface model (DSM) and supporting visual detection of crop diseases; the latter, after radiometric calibration, provides spectral reflectance across visible and non-visible bands, including red edge (RE) and near-infrared (NIR), enabling the construction of VIs such as the Normalized Difference Vegetation Index (NDVI) for assessing crop physiological status (Joshi et al., 2024). Beyond spectral characteristics, the latter contain spatial information, which can be quantified through texture features (TFs) derived from the gray-level co-occurrence matrix (GLCM) that capture local gray-level relationships, reflecting properties such as coarseness, smoothness, and regularity (Guo et al., 2020; Haralick et al., 1973). Studies have shown that incorporating TFs alongside spectral features improves the characterization of crop diseases, enhancing monitoring accuracy (Liu et al., 2024c; Zhu et al., 2022). Moreover, for diseases like PSB, which causes plant wilting and reduces canopy height (see Fig. 1), DSM can capture height differences between diseased and healthy plants, thereby providing additional structural information for disease monitoring.

Given the limited spatial coverage of UAV observations and the existing gap in accurate regional-scale monitoring of PSB (Guo et al., 2023, 2024; Sun et al., 2024; Wen et al., 2024), this study proposes a ML-based integrated UAV-satellite framework that differs from conventional disease monitoring approaches at the regional scale. Most existing studies treat satellite pixels as homogeneous units, ignoring that they may contain mixtures of healthy and diseased plants, and thus failing to capture within-pixel disease heterogeneity. To address this limitation, we introduce PSB fractional coverage (FCover), a sub-pixel metric representing the proportion of diseased peanut canopy, thereby quantifying the average disease severity within each satellite pixel. Unlike commonly used sub-pixel metrics based on spectral unmixing, in which fractional abundances are estimated through Linear Spectral Unmixing or Spectral Mixture Analysis using predefined endmembers (Shi and Wang, 2014), FCover is directly derived from high-resolution UAV-based disease classification maps, enabling aggregation to satellite pixels and bridging fine-scale UAV observations with regional-scale remote sensing. Building on this framework, the objectives of this study are to: (1) generate an accurate peanut distribution map at the regional scale; (2) estimate PSB FCover at the satellite pixel level using UAV-based classification; and (3) develop regression models to predict FCover from satellite-derived features, thereby contributing to multi-scale crop disease monitoring.

---

[3] https://developers.google.com/earth-engine/datasets/catalog (accessed June 30, 2025).

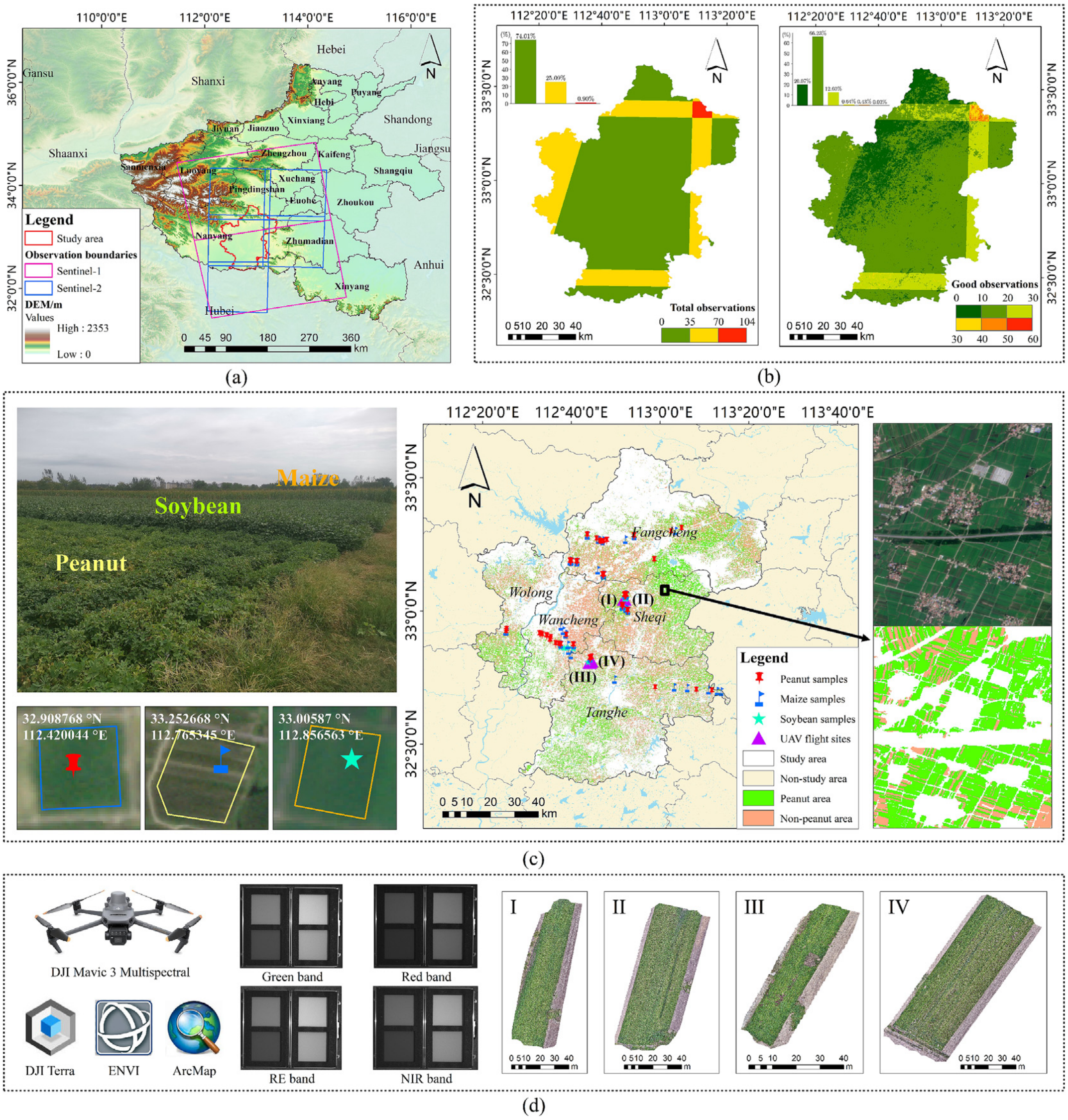


Fig. 2. Study area and multi-source information. (a) Geographic location, satellite observation boundaries and digital elevation model. (b) Sentinel-2 total and good observation frequencies between Jun 11 and Oct 20, 2024. (c) Example field photo and sample field boundaries; spatial distributions of samples, UAV flight sites, and peanut/non-peanut crops (maize and soybean). (d) UAV platform, processing software, example radiometric calibration panel images captured by the multispectral sensors, and UAV RGB orthomosaics.

2. Materials and methods

2.1. Study area

The study area, located in southwestern Henan Province, China (Fig. 2a, 112°17′-113°24′E, 32°20′-33°36′N), forms an important part of the Nanyang Basin. It encompasses two districts (Wancheng and Wolong) and three counties (Fangcheng, Sheqi, and Tanghe), with a total area of 8178 km^2, of which approximately 4854.6 km^2 is cultivated. The terrain is dominated by alluvial plains, with an average elevation of about 152 m. The region has a northern subtropical humid monsoon climate, characterized by hot and wet summers, which support the growth of autumn crops such as maize, peanut, and soybean. In 2023, the planting areas of maize, peanut, and soybean were 2472.8, 1579.2, and 188.3 km^2, respectively.[4] Peanut is cultivated in two seasonal cycles: spring peanuts are sown from late March to mid-April and harvested from late August (Aug) to early September (Sep); summer peanuts are sown in mid- to late June (Jun) and harvested from late Sep to mid-October (Oct). In recent years, climatic fluctuations during the growing season, combined with the absence of resistant peanut cultivars, have favored frequent PSB outbreaks, posing a persistent threat to peanut yield (Li et al., 2021).

2.2. Data and methodology

To achieve the objectives of this study, the overall research workflow consists of four key steps: (1) data acquisition and processing, (2) peanut spatial distribution extraction, (3) PSB FCover estimation using UAV-based classification, and (4) modeling FCover from satellite-derived features. Detailed procedures are illustrated in Fig. 3.

2.2.1. Data acquisition and processing

2.2.1.1. Field data. A field survey was conducted in the study area on Sep 11–12, 2024, during which 28 soybean, 82 maize, and 121 peanut samples were collected (Fig. 2c). Sample coordinates were recorded using OvitalMap (version 10.1.5, Beijing Ovital Software, China). PSB incidence was determined as the proportion of infected peanut plants

---

[4] https://tj.nanyang.gov.cn/2025/02-11/940805.html (accessed June 10, 2025).

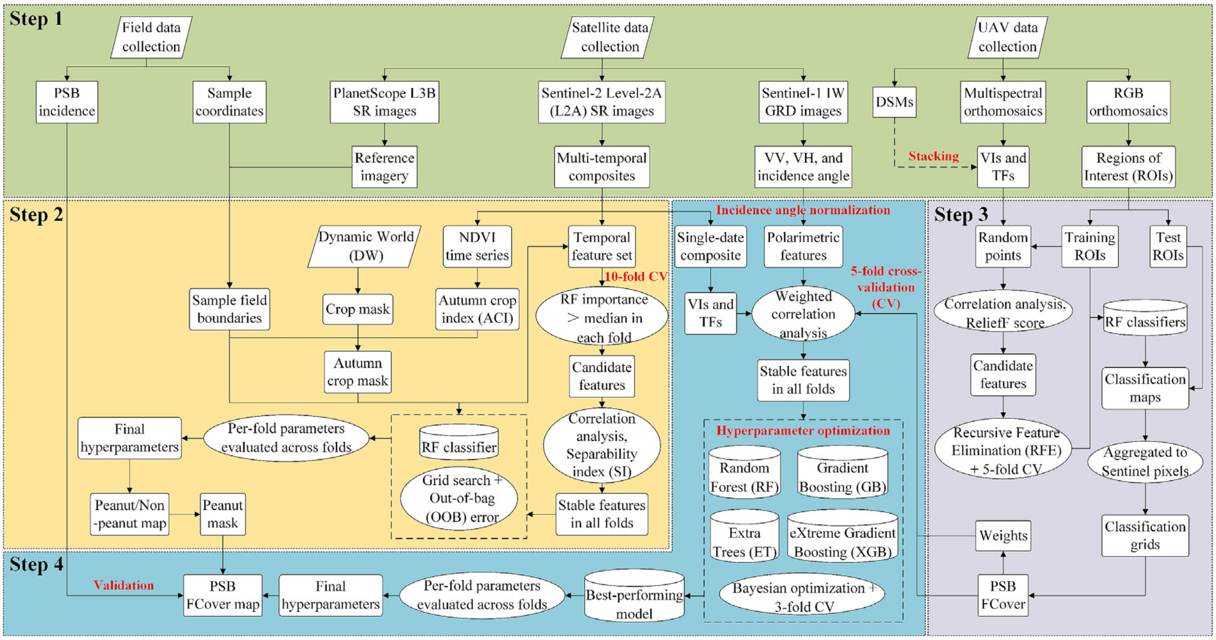


Fig. 3. Research workflow for PSB FCover mapping. Note: Abbreviations appearing in the figure, including L3B, SR, IW, and GRD, are explained later in the main text.

exhibiting white mycelia on the basal stem at the soil surface within a 10 m × 10 m plot centered on each peanut sample.

2.2.1.2. Satellite data

2.2.1.2.1. PlanetScope data. Forty-two PlanetScope Level-3B (L3B)[5] surface reflectance (SR) 4-band (NIR and RGB) images with cloud cover below 10%, acquired on Aug 13, 2024, were accessed, mosaicked, and clipped via Planet Explorer[6] to generate reference imagery covering 98.49% of the study area. These images were captured by SuperDove sensors onboard multiple CubeSats operated by Planet Labs, with a spatial resolution of 3 m (Kim et al., 2022; Rahali et al., 2025).

2.2.1.2.2. Sentinel-2 data. Between Jun 11 and Oct 20, 2024, Sentinel-2 A/B (S2 A/B) satellites provided an average of 33.2 total observations per pixel over the study area, of which 14.3 were good (unoccluded) (Fig. 2b), indicating that over half of the observations were affected by cloud cover. To address this, Sentinel-2 Level-2 A (L2A) SR 10-band (Table 1) images were accessed from the ImageCollection ("COPERNICUS/S2_SR_HARMONIZED"). Bands originally at 20 m spatial resolution (B5, B6, B7, B8A, B11, and B12) were resampled to 10 m using nearest-neighbor interpolation to harmonize all bands for subsequent feature extraction. The images were then processed in GEE through three steps:

- Cloud masking: Each L2A image was linked via unique IDs to the Cloud Score + S2_HARM ONIZED ImageCollection ("GOOGLE/CLOUD_SCORE_PLUS/V1/S2_HARMONIZED"), and a per-pixel clear likelihood band (cs_cdf) was added. Within the recommended range of cs_cdf (0.50–0.65),[7] we tested four thresholds (0.50, 0.55, 0.60, and 0.65) to evaluate cloud masking performance. As shown in Fig. 4, lower thresholds retained more pixels but left residual cloud contamination, while higher thresholds more effectively removed thin clouds and cloud shadows, albeit at the cost of reduced valid data coverage (Pasquarella et al., 2023). Based on these comparisons, we selected a cs_cdf threshold of 0.6, such that pixels with cs_cdf < 0.6 were masked, achieving a balance between cloud masking efficacy and valid data coverage.
- Temporal compositing: Cloud-masked L2A images within each 10-day window between Jun 11 and Oct 20, 2024, were aggregated on a pixel-by-pixel basis. The median was calculated independently for each band, producing one composite image per window to reduce residual cloud contamination while preserving valid data coverage. The composite date was assigned as the middle date of each 10-day window, resulting in fourteen composites on Jun 16, 26; Jul 06, 16, 26; Aug 05, 15, 25; Sep 04, 14, 24; and Oct 04, 14, 24, 2024.
- Gap filling: Residual gaps were filled using linear interpolation (LI), calculated as (Che et al., 2024):

$$\rho_t = \rho_{t_{\text{prev}}} + \frac{\rho_{t_{\text{next}}} - \rho_{t_{\text{prev}}}}{t_{\text{next}} - t_{\text{prev}}} \times (t - t_{\text{prev}}) \tag{1}$$

where ρ_t is missing pixel value on date t; $\rho_{t_{\text{prev}}}$ and $\rho_{t_{\text{next}}}$ are nearest valid pixel values within preceding and following three composites.

The potential impact of LI on time-series data quality was evaluated using NDVI as an indicator. As shown in Fig. 5, several 10-day composites exhibited substantial data gaps prior to interpolation, with missing-pixel proportions exceeding 60% on some dates (e.g., Jul 16, Sep 14, and Oct 14). After applying LI, these gaps were largely eliminated, resulting in spatially continuous NDVI composites. To further assess temporal consistency, NDVI time series were extracted from five evenly distributed points across the study area. Before interpolation, severe data gaps led to missing values in the NDVI trajectories, whereas

Table 1
Details of 10 Sentinel-2 spectral bands used in this study.

Band number	Description	Central wavelength (nm)	Spatial resolution (m)
B2	Blue	496.6 (S2A) / 492.1 (S2B)	10
B3	Green	560 (S2A) / 559 (S2B)	10
B4	Red	664.5 (S2A) / 665 (S2B)	10
B5	RE 1	703.9 (S2A) / 703.8 (S2B)	20
B6	RE 2	740.2 (S2A) / 739.1 (S2B)	20
B7	RE 3	782.5 (S2A) / 779.7 (S2B)	20
B8	NIR	835.1 (S2A) / 833 (S2B)	10
B8A	Narrow NIR	864.8 (S2A) / 864 (S2B)	20
B11	SWIR 1	1613.7 (S2A) / 1610.4 (S2B)	20
B12	SWIR 2	2202.4 (S2A) / 2185.7 (S2B)	20

SWIR: Short-wave infrared.

---

[5] https://earth.esa.int/eogateway/missions/planetscope (accessed June 15, 2025).

[6] https://www.planet.com/explorer/ (accessed June 30, 2025).

[7] https://developers.google.com/earth-engine/datasets/catalog/GOOGLE_CLOUD_SCORE_PLUS_V1_S2_HARMONIZED (accessed June 30, 2025).

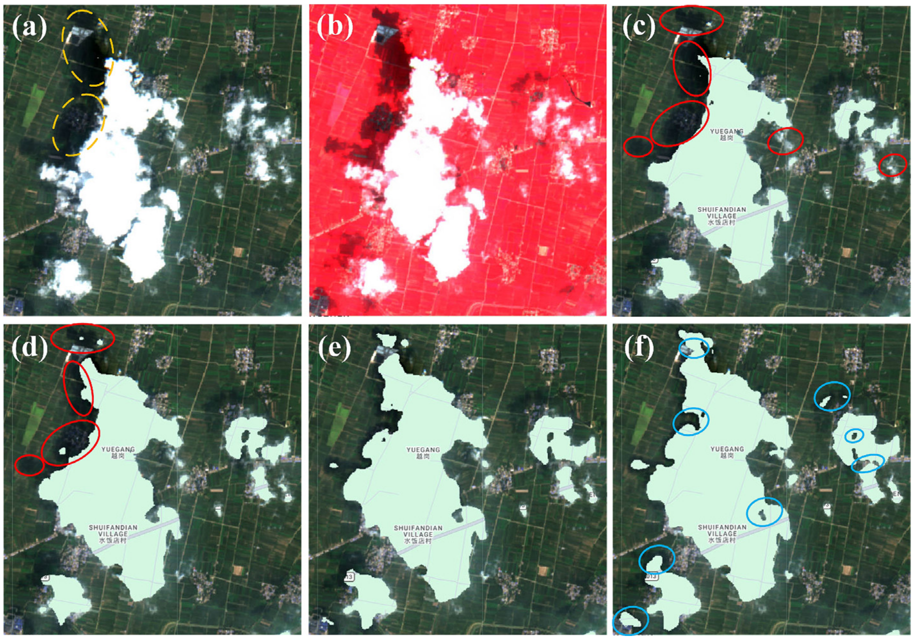


Fig. 4. Comparison of images before and after cloud masking. (a) Original RGB image. (b) False-color image (NIR, Red, and Green). (c–f) Images masked at cs_cdf thresholds of 0.50, 0.55, 0.60, and 0.65, respectively. Note: Orange, red, and blue circles indicate cloud shadows, residual clouds or shadows not masked, and over-masked pixels, respectively. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

after interpolation the NDVI curves became continuous and exhibited smooth, phenologically reasonable temporal patterns. These results indicate that LI effectively mitigated data gaps while preserving the underlying seasonal dynamics of land cover (LC).

2.2.1.2.3. Sentinel-1 data. Two Sentinel-1 Interferometric Wide-swath (IW) Ground Range Detected (GRD) 3-band (VV, VH, and incidence angle) images, acquired on Sep 14, 2024, were accessed from the ImageCollection ("COPERNICUS/S1_GRD") in GEE (Fig. 2a). These images, captured by dual-polarization C-band Synthetic Aperture Radar (SAR) onboard Sentinel-1 A on ascending orbits, were preprocessed using the standard Sentinel-1 Toolbox workflow, which included orbital file correction, GRD border and thermal noise removal, radiometric calibration, and terrain correction, before mosaicking and clipping to the study area (Yang et al., 2023). The spatial resolution is 10 m. To correct for incidence angle variability, all backscatter values were converted from dB to linear scale and divided by the cosine of

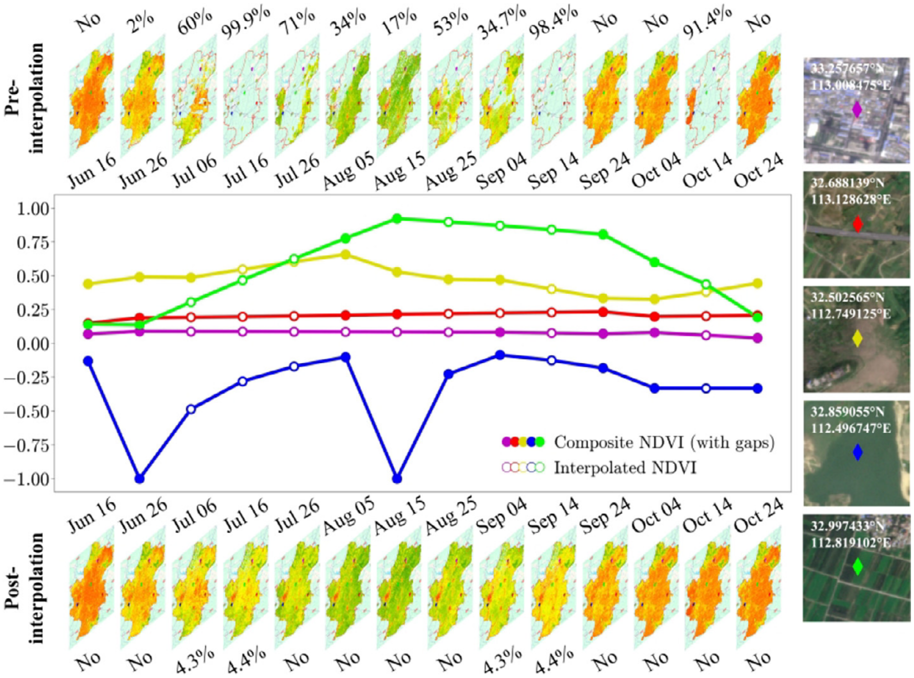


Fig. 5. Comparison of NDVI composites and time series of five evenly distributed points within the study area, pre- and post-interpolation. Note: 'No' and percentages indicate no missing pixels and the proportion of missing pixels, respectively.

the local incidence angle (i.e., incidence angle normalization) (Najem et al., 2024). Subsequent polarimetric indices (VV_plus_VH, VV_minus_VH, VH_VV_ratio, and ND_VVVH) were calculated based on these incidence-angle–corrected values.

2.2.1.3. UAV data. During the field survey, a DJI Mavic 3 Multispectral UAV was deployed to acquire imagery over four typical peanut fields affected by PSB (labeled I-IV) (Fig. 2d). The UAV is equipped with a 4/3″ CMOS RGB camera (20 MP, 24 mm equivalent, 84 °FOV) and four 1/2.8″CMOS multispectral sensors (5 MP each, 25 mm equivalent, 73.9 °FOV), capturing the Green (560 ± 16 nm), Red (650 ± 16 nm), RE (730 ± 16 nm), and NIR (860 ± 26 nm) bands. The onboard RTK-GNSS system provides centimeter-level positioning accuracy (± 0.1 m), eliminating the need for ground control points (GCPs). Prior to flight, radiometric calibration panels were placed horizontally under unobstructed sunlight (solar elevation >30°) and imaged by the multispectral sensors. The UAV then flew at 12 m altitude with 80% forward and 70% side overlap.

The acquired images were processed in DJI Terra (version 3.6.8, Shenzhen Dajiang Innovation Technology, China) for radiometric calibration and stitching, generating the RGB orthomosaics, multispectral orthomosaics, and DSMs with spatial resolutions of 0.35, 0.59, and 0.69 cm, respectively. In addition, five VIs—NDVI, Green Normalized Difference Vegetation Index (GNDVI), Normalized Difference RE Index (NDRE), Leaf Chlorophyll Index (LCI), and Optimized Soil Adjusted Vegetation Index (OSAVI)—were automatically calculated by the software based on the multispectral orthomosaics, which support the assessment of plant health, biomass, and chlorophyll content (Gitelson and Merzlyak, 1994; Gitelson et al., 1996; Pu et al., 2008; Rondeaux et al., 1996; Tucker, 1979).

To quantify geometric misregistration between UAV mosaics and Sentinel-2 imagery, five ground control points (GCPs) per field—four corners and the center—were manually identified at corresponding locations in both UAV RGB orthomosaics and the Sentinel-2 composite (Red, Green, and Blue; Sep 04). For each field, geometric residuals were computed as the Euclidean distance between corresponding UAV and Sentinel-2 GCPs, and the root mean square (RMS) error of the five GCPs quantified the field's overall misregistration. RMS errors for the four fields were 2.47 m, 2.28 m, 1.21 m, and 3.76 m, all below the 10 m Sentinel-2 pixel size, indicating minimal misregistration. To correct remaining misalignment, an affine transformation matrix was derived from the GCPs for each field and applied to all UAV-derived products (RGB orthomosaic, multispectral and VI orthomosaics, and DSM), ensuring spatial co-registration with Sentinel-2 imagery.

Subsequently, eight TFs—Mean (MEA), Variance (VAR), Homogeneity (HOM), Contrast (CON), Dissimilarity (DIS), Entropy (ENT), Second Moment (SM), and Correlation (COR)—were extracted from the multispectral orthomosaics in ENVI (version 5.6, Harris Geospatial Solutions, USA) using the Co-occurrence Measures tool with a 3 × 3 window. Finally, DSMs were resampled to 0.59 cm to match the spatial resolution of the multispectral orthomosaics and stacked with VIs and TFs.

Polygonal regions of interest (ROIs) for the six LC classes, including healthy peanut, diseased peanut, maize residue, weeds, bare soil, and others (tree canopy, soybean, or stone), were manually delineated in ENVI based on the RGB orthomosaics, ensuring clear boundaries and sufficient within-class variability (Table 2).

2.2.2. Extraction of peanut spatial distribution

2.2.2.1. Autumn crop mask generation. Dynamic World (DW), available in GEE (ImageCollection: “GOOGLE/DYNAMICWORLD/V1”), is a 10-m NRT (updated every 2–5 days) LC dataset produced from Sentinel-2 Level-1C images with CLOUDY_PIXEL_PERCENTAGE ≤ 35% (Wang et al., 2023c; Xu et al., 2024). It provides pixel-level predictions of nine LC classes, each assigned a label: water (0), trees (1), grass (2), flooded vegetation (3), crops (4), shrub and scrub (5), built area (6), bare ground (7), and snow and ice (8) (Brown et al., 2022). In this study, an LC composite was generated by mode-aggregating DW class labels between Jun 11 and Oct 20, 2024, from which only pixels labeled 4 were retained as the crop mask.

NDVI time series were calculated from Sentinel-2 composites for Autumn Crop Index (ACI) development, defined as follows (Qiu et al., 2017):

Table 2
Number of LC ROIs and example images.

LC class	Field I	Field II	Field III	Field IV	Example image
Healthy peanut	248	215	119	460	
Diseased peanut	251	204	235	500	
Maize residue	207	111	165	241	
Weeds	223	137	160	133	
Bare soil	117	178	100	276	
Others	85	93	0	103	

Note: Example image of ‘Others’ class is soybean.

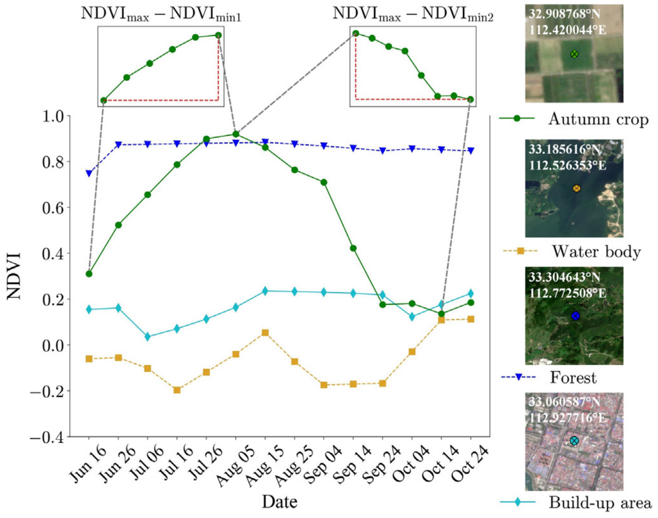


Fig. 6. NDVI time-series profiles of four LC types. Note: Autumn crop is represented by peanut to illustrate its NDVI variations.

$$\text{ACI} = (\text{NDVI}_{\text{max}} - \text{NDVI}_{\text{min1}}) \times (\text{NDVI}_{\text{max}} - \text{NDVI}_{\text{min2}}) \tag{2}$$

where NDVI_{max} is global NDVI maximum occurring between two local NDVI minima ($\text{NDVI}_{\text{min1}}$ and $\text{NDVI}_{\text{min2}}$). Fig. 6 illustrates how ACI enhances the NDVI signal of autumn crops.

ACI was computed for each pixel within the crop mask, and a threshold range—defined as the minimum and maximum ACI values within sample field boundaries manually delineated in ENVI based on in-situ sample points and PlanetScope reference imagery (Fig. 2c)—was applied to retain pixels within this range, generating the autumn crop mask.

2.2.2.2. Feature construction and selection. Spectral features (Table 3) for each Sentinel-2 composite were calculated and combined into a temporal feature set. These features were chosen to capture key crop traits, including Sentinel-2 composite bands, VIs (DVI, EVI, GCC, RDVI) sensitive to vegetation biochemistry and biomass, water-sensitive indices (GWCCI, LSWI), and chlorophyll- and soil-related indices (MCARI, NDSI, S2REP), all previously reported as informative for crop classification and monitoring. By integrating these features across time, the resulting spectral-temporal set is intended to support peanut detection throughout the growing season.

Feature values and corresponding labels (peanut: 2; non-peanut crops: 3) were extracted from pixels within sample field boundaries to construct a classification dataset. The dataset was randomly and equally partitioned into 10 subsets for cross-validation (CV), with one subset used as the validation set and the remaining nine subsets as the training set in each fold. Feature selection was fully nested within the CV framework and conducted exclusively on the training set of each fold to avoid information leakage. A two-step feature selection procedure was applied. First, within each training set, features were ranked by RF importance, and those with importance values above the fold-specific median were retained (Breiman, 2001). Only features consistently retained across all 10 folds were carried forward as candidate features. Second, redundant features were further removed within each training set using the Separability Index (SI; Eq. (3)) and Pearson correlation analysis (Benesty et al., 2009; Somers and Asner, 2013). The SI threshold was set based on the distribution of SI values among candidate features. Features with SI values below the threshold were first excluded, and then for feature pairs with absolute Pearson correlation coefficients ($|r|$) > 0.9, the feature with the higher SI was retained. Stable features consistently retained across all 10 folds after redundancy removal were subsequently used for hyperparameter optimization.

$$\text{SI} = \frac{|\bar{x}_1 - \bar{x}_2|}{1.96 \times (s_1 + s_2)} \tag{3}$$

where $\bar{x}_1$ and $\bar{x}_2$ are means of specific feature for peanut and non-peanut crops, respectively; s_1 and s_2 are corresponding standard deviations. A higher SI suggests stronger class separability for the feature.

2.2.2.3. Model development and mapping. For each fold, an RF classifier was constructed based on selected features and corresponding labels from the training set (Belgiu and Drăguţ, 2016). To optimize model performance, a grid search was conducted over combinations of n_estimators (10, 50, 100, 150, 200, 500, and 1000) and max_features ('sqrt', 'log2', or integers from 1 to N, where N is the number of selected features). For each combination, the classifier was trained using bootstrap sampling and evaluated via out-of-bag (OOB) error, which estimates generalization performance without requiring a separate validation set (Bylander, 2002). The combination yielding the lowest OOB error in each fold was selected as that fold's optimal hyperparameter set, which was then applied to the corresponding validation set to evaluate model performance using OA and Kappa coefficient.

$$\text{OA} = \frac{\sum_{i=1}^{k} n_{ii}}{N} \tag{4}$$

$$\text{Kappa} = \frac{P_o - P_e}{1 - P_e}, P_o = \text{OA}, P_e = \frac{\sum_{i=1}^{k} (n_{i+} \times n_{+i})}{N^2} \tag{5}$$

where k and N are total number of classes and samples, respectively; n_{ii}, n_{i+}, and n_{+i} are number of correctly classified samples, total actual samples, and total predicted samples of class i, respectively.

To determine the final hyperparameter set for mapping, each fold's optimal set was applied to train the classifier across all 10 folds, and performance was evaluated on each fold's validation set. The mean and standard deviation (std) of OA and Kappa across all folds were

Table 3
Summary of spectral features.

Feature type	Feature name	Formula or descriptor	Reference
SR	B2, B3, B4, B5, B6, B7, B8, B8A, B11, B12	Sentinel-2 composite bands	–
Derived indices	Difference Vegetation Index (DVI)	$B8 - B4$	Roujean and Breon (1995)
	Enhanced Vegetation Index (EVI)	$2.5 \times \frac{B8 - B4}{B8+6B4 - 7.5B2+1}$	Huete et al. (2002)
	Green Chromatic Coordinate (GCC)	$\frac{B3}{B4+B3+B2}$	Gillespie et al. (1987)
	Greenness and Water Content Composite Index (GWCCI)	$B11 \times \frac{B8 - B4}{B8+B4}$	Chen et al. (2023a)
	Land Surface Water Index (LSWI)	$\frac{B8 - B11}{B8+B11}$	Xiao et al. (2005)
	Modified Chlorophyll Absorption Ratio Index (MCARI)	$(B5 - B4) - \frac{0.2\times(B5 - B3)\times B5}{B4}$	Daughtry et al. (2000)
	Normalized Difference Soil Index (NDSI)	$\frac{B12 - B3}{B12+B3}$	Deng et al. (2015)
	Renormalized Difference Vegetation Index (RDVI)	$\frac{B8 - B4}{\sqrt{B8+B4}}$	Roujean and Breon (1995)
	Sentinel-2 RE Position (S2REP)	$705 + \frac{35\times(0.5\times(B7+B4) - B5)}{B6 - B5}$	Frampton et al. (2013)

calculated, and the set with the highest mean OA and Kappa and relatively low stds was used to train the final classifier, which was then applied to the entire dataset for peanut/non-peanut crop mapping across the autumn crop mask (Section 2.2.2.1).

2.2.3. UAV-based classification and PSB FCover estimation

For each field, ROIs were stratified by LC class and randomly split 80/20 into training and test sets. Within the training ROIs, three random points per ROI were generated using the Create Random Points tool in ArcMap (version 10.8, Esri, USA), and values of UAV-derived VIs, TFs, and DSM were extracted at these points. Feature selection was performed in two stages. First, redundant feature pairs ($|r| > 0.9$) were identified by correlation analysis, and for each pair, the feature with the higher ReliefF importance score was retained. ReliefF evaluates features based on their ability to distinguish neighboring samples from different classes and is robust to noise, making it suitable for multiclass problems (Kononenko, 1994; Robnik-Šikonja and Kononenko, 2003; Urbanowicz et al., 2018). Second, candidate features from the previous stage were further refined using Recursive Feature Elimination (RFE) with 5-fold CV (Liu et al., 2024b). Specifically, an RF classifier served as the base model. Starting with the full set of candidate features, the least important feature, based on the classifier's feature importance score, was removed at each iteration. The classifier was then retrained on the reduced subset. This process was repeated until all features were eliminated and the subset achieving the highest mean CV accuracy across all 5 folds was selected as the optimal combination.

Selected features and training ROIs were then used to train an RF classifier in ENVI (n_estimators = 100, criterion = 'gini') to generate the classification map. Classification accuracy was evaluated using test ROIs, with metrics including OA (Eq. (4)), omission error (OE; Eq. (6)), commission error (CE; Eq. (7)), and F1-score (F1; Eq. (8)).

$$\mathrm{OE}_i = 1 - \mathrm{PA}_i = 1 - \frac{T_{ii}}{\sum\limits_{j=1}^{k} T_{ij}} = \frac{\sum\limits_{j=1, j\neq i}^{k} T_{ij}}{\sum\limits_{j=1}^{k} T_{ij}} \tag{6}$$

$$\mathrm{CE}_i = 1 - \mathrm{UA}_i = 1 - \frac{T_{ii}}{\sum\limits_{j=1}^{k} T_{ji}} = \frac{\sum\limits_{j=1, j\neq i}^{k} T_{ji}}{\sum\limits_{j=1}^{k} T_{ji}} \tag{7}$$

$$\mathrm{F1}_i = 2 \times \frac{\mathrm{UA}_i \times \mathrm{PA}_i}{\mathrm{UA}_i + \mathrm{PA}_i} \tag{8}$$

where PA_i and UA_i are producer's accuracy and user's accuracy of class i, respectively; T_{ii} is number of correctly classified pixels of class i; $\sum\limits_{j=1}^{k} T_{ij}$ and $\sum\limits_{j=1}^{k} T_{ji}$ are total number of pixels actually belonging to and classified as class i, respectively; $\sum\limits_{j=1, j\neq i}^{k} T_{ij}$ and $\sum\limits_{j=1, j\neq i}^{k} T_{ji}$ are number of pixels actually belonging to class i but misclassified and classified as class i but belonging to other classes, respectively.

PSB FCover was estimated from the UAV-based classification map as the proportion of diseased peanut pixels relative to the total number of UAV pixels that theoretically cover each Sentinel pixel:

$$\mathrm{FCover}_i = \frac{\mathrm{DPP}_i}{2872738} \tag{9}$$

where DPP_i is number of diseased peanut pixels within Sentinel pixel i. Each Sentinel pixel thus served as a sample.

Although the FCover samples were obtained from only four fields, these fields are located in the northeast and southwest of the study area (Fig. 2c), which are major peanut production zones and exhibit variation in environmental conditions such as soil type, climate, and management practices. This ensures that these FCover samples are representative of the region's peanut-growing conditions and suitable for regional-scale modeling.

Considering the differences in sample counts across fields and variability of FCover values within each field, each Sentinel-pixel FCover sample was assigned a weight:

$$\omega_i = \frac{1}{A_i} \times \sqrt{\mathrm{FCover}_i + 0.1} \times 10 \tag{10}$$

where A_i is total number of FCover samples (Sentinel pixels) in the field containing sample i. This weighting prevents larger fields from dominating the training and gives proportionally more influence to samples with higher FCover values.

2.2.4. Modeling the relationship between satellite-derived features and PSB FCover

The associated features (Table 4) were constructed using Sentinel-1 SAR imagery (Sep 14) and Sentinel-2 composite (Sep 04) close to the field survey period to ensure temporal correspondence with PSB FCover. These features were chosen to capture complementary aspects of peanut canopy condition relevant to PSB infestation. VIs (e.g., TNDVI, PSRI, and NDRE2) are sensitive to canopy greenness, chlorophyll content, and stress responses associated with PSB damage. TFs (e.g., COR and ENT) quantify spatial heterogeneity in the peanut canopy, reflecting the uneven distribution of disease symptoms. In addition, Sentinel-1 polarimetric features (e.g., VH) provide structural and moisture-related information that complements the optical data. Together, these features provide a comprehensive representation of peanut health and PSB infestation patterns. Feature values were extracted from the Sentinel-1/2 pixels corresponding to the FCover samples to construct a weighted regression dataset. To mitigate the risk that the model learns field-specific background patterns rather than PSB-related signals, the dataset was randomly and equally partitioned into 5 subsets for CV, with samples from all four fields evenly distributed across the subsets. In each fold, one subset was used as the validation set and the remaining four subsets as the training set, taking sample weights into account. While leave-one-field-out validation could more strictly assess generalization to unseen fields, using one entire field as the test set would result in very small and uneven training sets given the limited number of fields, which could lead to unstable performance. Therefore, 5-fold CV represents a practical compromise given the available data.

In each training set, weighted Pearson correlations (r_{weighted}) between features and FCover were computed using sample weights. Features with $|r_{\mathrm{weighted}}| < 0.2$ were removed. For feature pairs with $|r| > 0.9$, only the feature more strongly correlated with FCover was then retained. r_{weighted} is defined as:

$$r_{\mathrm{weighted}} = \frac{\sum_{i=1}^{n} \omega_i (X_i - \overline{X}_\omega)(Y_i - \overline{Y}_\omega)}{\sqrt{\sum_{i=1}^{n} \omega_i (X_i - \overline{X}_\omega)^2} \sqrt{\sum_{i=1}^{n} \omega_i (Y_i - \overline{Y}_\omega)^2}} \tag{11}$$

$$X_\omega = \frac{\sum_{i=1}^{n} \omega_i X_i}{\sum_{i=1}^{n} \omega_i}, Y_\omega = \frac{\sum_{i=1}^{n} \omega_i Y_i}{\sum_{i=1}^{n} \omega_i} \tag{12}$$

where X_i and Y_i are values of features and FCover, respectively, for sample i; $\overline{X}_\omega$ and $\overline{Y}_\omega$ are weighted means of feature X and FCover Y, respectively; n is number of samples. This approach ensures features are evaluated based on their significance to FCover, considering sample weighting.

Stable features consistently retained across all 5 folds were subsequently used for hyperparameter optimization based on four ensemble learning models: RF, Extra Trees (ET), Gradient Boosting (GB), and eXtreme Gradient Boosting (XGB). These ensemble methods are particularly effective in modeling nonlinear relationships, managing

Table 4
Summary of Sentinel-1/2-derived features.

Feature type	Feature name	Formula or descriptor	Source	Reference
VIs	NDVI	$\frac{B8-B4}{B8+B4}$	Sentinel-2	Tucker (1979)
	GCC	see Table 3		Gillespie et al. (1987)
	LSWI	see Table 3		Xiao et al. (2005)
	MCARI	see Table 3		Daughtry et al. (2000)
	RDVI	see Table 3		Roujean and Breon (1995)
	RE Chlorophyll Index (RECI)	$\frac{B8}{B5}-1$		Gitelson et al. (2003)
	Excess Green Vegetation Index (ExG)	$2B3-B4-B2$		Mao et al. (2003)
	Inverted RE Chlorophyll Index (IRECI)	$\frac{(B7-B4)\times B6}{B5}$		Frampton et al. (2013)
	Modified Simple Ratio (MSR)	$\frac{B8/B4-1}{\sqrt{B8/B4+1}}$		Chen (1996)
	Normalized Difference Infrared Index (NDII)	$\frac{B8-B12}{B8+B12}$		Hardisky et al. (1983)
	Normalized Difference RE 1 Index (NDRE1)	$\frac{B8-B5}{B8+B5}$		Gitelson and Merzlyak (1994)
	Normalized Difference RE 2 Index (NDRE2)	$\frac{B8-B6}{B8+B6}$		Gitelson and Merzlyak (1994)
	Normalized Difference RE 3 Index (NDRE3)	$\frac{B8-B7}{B8+B7}$		Gitelson and Merzlyak (1994)
	Plant Senescence Reflectance Index (PSRI)	$\frac{B4-B3}{B6}$		Merzlyak et al. (1999)
	Transformed Normalized Difference Vegetation Index (TNDVI)	$\sqrt{\frac{B8-B4}{B8+B4}+0.5}$		Tucker (1979)
TFs	Sum Average (SA), VAR, Inverse Difference Moment (IDM), CON, DIS, ENT, Angular Second Moment (ASM), COR	GLCM	Sentinel-2	Haralick et al. (1973)
Polarimetric	VV	VV backscatter coefficients (incidence-angle–corrected)	Sentinel-1	–
	VH	VH backscatter coefficients (incidence-angle–corrected)		–
	VV_plus_VH	$VV+VH$		–
	VV_minus_VH	$VV-VH$		–
	VH_VV_ratio	$\frac{VH}{VV}$		–
	ND_VVVH	$\frac{VV-VH}{VV+VH}$		–

Note: TFs were extracted from Sentinel-2 composite bands using the glcmTexture function in GEE with a 3 × 3 window.

multicollinearity, and handling high-dimensional feature spaces, thereby improving generalization and reducing the risk of overfitting (Dietterich, 2000; Huang et al., 2025; Zhou, 2012). RF and ET follow the bagging strategy, constructing multiple decision trees in parallel to reduce variance (Breiman, 1996). ET introduces additional randomness by randomly selecting split points for each feature, further reducing variance at the cost of interpretability (Breiman, 2001; Geurts et al., 2006). In contrast, GB and XGB adopt the boosting strategy, sequentially adding trees to minimize errors from prior learners (Friedman, 2001). XGB improves upon GB by incorporating regularization and efficient techniques, such as parallel tree construction and sparsity-aware learning, which enhance scalability and predictive robustness (Chen and Guestrin, 2016).

To optimize model performance, hyperparameter tuning was conducted using Bayesian optimization in each training set. As a global optimization algorithm, it probabilistically models the objective function, efficiently exploring the search space to identify optimal configurations with fewer evaluations (Snoek et al., 2012). For each set of hyperparameters, 3-fold CV was used to assess performance, improving stability and reducing selection bias. The objective function was defined as the average CV negative mean squared error (−MSE), with sample weights considered in its calculation, allowing Bayesian optimization to act as a maximization procedure. Each model underwent 100 trials, with the configuration yielding the highest average − MSE selected as optimal. MSE was computed as:

$$\mathrm{MSE}=\frac{1}{n}\sum_{i=1}^{n}(y_i-\widehat{y}_i)^2 \tag{13}$$

where y_i and $\widehat{y}_i$ are estimated and predicted FCover values, respectively; n is number of samples in each validation set for 3-fold CV.

To identify the best-performing model, the coefficient of determination (R^2; Eq. (14)) and root mean squared error (RMSE; Eq. (15)) were computed for each model on its corresponding validation set using the optimal hyperparameters from the five training sets. The overall R^2 and RMSE across all validation sets were then calculated to summarize predictive performance and compare the four ensemble models.

$$R^2=1-\frac{\sum_{i=1}^{n}(y_i-\widehat{y}_i)^2}{\sum_{i=1}^{n}(y_i-\overline{y}_i)^2} \tag{14}$$

$$\mathrm{RMSE}=\sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i-\widehat{y}_i)^2} \tag{15}$$

where $\overline{y}_i$ is mean of estimated FCover values; n is number of samples in each validation set for 5-fold CV.

To determine the final hyperparameters for PSB FCover mapping, the optimal hyperparameters from each fold were applied to train the selected model across all 5 folds, considering sample weights. The overall R^2 and RMSE were calculated across all validation sets. System error was further assessed using residuals (i.e., the difference between estimated and predicted FCover values), across both the overall and extreme FCover ranges. The hyperparameters that maximized R^2, minimized RMSE, and resulted in a mean residual (μ) close to 0 with a relatively small standard deviation (σ) were selected and used to train the weighted model on the entire dataset for mapping. The mapping results were subsequently validated through PSB incidence (Section 2.2.1.1).

3. Results

3.1. Peanut spatial distribution mapping

Within the crop mask, ACI was calculated for each pixel, and only those with values between 0.267 and 0.604 were retained to generate the autumn crop mask (Fig. 7). Next, 83 candidate features were identified from the temporal feature set. As shown in Fig. 8, features from Jun

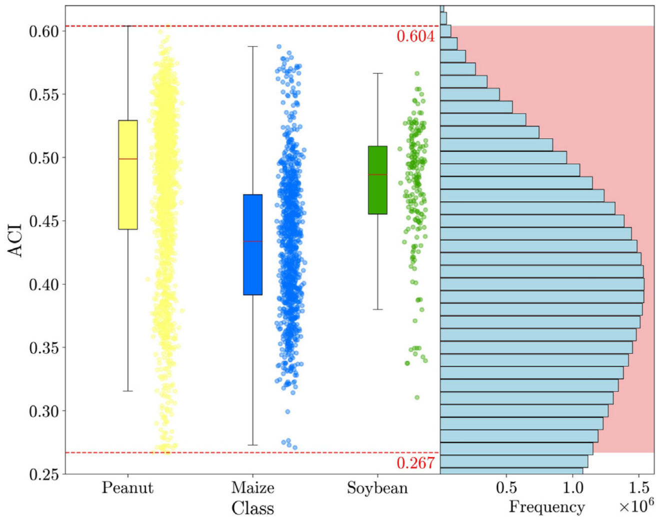


Fig. 7. Distribution of ACI values for pixels inside sample field boundaries.

16, Aug 05, Aug 15, and Aug 25 were most frequently retained across all composite dates, with cumulative selection frequencies of LSWI exceeded 7, highlighting its stable contribution to RF performance. In contrast, features from Sep 24, Oct 04, Oct 14, and Oct 24 were not retained, and among the retained features, SR and derived indices accounted for 34.9% and 65.1%, respectively, indicating that the latter played a dominant role in model performance. SI values of these candidate features were then calculated within the training sets of 10-fold CV. As shown in Fig. 9, most SI values fell within the 0.3–0.5 range, with 74.7% exceeding 0.3, whereas only 44.6–47% exceeded 0.4 across folds. Therefore, a threshold of 0.3 was applied to retain the majority of discriminative features while removing weakly separable ones, as a higher threshold would exclude many informative features; this threshold was subsequently used to determine feature retention for pairs with $|r| > 0.9$. This criterion follows previous studies using SI (e.g., Liu et al., 2024b).

Following the above procedures, features were selected independently in each fold. Fig. 10 shows the number of folds in which each feature was retained, indicating how consistently each feature was selected across all 10 folds. Based on these 30 features that were consistently retained across all folds, fold-specific RF classifiers were trained and evaluated, and the resulting optimal hyperparameter sets are listed in Table 5. Among these, the 150–8 combination achieved the highest mean OA (0.9845) and Kappa (0.9678) with relatively low stds (0.0066 and 0.0137, respectively) in the 10-fold CV (Fig. 11), and was therefore selected to train the final RF classifier. This classifier was applied across the autumn crop mask to extract the spatial distribution of peanut and non-peanut crops (Fig. 2c), with the peanut distribution used to generate the peanut area mask.

3.2. UAV-based classification and PSB FCover estimation

In Fields I-IV, 23, 21, 19, and 22 candidate features were filtered from UAV-derived features (VIs + TFs + DSM) based on correlation analysis and ReliefF importance scores, respectively, and further refined by RFE with 5-fold CV to retain the subset that maximized the mean CV accuracy of the RF base model (Fig. 12). Selected features for each field, along with their corresponding ReliefF importance scores, are shown in Fig. 13.

These selected features, together with training ROIs, were used to train the RF classifiers for classification mapping (Fig. 14), with accuracies evaluated on test ROIs. As shown in Table 6, performance differed among LC classes. Peanut classes (healthy and diseased) achieved F1s of 80.3%–92.7%, with higher CEs and OEs for diseased peanut in Fields II and III, reflecting misclassification of diseased pixels under variable field conditions. Bare soil was consistently well-classified in Fields I-III (F1 = 95.9–97.3%, CEs and OEs ≤ 7%), but performance declined markedly in Field IV (F1 = 70.3%, CE = 30.7%, OE = 28.8%) due to strong spectral confusion. In contrast, maize residue, weeds, and others showed more variable F1s (37.2–90.9%) and higher error rates, indicating misclassification among these spectrally similar or heterogeneous classes (Foody, 2002; Lillesand et al., 2015). OA ranged from 76.7% in Field IV to approximately 89% in Fields I-III. These results indicate that the overall classification was generally good, with peanut and bare soil classified consistently, whereas maize residue, weeds, and others were less reliably classified.

Based on UAV-based classification maps, PSB FCover was estimated for a total of 218 Sentinel pixels, with 46, 59, 27, and 86 pixels assigned to Fields I-IV, respectively. FCover weights were then calculated using Eq. (10). The FCover estimation procedure is shown in Fig. 15.

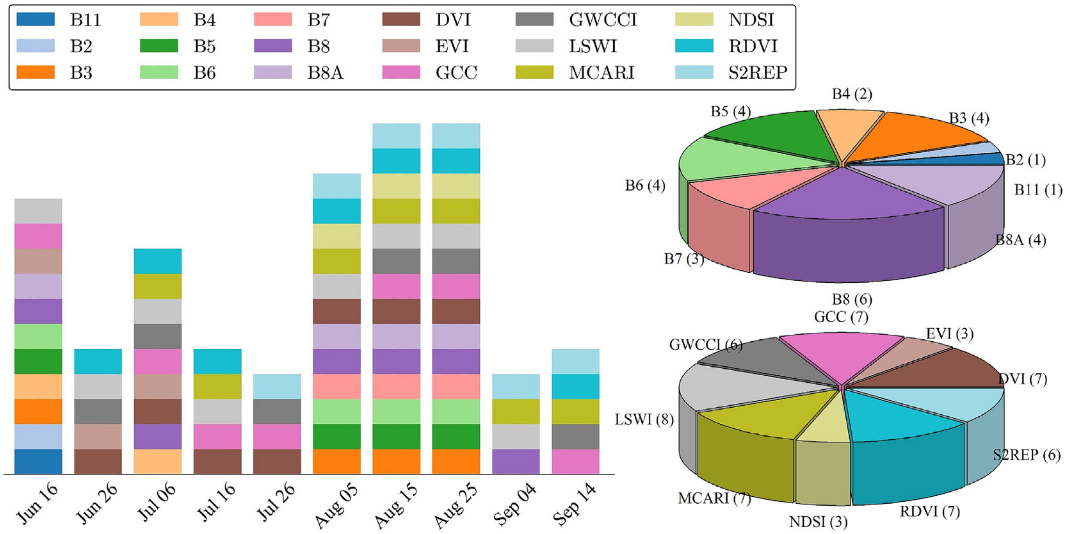


Fig. 8. Temporal distribution and cumulative selection frequencies of candidate features.

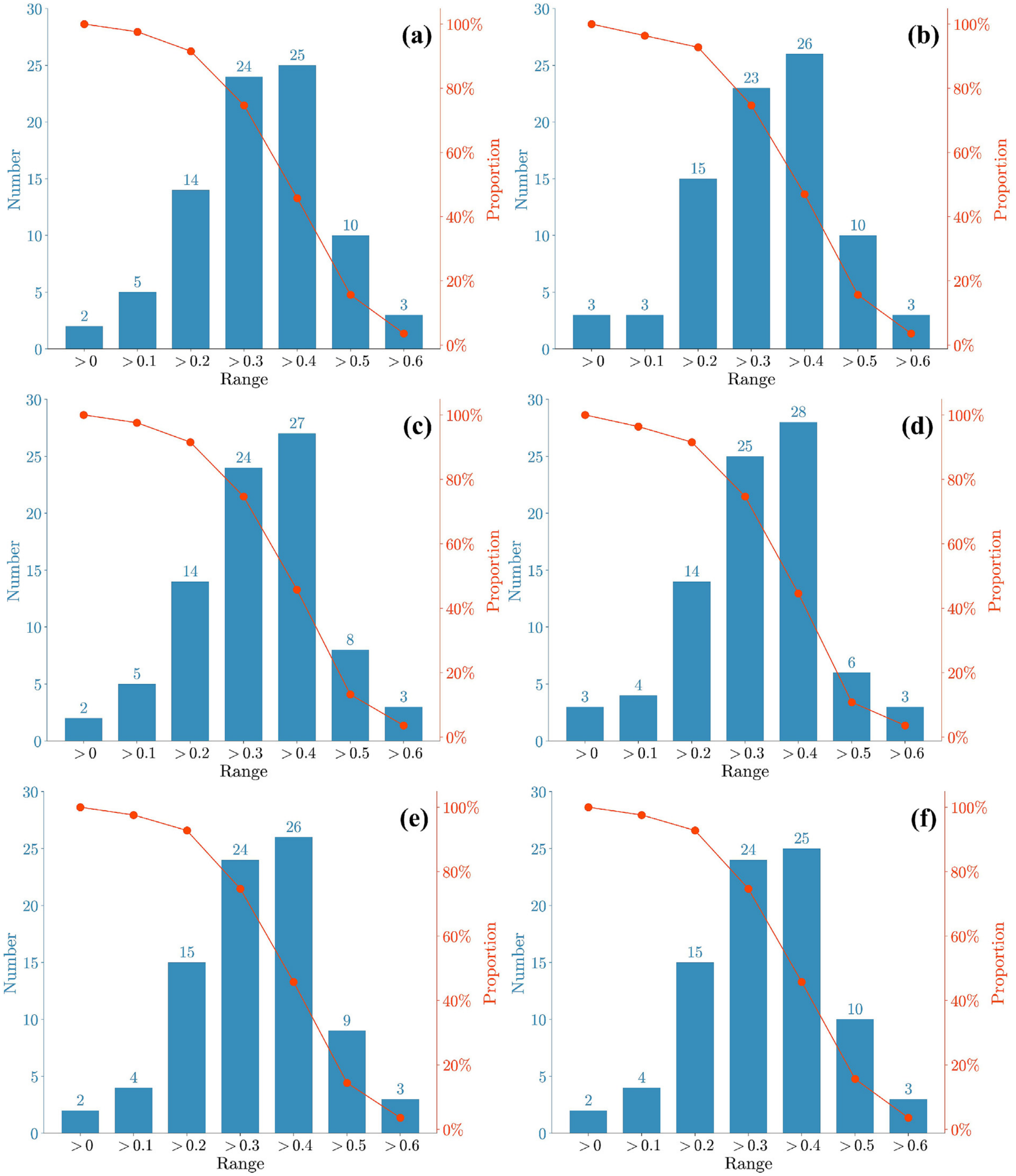


Fig. 9. Distribution and proportions of SI values for candidate features across all training sets of 10-fold CV. (a) Folds 1, 2, 5, 8, and 9. (b) Fold 3. (c) Fold 4. (d) Fold 6. (e) Fold 7. (f) Fold 10. Note: Bars with abscissa >0.3 represent the number of SI values between 0.3 and 0.4.

3.3. Modeling PSB FCover from satellite-derived features

Based on the weighted correlation analysis in the 5-fold CV training sets, features were selected in each fold (Figs. 16a-e). The number of folds in which each feature was retained is shown in Fig. 16f. Using the 17 stable features consistently retained across all 5 folds, hyperparameter optimization was performed for four ensemble models (RF, ET, GB, and XGB) within each training set, with FCover weights considered. Model performance was then evaluated on the corresponding validation sets, and the resulting optimal hyperparameters and performance are summarized in Table 7.

Hyperparameter optimization revealed distinct preferences across four models. RF favored a relatively large number of trees (median 530; range 185–753) and deeper trees (median depth 40; range 5–48), whereas XGB selected fewer trees (median 220; range 98–576) with shallower depths (median 5; range 3–40). GB and XGB preferred small learning rates (GB: 0.0314 [0.0138–0.0935]; XGB: 0.0663 [0.0063–0.196]) and high subsample ratios (GB: 0.8417 [0.3851–0.9996]; XGB: 0.721 [0.3502–0.9821]). Feature sampling strategies also differed: across folds, RF and ET had median fractional values of 0.511 and 0.7759, respectively, whereas GB predominantly selected categorical options (auto, sqrt, log2). Predictive performance,

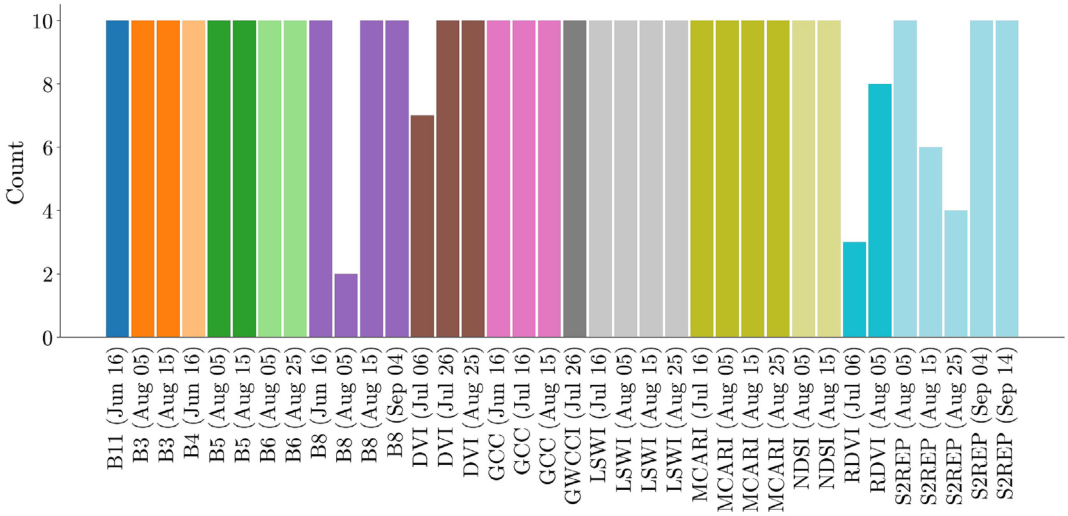


Fig. 10. Cumulative selection frequencies of the final features across 10-fold CV.

Table 5
Optimal hyperparameter set and model performance for each fold during CV.

Fold	n_estimators	max_features	OA	Kappa
1	150	log2	0.9670	0.9312
2	150	log2	0.9967	0.9931
3	500	9	0.9835	0.9658
4	500	7	0.9802	0.9590
5	500	sqrt	0.9901	0.9794
6	150	6	0.9868	0.9725
7	100	sqrt	0.9834	0.9658
8	150	8	0.9834	0.9656
9	150	8	0.9834	0.9659
10	500	9	0.9768	0.9521

Note: sqrt = square root of the number of features; log2 = log base 2 of the number of features.

summarized by R^2 and RMSE across folds, indicated that XGB achieved the highest accuracy (median $R^2 = 0.7412$, RMSE = 0.121) with relatively narrow fold-to-fold variation (R^2 range 0.6086–0.7511; RMSE 0.1128–0.1403), whereas RF exhibited the lowest median accuracy ($R^2 = 0.5103$, RMSE = 0.1619) and the largest fold-to-fold variability. Across all validation sets, XGB achieved an overall R^2 of 0.718 and RMSE of 0.123, outperforming all other models, thereby confirming XGB as the best-performing model (Fig. 17).

Analysis of cross-fold XGB validation residuals revealed consistent biases at extreme FCover values (Fig. 18). In the very low range (<10%), all fold-specific hyperparameter configurations exhibited negative mean residuals, with fold 1 showing the largest overestimation ($\mu = -9.2\%$, $\sigma = 6.8\%$). In the very high range (>60%), all configurations exhibited positive mean residuals, with fold 1 showing the greatest underestimation ($\mu = 12.7\%$, $\sigma = 10.5\%$). Despite these extreme-value biases, the overall mean residuals across folds remained near 0, with standard deviations between 12.1% and 12.9%, indicating minimal impact on overall model performance. Among all configurations, fold 5 achieved the highest overall R^2 (0.729) and lowest RMSE (0.121) across validation sets and was therefore selected for weighted training of XGB on the entire dataset to map PSB FCover across the peanut area mask (Section 3.1).

Pixel-based statistics of the PSB FCover map (Fig. 19) reveal that most pixel values fell within the 10–30% range, with 10.37% being the most frequent value. Negative values were treated as 0 to represent the absence of PSB infection. Across the five administrative divisions, approximately 2/5 to 3/5 of the pixels had FCover values ≥20%, with Tanghe showing the highest proportion and Fangcheng the lowest,

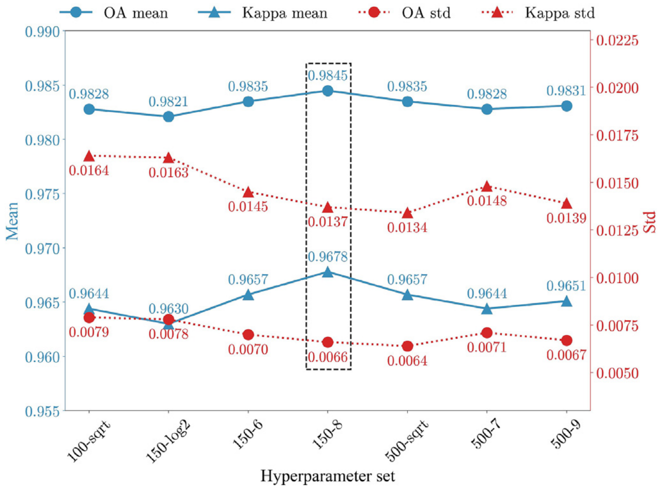


Fig. 11. Performance comparison of the optimal hyperparameter sets across 10-fold CV.

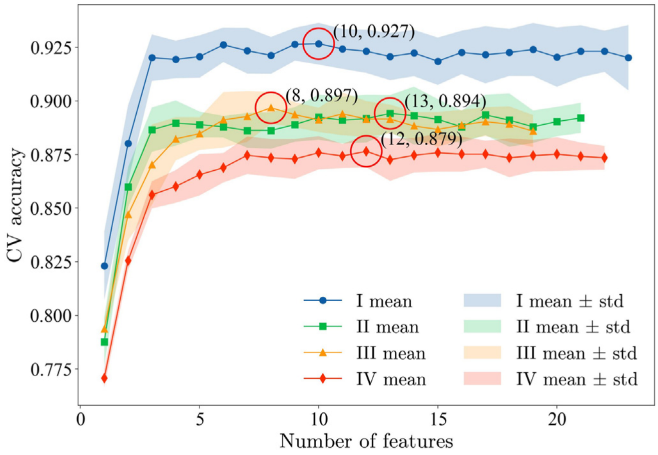


Fig. 12. Variation in 5-fold CV accuracy (mean ± std) of the RF base model during RFE for Fields I-IV.

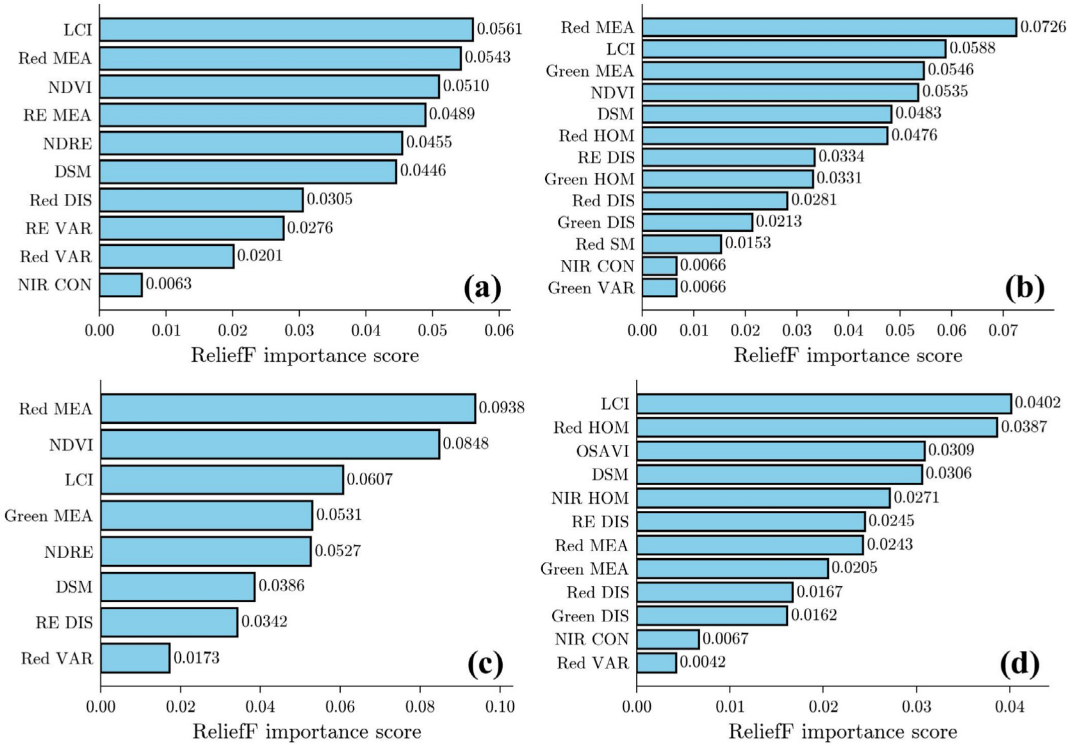


Fig. 13. ReliefF importance scores of selected features. (a) Field I. (b) Field II. (c) Field III. (d) Field IV.

indicating widespread PSB occurrence, particularly severe in Tanghe. Predicted FCover showed strong agreement with field-surveyed PSB incidence ($R^2 = 0.89$), demonstrating the reliability of the PSB FCover map (Fig. 20).

Based on the spatial patterns revealed by the PSB FCover map, the map can be directly applied to guide precision disease management. Areas with FCover values exceeding 20% can be prioritized for fungicide application, allowing targeted treatment of high-risk zones while minimizing unnecessary chemical use. In addition, the map can support zoning strategies by dividing the fields into low-, medium-, and high-risk areas according to FCover thresholds, facilitating optimized scheduling and resource allocation for disease control. By providing a spatially

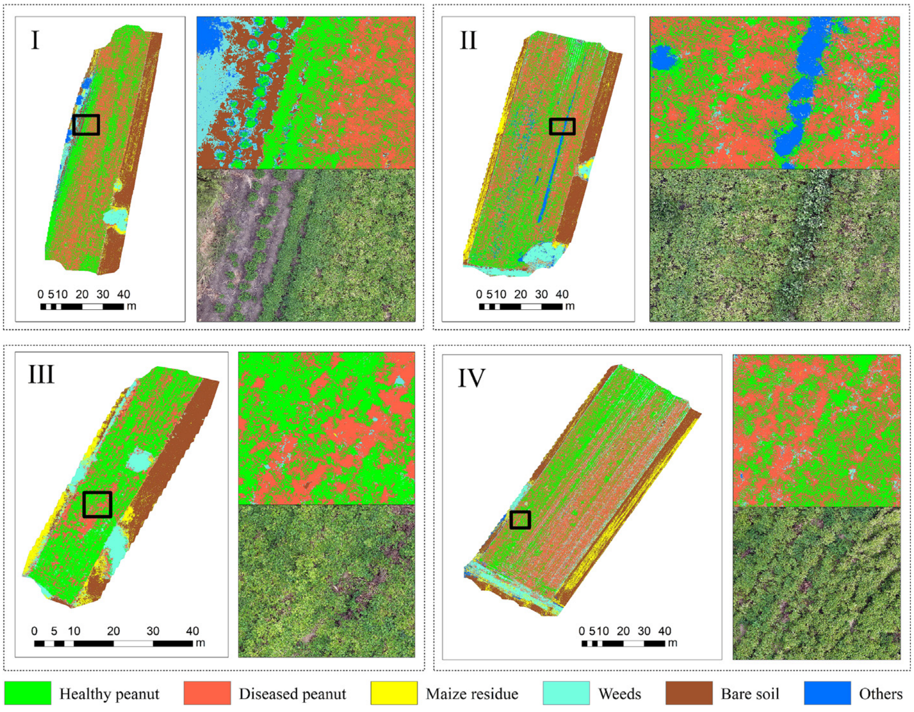


Fig. 14. UAV-based classification maps for Fields I-IV.

Table 6
Summary of LC classification performance across fields.

Field	LC class	F1	CE	OE	OA
I	Healthy peanut	92.7%	3.6%	10.8%	89.8%
	Diseased peanut	91.8%	2.1%	13.6%	
	Maize residue	70.1%	43.5%	7.9%	
	Weeds	72.8%	37.3%	13.1%	
	Bare soil	97.3%	1.5%	3.9%	
	Others	87.9%	17.8%	5.5%	
II	Healthy peanut	84.1%	16.9%	14.8%	89%
	Diseased peanut	83.8%	16%	16.4%	
	Maize residue	76.5%	34.7%	7.7%	
	Weeds	79%	28.1%	12.4%	
	Bare soil	95.9%	1.1%	7%	
	Others	90.9%	3.5%	14.1%	
III	Healthy peanut	87%	15.4%	10.6%	89.9%
	Diseased peanut	80.3%	19.4%	20%	
	Maize residue	87.8%	15.8%	8.2%	
	Weeds	90.7%	6.7%	11.8%	
	Bare soil	96.6%	2.2%	4.5%	
IV	Healthy peanut	85.7%	18%	10.4%	76.7%
	Diseased peanut	86.9%	10%	15.9%	
	Maize residue	59.1%	48.8%	30.2%	
	Weeds	77%	15.3%	29.5%	
	Bare soil	70.3%	30.7%	28.8%	
	Others	37.2%	76.8%	6.8%	

explicit and scalable assessment of PSB infestation, this approach bridges fine-scale UAV observations with regional-scale management, enhances decision-making for integrated pest management, and substantially improves the practical value of remote sensing-based disease monitoring.

4. Discussion

4.1. Advantages of the proposed method for peanut spatial distribution mapping

Accurate pixel-level mapping of the target crop is the first step before conducting regional-scale crop disease monitoring. Achieving this goal relies on the extensive data acquisition capacity of satellite remote sensing and the powerful computational environment provided by platforms like GEE (Dong and Xiao, 2016). Building on this foundation, crop mapping typically follows three approaches: classifier-, phenology-, and index-based (Li et al., 2024). Classifier-based methods rely on supervised models trained with large, high-quality samples, but their accuracy is often limited to the regions represented by those samples. Phenology-based methods rely on seasonal variations in VIs without training samples, but they require full-season time series and are sensitive to clouds, limiting generalization. Index-based methods rely on spectral band combinations, but their reliance on fixed thresholds and sensitivity to spatiotemporal variability restricts applicability across regions. To overcome these limitations, this study integrates the three methods, leveraging their complementary strengths for more robust and transferable peanut mapping.

Firstly, cloud masking, temporal compositing, and LI were applied to mitigate cloud contamination and fill temporal gaps, ensuring continuous, reliable time-series data for phenology analysis (Wang et al., 2023a). Compared with the commonly used QA60 band for cloud masking in Sentinel-2 imagery, the Cloud Score + method has been reported to perform better under high cloud-cover conditions (Gao et al., 2024; Rodríguez-Puerta et al., 2024), supporting its adoption in this study. In addition, median compositing, by reducing the impact of extreme values and suppressing residual clouds and shadows, has become a generally applicable approach for temporal aggregation in GEE (Xu et al., 2025).

Secondly, crop type discrimination was conducted within the predefined crop mask, considering the increasing availability of high-resolution LC datasets, including crop class (Zhang et al., 2025), and the challenges of distinguishing peanut from all other LC classes solely based on phenology analysis. Among all global 10-m LC datasets derived from Sentinel-2 imagery, only DW updates NRT with each new acquisition, giving it unique value for preliminary crop extent monitoring during the peanut growing season (Venter et al., 2022). As noted by Li et al. (2025a), while DW exhibits strong classification performance in homogeneous environments, it still shows biases in boundary delineation, particularly in fragmented landscapes. Therefore, in this study, a phenology-based index (ACI) was developed within the crop mask defined by DW, and threshold ranges were determined using sample fields to compensate for DW's boundary biases and further refine the crop extent.

Thirdly, a multi-step strategy, including expansion of sample points to polygons (field boundaries), nested feature selection and hyperparameter optimization within 10-fold CV, was employed to ensure the accuracy and robustness of the final peanut mapping. Polygons, compared with points, better capture spatial extent and heterogeneity, enhancing the quantity and quality of training samples (Hu et al., 2020; Karasiak et al., 2022; Zhang et al., 2023). Among the stable features consistently retained across all folds, features from Jun, Jul, Aug, and Sep accounted for 13.3%, 16.7%, 60%, and 10%, respectively, indicating that the dominance of Aug features highlights the peak growth stage of peanut as the most discriminative period for crop identification, when canopy structure and spectral contrast between peanut and other crops are maximized. Features from Jun and Jul capture early growth characteristics and phenological differences, while Sep features reflect late-season conditions. Together, these features likely provide robust discriminative information for peanut classification across the entire growing season. Notably, most retained features were derived indices rather than SR (63.3% vs. 36.7%), highlighting the importance of integrated spectral metrics that capture canopy structure, chlorophyll content, and moisture conditions. Across all 10 folds, despite differences in training and validation sets, model performance remained consistently high (OA: 0.9670–0.9967; Kappa: 0.9312–0.9931; Table 5), demonstrating that the RF classifiers were robust to fold-specific data variations and that hyperparameter optimization yielded stable models across folds. Previous studies have shown that RF classifier performance is generally more sensitive to max_features than to n_estimators (Belgiu and Drăguţ, 2016). Our results, obtained by applying each fold's optimal hyperparameter set across all folds (Fig. 11), support this pattern: when n_estimators was fixed at 150, increasing max_features resulted in larger performance gains (mean OA +0.0024, mean Kappa

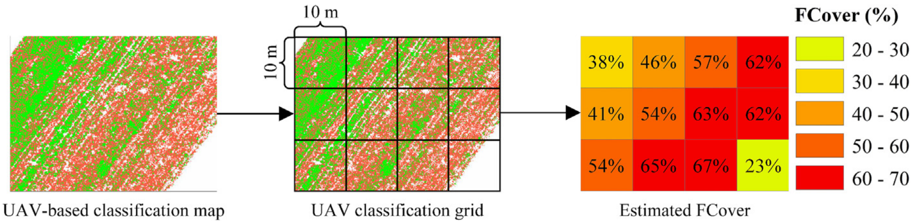


Fig. 15. Workflow of PSB FCover estimation. Note: Part of Field IV is illustrated as an example.

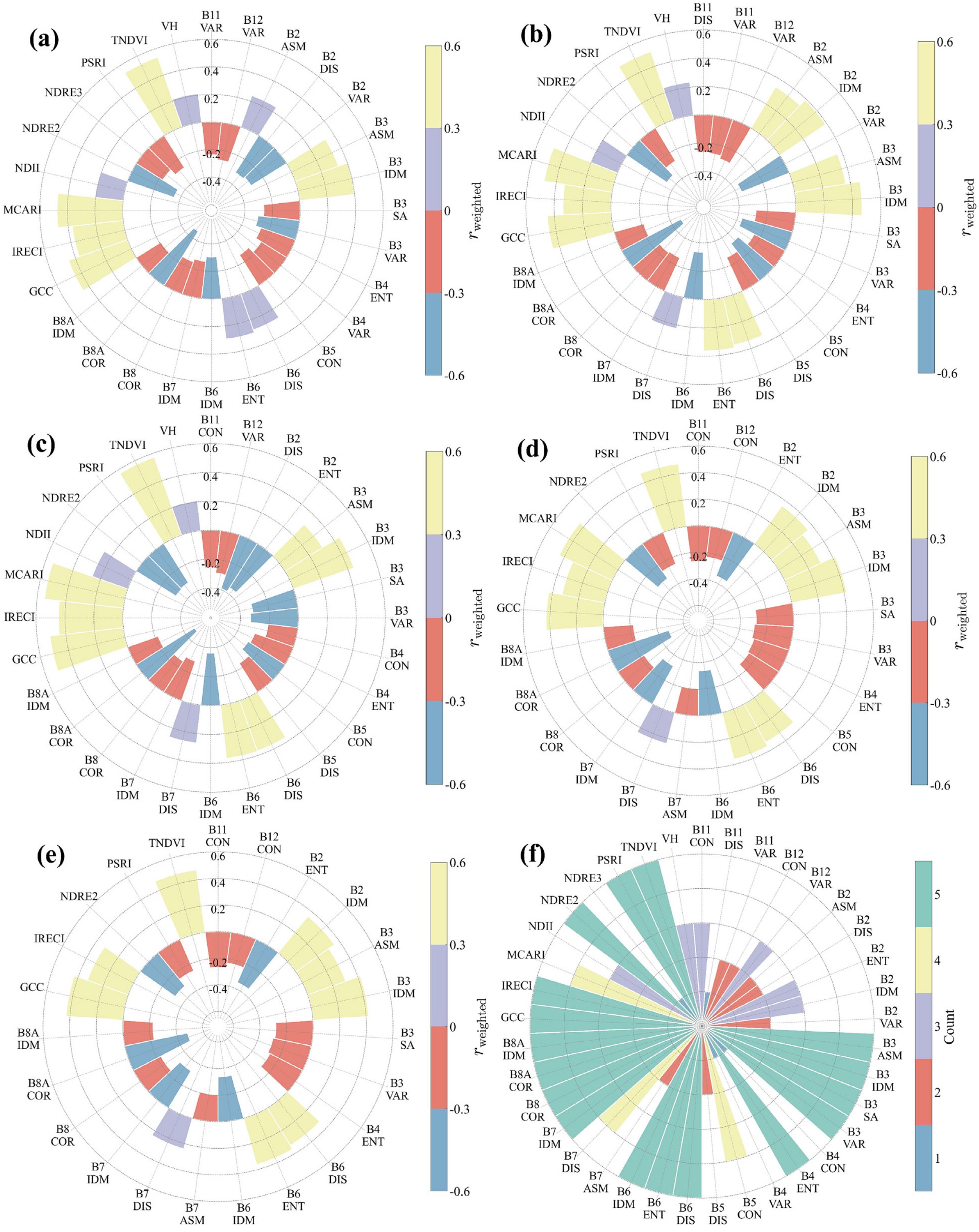


Fig. 16. Overview of selected features across 5-fold CV. (a–e) r_{weighted} between selected features of folds 1–5 and FCover. (f) Feature retention frequency across all 5 folds.

+0.0048) than those observed when varying n_estimators from 100 to 500 at a fixed max_features (sqrt) (mean OA +0.0007, mean Kappa +0.0013). As an additional observation, at n_estimators = 500, increasing max_features produced no performance gains, and mean OA and Kappa decreased by no more than 0.0007 and 0.0013, respectively, suggesting that a sufficiently large number of trees confers robustness to changes in max_features.

4.2. Spatial and temporal transferability of the ACI

To evaluate the spatial and temporal transferability of the ACI for autumn crop delineation, we applied the ACI threshold range (0.267–0.604) derived from 2024 sample field boundaries to other years within the study area (2023 and 2025) and to other regions in 2024. Corresponding sample field boundaries, manually delineated

Table 7
Optimal hyperparameters and performance of RF, ET, GB, and XGB.

Hyperparameter/ Performance	Search range	Description	RF	ET	GB	XGB
n_estimators	[50, 1000]	Number of trees	530 (185–753)	286 (110–957)	844 (623–955)	220 (98–576)
max_depth	[3, 50]	Maximum depth of each tree	40 (5–48)	29 (10–45)	28 (11–50)	5 (3–40)
max_features	{auto, sqrt, log2} or [0.2, 1]	Feature sampling strategy at each split	0.511 (0.4071–0.5937), log2 (2/5)	0.7759 (0.756–0.9427), sqrt (1/5), log2 (1/5)	auto (1/5), sqrt (2/5), log2 (2/5)	
learning_rate	[0.005, 0.3] (log scale)	Shrinks contribution of each tree			0.0314 (0.0138–0.0935)	0.0663 (0.0063–0.196)
subsample	[0.3, 1]	Fraction of samples for each tree			0.8417 (0.3851–0.9996)	0.721 (0.3502–0.9821)
loss	{squared_error, absolute_error, huber}	Loss function			squared_error (1/5), absolute_error (2/5), huber (2/5)	
colsample_bytree	[0.3, 1]	Fraction of features per tree				0.5253 (0.5029–0.9292)
reg_alpha	[0,2]	L1 regularization term				0.1076 (0.0004–0.1361)
reg_lambda	[0, 2]	L2 regularization term				0.3558 (0.0575–1.3886)
R^2			0.5103 (0.3905–0.5551)	0.6016 (0.4498–0.6718)	0.6233 (0.4763–0.7187)	0.7412 (0.6086–0.7511)
RMSE			0.1619 (0.1485–0.166)	0.1471 (0.1325–0.1663)	0.1453 (0.1227–0.1622)	0.121 (0.1128–0.1403)

Note: auto = total number of features. For numeric hyperparameters and performance metrics, values are summarized as the median (minimum–maximum) of all numeric values identified across all 5 folds. For categorical hyperparameters, the selection frequency across all 5 folds is reported. Empty cells indicate that the parameter is not applicable to the model.

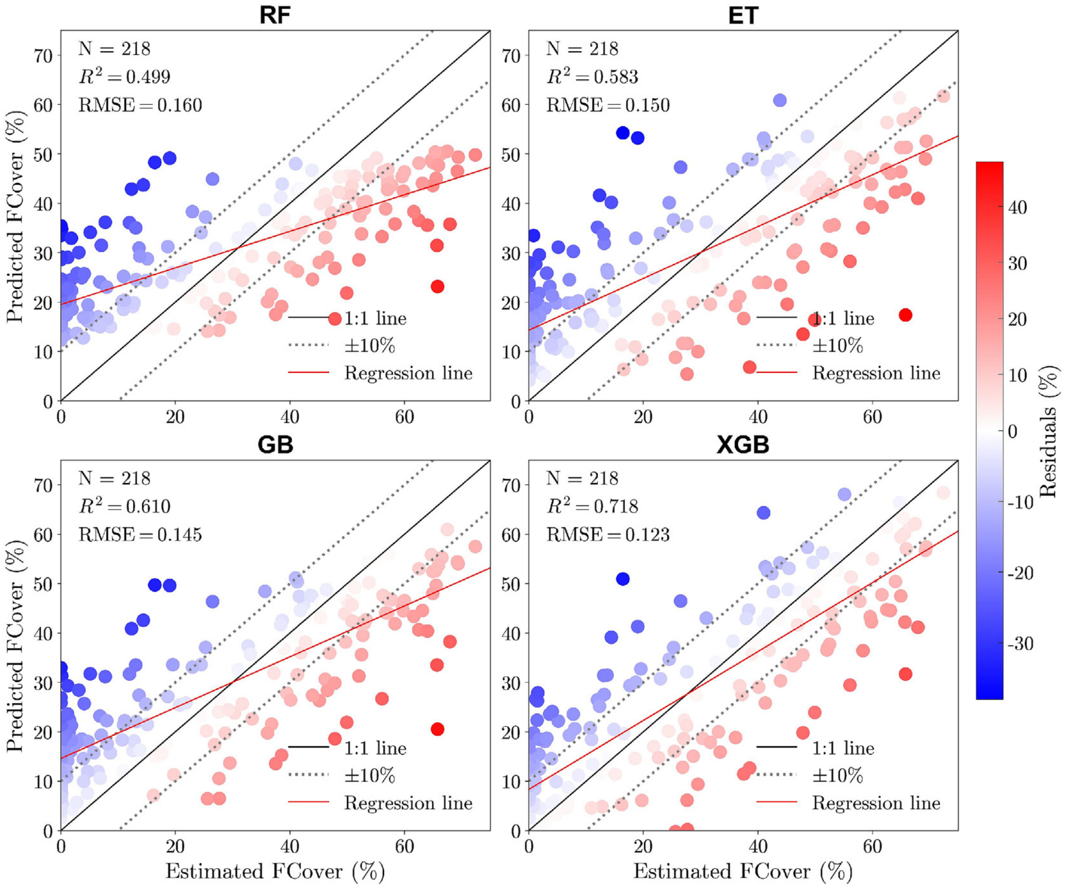


Fig. 17. Scatter plots of predicted vs. estimated FCover for all validation sets for RF, ET, GB, and XGB. Note: Points are colored by residuals, with the depth of red and blue indicating the magnitude of underestimation and overestimation, respectively. Dashed lines indicate ±10% deviation from the 1:1 line. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

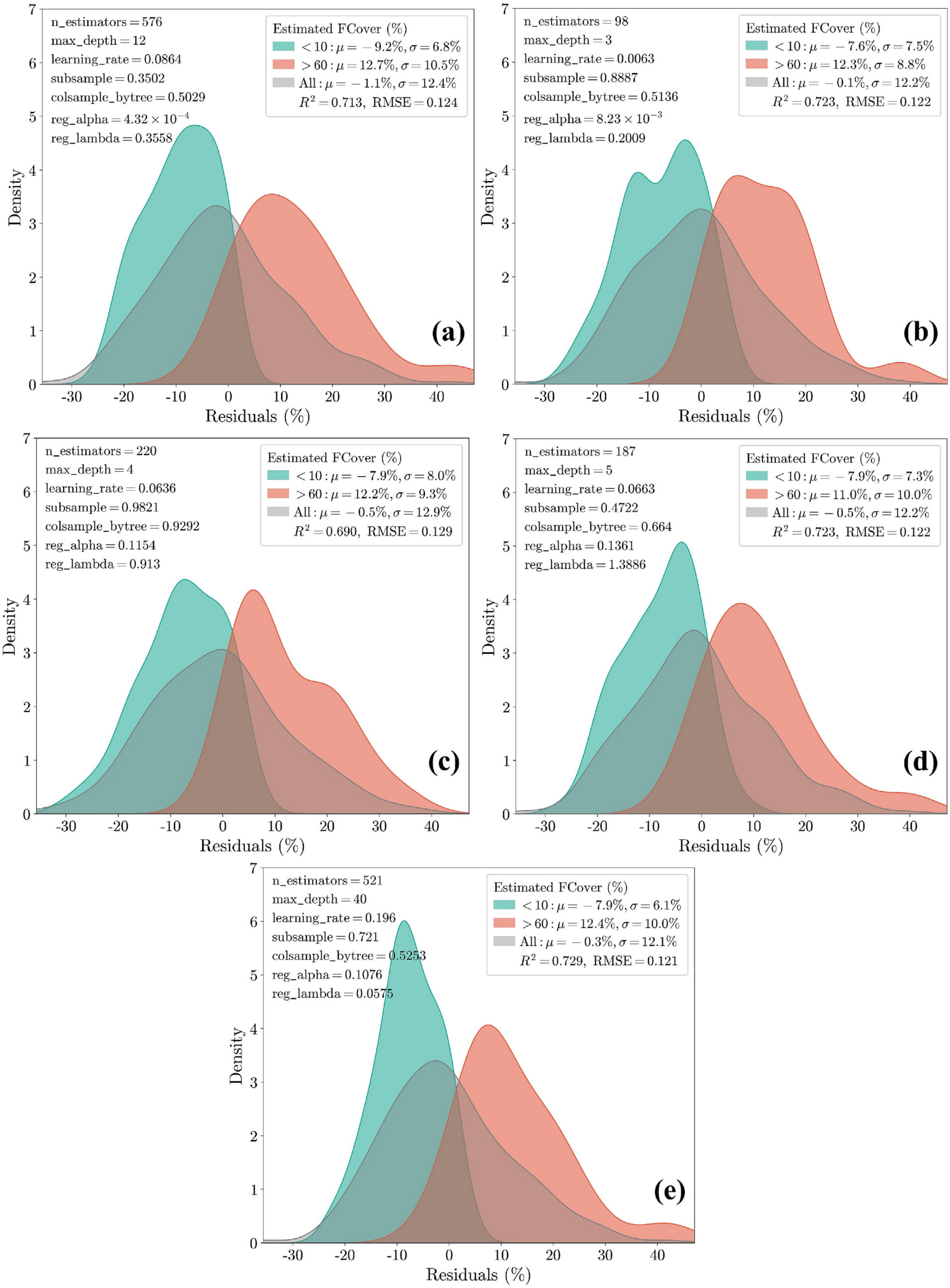


Fig. 18. Residual distributions of cross-fold XGB validation predictions. (a–e) Residuals for FCover <10%, >60%, and overall, based on folds 1–5 optimal hyperparameters applied across all folds, respectively.

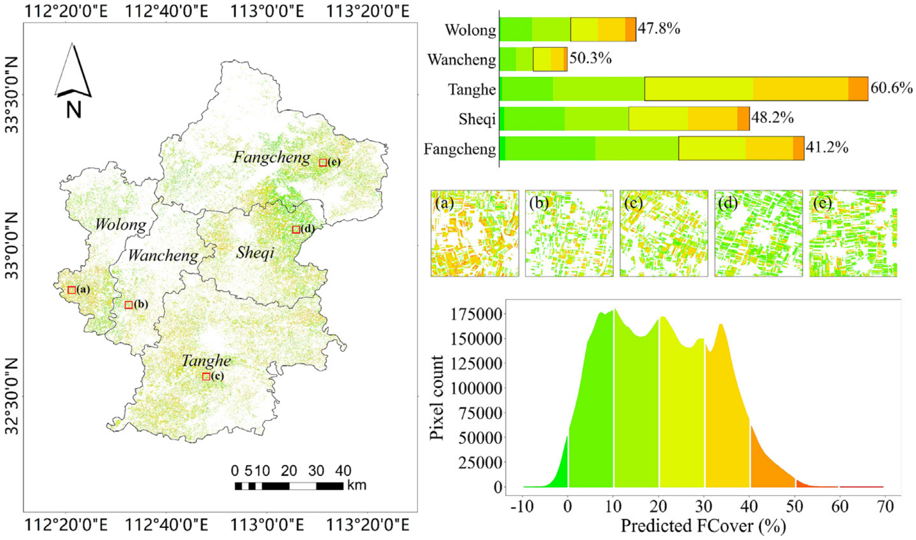


Fig. 19. PSB FCover map, with pixel-based statistics shown by 10% equal-interval categories. Note: Rectangular frames indicate stacked bar segments where FCover ≥20%, with proportions shown to the right.

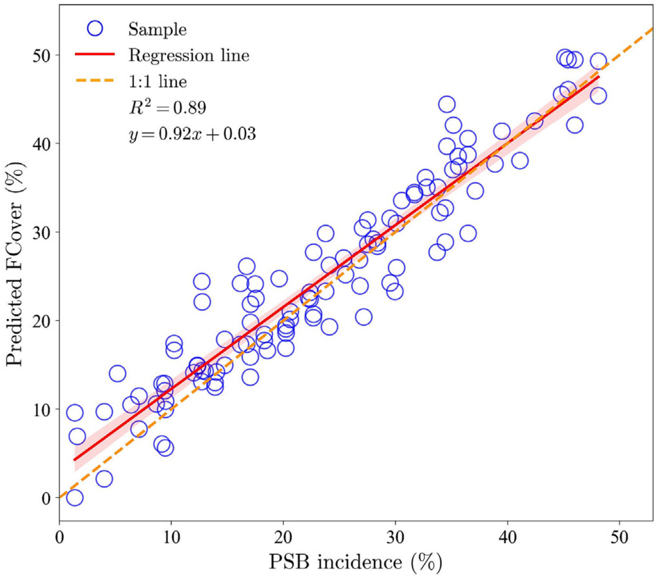


Fig. 20. Scatter plot of predicted FCover vs. field-surveyed PSB incidence.

based on in-situ sample points and PlanetScope reference imagery, were necessary to calculate PA and OE, which measure the proportion of autumn crop pixels correctly retained or missed. As some years and regions lacked in-situ sample points (study area: 2025; other regions: 2024), we assumed that autumn crop planting patterns remain largely stable over a short period (e.g., three years) and used 2023–2024 sample points to represent 2025 in the study area and other-year sample points from other regions to represent 2024 (Table 8 and Figs. 21c–f). Independently, ACI was calculated using the DW crop mask and Sentinel-2 composites of the corresponding years for other regions in 2024 and the study area in 2023 and 2025, following the same processing strategy as in Sections 2.2.1.2 and 2.2.2.1.

The transferability of the ACI threshold range across years and regions showed substantial variability (Fig. 21g). For the study area, PA was high (85.5%) and OE was low (14.5%) in 2023, whereas in 2025 PA dropped sharply to 1.9% and OE rose to 98.1%. This drastic decline in performance can be largely attributed to the lower average proportion of good Sentinel-2 observations per pixel in 2025 (24.7%) compared to 2023 (36.9%) (Table 8), which explains the threshold range failure, as reflected in Fig. 21c. For other regions in 2024, the threshold range showed variable performance, with PA ranging from 65.0% to 92.5% and OE from 7.5% to 35.0%. Differences among regions were influenced by both per-pixel average good observations and crop composition. Region A, dominated by maize and peanut, exhibited the highest PA and lowest OE. In contrast, in more heterogeneous systems like Region C, crop composition appeared to be the dominant factor limiting threshold range performance, as reflected by its lowest PA despite the highest average proportion of good observations. This is further illustrated in Fig. 21f, where the ACI values within the DW crop mask exhibit a bimodal distribution, with the unselected peak largely corresponding to the southern rice-dominated area, accounting for most extraction omissions. Overall, these results indicate that the empirical ACI threshold range is reasonably transferable across regions with similar crop types and adequate per-pixel good observations, but its performance is limited when applied to years with sparse observations or to regions

Table 8
Overview of regions for ACI application and in-situ sample fields.

Region	ACI application year	Main autumn crops	Sentinel-2 average good observations (%)	Sample points available?	Sample points migrated from	PlanetScope reference imagery date
Study area	2023	Maize, peanut, soybean	36.9%	Yes	–	Aug 31, 2023
Study area	2025		24.7%	No	2023 and 2024	Aug 16, 2025
Region A	2024	Maize, peanut	41%	No	2023	Aug 11, 2024
Region B	2024	Maize, peanut, soybean	47.9%	No	2025	Aug 11, 2024
Region C	2024	Maize, rice, peanut, soybean	48.4%	No	2023	Aug 11, 2024

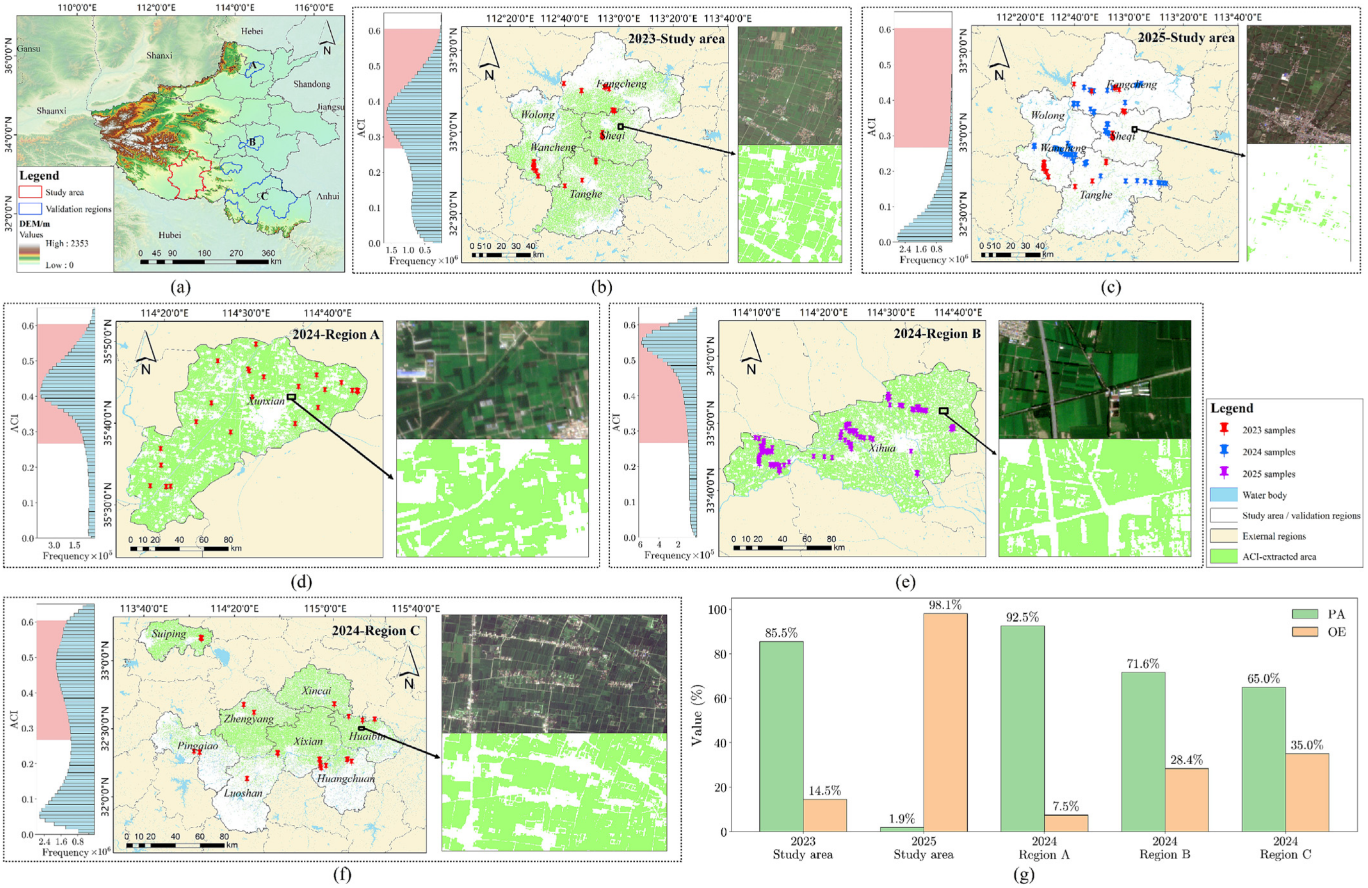


Fig. 21. Application and evaluation of the ACI threshold range across years and regions. (a) Geographic location of the study area and validation regions. (b–f) Application results of the ACI threshold range for the study area in 2023 and 2025, and for other regions in 2024, with sample point distributions overlaid. (g) Comparison of PA and OE across regions and years.

with more heterogeneous cropping systems. These findings highlight the importance of considering both imagery availability and crop composition when applying the ACI across years and regions.

4.3. Challenges and uncertainties in PSB FCover estimation

Accurate estimation of PSB FCover relies on reliable UAV-based pixel-level classification. In natural fields, PSB identification is challenging because root collar infections often precede visible canopy symptoms, and other stressors (e.g., nutrient deficiencies, drought, or co-occurring diseases) can produce similar visual or spectral responses (Zhao et al., 2023), potentially leading to underestimation of actual infection. To address this, this study assumes that PSB is the primary stressor in the surveyed fields. While it is not possible to completely rule out other stresses, UAV-labeled diseased-peanut pixels were delineated based on typical PSB symptoms (canopy chlorosis or wilting) and verified through field inspection, ensuring that marked pixels predominantly represent PSB infection. UAV flight altitude was adjusted to improve visual discrimination of infected plants.

While ultra-high-resolution UAV imagery provides fine spatial detail, it can accentuate spectral variability within classes ("same-object, different-spectrum") and similarity between classes ("different-object, same-spectrum"), introducing uncertainty in feature selection and classification (Chen et al., 2023b). Random points sampled within training ROIs, combined with two-step feature selection of VIs, TFs, and DSM (see Section 2.2.3 for details), helped capture spectral variations and enhance class discriminability to reduce this uncertainty (Rommel et al., 2022; Zheng et al., 2024). Among selected features, several (i.e., Red MEA, LCI, and DSM) were consistently retained across all four fields, with cumulative ReliefF importance scores of 0.245, 0.216, and 0.162, respectively, indicating robust discriminative power under complex field conditions. Additional VIs and TFs (i.e., NDVI, Green MEA, Red DIS, Red VAR, RE DIS, and NIR CON) were retained in three or fewer fields, providing auxiliary discriminative information under certain field conditions. It is worth noting that GNDVI and TFs associated with COR and ENT were not retained in any field, indicating limited contribution to class discrimination under current UAV and field conditions.

Despite rigorous ROIs delineation and feature selection, classification errors persisted for certain classes, particularly for diseased peanut and spectrally similar background classes (Table 6), highlighting the need to further methodological improvements. These include refining ROI sampling strategies (random plus field-verified points), expanding features (e.g., VIs, color indices, textural indices, and canopy height model), optimizing textural parameters (e.g., window size), and integrating multi-source or multi-temporal data to reduce spectral confusion and enhance classification robustness (Guo et al., 2024; Lu et al., 2022; Ma et al., 2024b; Xiao et al., 2021; Xu et al., 2023).

4.4. Predictive performance, limitations, and transferability of PSB FCover models

Satellite-derived features for modeling PSB FCover were obtained from Sentinel-1 SAR imagery (Sep 14) and Sentinel-2 composite (Sep 04), chosen considering Sentinel-1's 12-day revisit interval, potential cloud contamination in Sentinel-2 imagery, and the temporal representativeness of the composite. To mitigate overfitting given the high feature dimensionality relative to the sample size (218 FCover samples vs. 101 initial features), we selected 17 stable features retained across all 5-fold CV training sets. Hyperparameter optimization for four ensemble models (XGB, RF, GB, ET) was performed within each training set using Bayesian optimization combined with 3-fold CV, with sample weights incorporated in the optimization process. Model performance was evaluated on each and all validation sets via R^2 and RMSE, and the best-performing model was identified. Finally, the optimal hyperparameters from each fold were applied to train the best-performing model across all training sets, considering sample weights. Model performance was then evaluated on all validation sets to ensure the selection of final hyperparameters for PSB FCover mapping, maximizing R^2, minimizing RMSE, and ensuring that the mean residuals were close to 0 with low std. This multi-layered approach—feature selection, hyperparameter tuning, and CV validation—effectively controlled overfitting, ensuring robust and generalizable PSB FCover predictions.

Notably, among the Sentinel-1 polarimetric features, only VH was retained in 3 of the 5 CV folds, indicating that polarimetric signals generally contributed weakly and inconsistently to FCover prediction. This likely reflects the localized impact of PSB near the root collar, whereas polarimetric features primarily capture broader structural and moisture signals; inherent speckle noise in Sentinel-1 imagery may further limit their reliability (Tian et al., 2025). In contrast, 17 features stably retained across all folds showed substantial cumulative weighted correlations (−2.38 to 2.56). Positive features (e.g., TNDVI, GCC, B3 IDM) were associated with PSB-induced declines in greenness and biomass, whereas negative features (e.g., NDRE2, B8A COR) were associated with healthy vegetation or background signals. These stable features indicate that VIs and TFs provide more informative signals than Sentinel-1 polarimetric features for FCover modeling.

A potential concern with XGB is that shallow trees may limit the model's ability to capture complex relationships. In this study, the optimal max_depth varied from 3 to 40 across all 5 folds (Table 7). Models with shallow trees (max_depth = 3–5) achieved consistent performance across all folds, with R^2 values of 0.690 and 0.723, and RMSE values of 0.122 and 0.129. Mean residuals were near 0, and residual stds were 12.2% and 12.9%, indicating that the model effectively captured key FCover signals without overfitting. The deepest configuration (fold 5, max_depth = 40) showed a slight improvement in performance ($R^2 = 0.729$; RMSE = 0.121), while residuals remained stable, suggesting that increased tree depth helped capture more complex relationships without sacrificing generalization. Overall, the consistency across folds in both validation accuracy and residual behavior demonstrates that the selected max_depth values successfully balanced model complexity and the risk of overfitting.

The FCover values used as ground truth were derived from UAV-based classification rather than direct field measurements. As a result, classification uncertainty may propagate through the model's training and prediction processes. To quantify this impact, we conducted 1000 Monte Carlo perturbations for each field's FCover samples, based on OEs of diseased peanut pixels in the UAV-based classification maps (Field I: 13.6%; Field II: 16.4%; Field III: 20%; Field IV: 15.9%). Perturbations were drawn from a normal distribution with σ determined by the product of OE and FCover:

$$\mathrm{FCover}_i^{(m)} \sim \mathcal{N}\left(\mathrm{FCover}_i, \left(\mathrm{OE}_f \times \mathrm{FCover}_i\right)^2\right), m = 1, 2, \cdots, 1000 \tag{16}$$

where $\mathrm{FCover}_i^{(m)}$ is perturbed FCover value for sample i in the m-th Monte Carlo iteration; FCover_i is original FCover value for sample i; OE_f is OE for corresponding field, representing UAV-based classification uncertainty of diseased peanut pixels. Perturbed FCover values were truncated to the range from 0 to 100% to ensure physically meaningful values. For each perturbed FCover sample, the weights were recalculated using Eq. (10) within the 5-fold CV in Section 3.3.

XGB models were trained on each fold's training set, incorporating 17 stable features, both original and perturbed FCover values, FCover weights, and the 5th-fold XGB optimal hyperparameters, and then evaluated on the corresponding validation sets via R^2 and RMSE to quantify performance variability. The mean R^2 and RMSE across all 5 folds reflect the overall predictive accuracy and consistency of the XGB models. The performance of XGB models under FCover perturbations is summarized in Table 9, where the median R^2 values for perturbed FCover ranged from 0.606 to 0.747, with an average of 0.701 (compared to 0.725 for the original FCover), and median RMSE ranged from 0.118 to 0.141, with an average of 0.126 (compared to 0.121 for the original FCover). These results indicate that the propagation of classification uncertainty from UAV-based FCover measurements slightly reduced predictive accuracy, with median decreases of 0.025 in R^2 and 0.005 in RMSE on average. Nevertheless, the relatively small magnitude of these changes demonstrates that the XGB models are robust to the observed level of UAV classification error. Furthermore, Fig. 22 highlights that the interquartile range (IQR) of R^2 and RMSE for the average across folds is smaller than that of individual folds, suggesting that although FCover perturbations introduce some variability in single folds, the overall predictive performance remains consistent. This is reflected in the shorter box lengths for the average values compared to individual folds. Overall, these results quantify how UAV-derived FCover uncertainty propagates through the model training and prediction processes and provide evidence that the selected 17 stable features and CV framework mitigate the impact of classification uncertainty on PSB FCover mapping.

Beyond assessing model reliability, the spatial distribution of PSB FCover also provides valuable information for practical disease management at the field scale. For example, FCover values exceeding a certain threshold (e.g., around 20%) may indicate moderate to severe infection and could serve as an operational trigger for targeted intervention. Areas with relatively high FCover values therefore represent zones where disease symptoms are more widespread and may require prioritized management actions, such as targeted fungicide application or intensified field inspection. In contrast, areas with low FCover values may only require routine monitoring rather than immediate treatment. By translating FCover map into spatial risk levels, farmers and agronomists can implement site-specific management strategies, focusing control efforts on high-risk zones while avoiding unnecessary pesticide use in

Table 9

Comparison of model performance (R^2 and RMSE) before and after perturbation of FCover values.

Fold	Perturbed R^2 (median ± IQR)	Original R^2	△R^2 (perturbed median – original)	Perturbed RMSE (median ± IQR)	Original RMSE	△RMSE (perturbed median – original)
1	0.692 ± 0.035	0.735	−0.043	0.131 ± 0.007	0.122	0.010
2	0.747 ± 0.035	0.762	−0.016	0.120 ± 0.008	0.116	0.004
3	0.606 ± 0.045	0.623	−0.017	0.141 ± 0.008	0.138	0.003
4	0.733 ± 0.036	0.755	−0.022	0.119 ± 0.008	0.114	0.005
5	0.728 ± 0.040	0.751	−0.023	0.118 ± 0.009	0.113	0.005
Average	0.701 ± 0.018	0.725	−0.025	0.126 ± 0.004	0.121	0.005

Note: IQR is defined as the difference between the 75th and 25th percentiles (Q3 − Q1).

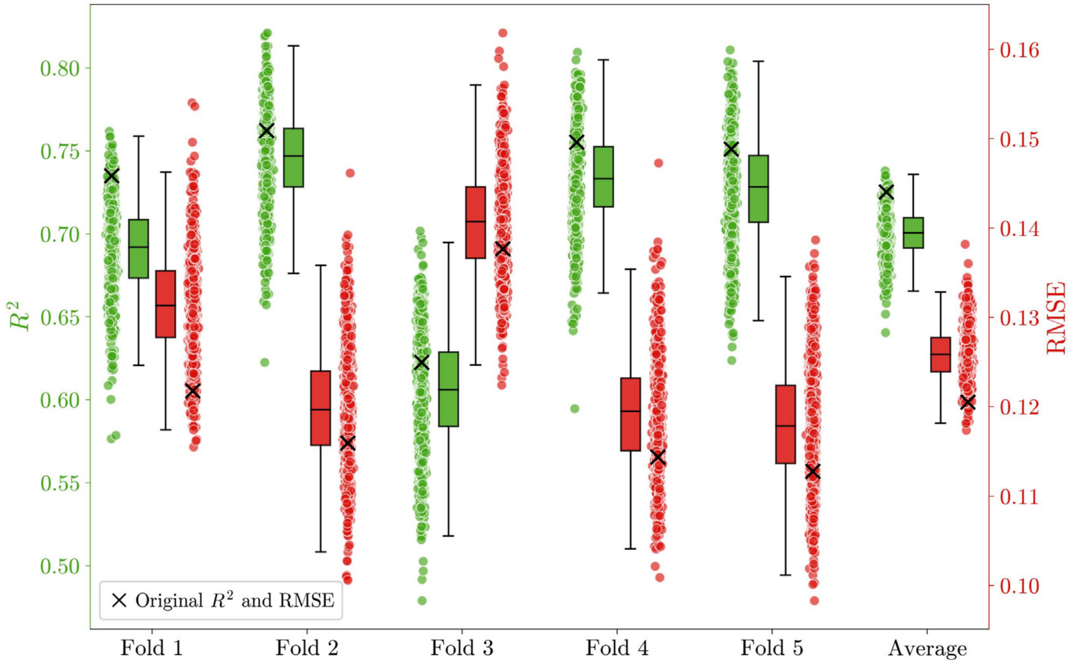


Fig. 22. Cross-fold R^2 and RMSE variability with average values (scatter-box plot), including original R^2 and RMSE markers, which refer to metrics computed using unperturbed FCover values.

lightly affected areas. Such spatially explicit information can improve disease control efficiency and support more sustainable crop protection practices.

While the UAV-satellite framework effectively addresses PSB monitoring, several limitations should be considered for further refinement. The dataset used in this study consists of only four fields, which may constrain the model's generalizability. Specifically, the risk remains that the model may learn field-specific background patterns rather than focusing solely on disease-related signals. Expanding the dataset to include a more diverse range of fields, both geographically and in terms of cropping systems, would likely improve model robustness and reduce overfitting to localized patterns. Additionally, the uncertainty inherent in UAV-derived FCover values, which are based on classification rather than direct field measurements, may propagate through the modeling process. Although techniques like 5-fold CV help mitigate overfitting, further exploration of other validation methods, such as leave-one-field-out, could offer additional insights into the model's ability to generalize to unseen fields. The framework's adaptability to regions with varying cropping structures, climate conditions, UAV parameters (e.g., flight altitude, sensor resolution), and satellite revisit frequency is still uncertain. These factors may influence model performance, so future research should focus on assessing their impact. Moreover, comparing this methodology with studies on other crop diseases, such as wheat rust or cotton root rot, would provide a clearer understanding of its broader applicability and limitations.

5. Conclusions

This study demonstrated the practical implementation of a ML-based UAV-satellite integration framework for mapping PSB FCover at the regional scale. The framework successfully achieved three key objectives: accurate regional-scale peanut mapping, derivation of 10 m pixel-level PSB FCover samples from UAV-based classification maps, and Sentinel-2-based prediction of PSB FCover within the mapped peanut area. Importantly, the approach provides a viable solution for regional-scale PSB monitoring, with XGB-based modeling demonstrating robust predictive performance across the study area. Future work will focus on improving UAV-based classification reliability, extending the framework to other regions and crop diseases, and developing predictive models that integrate meteorological factors and crop growth stages to support more effective disease management (Li et al., 2025b).

Data statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

CRediT authorship contribution statement

Shangzhou Li: Writing – original draft, Visualization, Validation, Methodology. **Juan Liu:** Investigation, Funding acquisition, Data curation. **Yapeng Zhang:** Visualization, Investigation. **Sifan Feng:** Software, Investigation. **Lu Liu:** Visualization, Software. **Jibo Yue:** Writing – review & editing, Resources, Formal analysis. **Yang Liu:** Writing – review & editing, Resources, Formal analysis. **Meiyan Shu:** Writing – review & editing, Resources, Formal analysis. **Fernando Auat Cheein:** Writing – review & editing. **Wei Guo:** Writing – review & editing, Supervision, Resources, Project administration, Funding acquisition, Conceptualization.

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work, the authors used ChatGPT and DeepSeek in order to improve readability and language. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (32271993), the International Science and Technology Cooperation Program of Henan Province (262102520021), the Henan Provincial Science and Technology Research and Development Joint Fund (222301420114, 222301420113), the Earmarked Fund for China Agriculture Research System (CARS-13), the Henan Provincial Key Science and Technology Program (242102111072), and the Major Emergency Response Special Project for Agricultural Production in Henan Province (2024ZDYJ001).

References

Akram, N.A., Shafiq, F., Ashraf, M., 2018. Peanut (*Arachis hypogaea* L.): a prospective legume crop to offer multiple health benefits under changing climate. Compr. Rev. Food Sci. Food Saf. 17 (5), 1325–1338. https://doi.org/10.1111/1541-4337.12383.

Amani, M., Ghorbanian, A., Ahmadi, S.A., Kakooei, M., Moghimi, A., Mirmazloumi, S.M., Moghaddam, S.H.A., Mahdavi, S., Ghahremanloo, M., Parsian, S., Wu, Q., Brisco, B., 2020. Google earth engine cloud computing platform for remote sensing big data applications: a comprehensive review. IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens. 13, 5326–5350. https://doi.org/10.1109/JSTARS.2020.3021052.

Aycock, R., 1966. Stem rot and Other Diseases Caused by *Sclerotium rolfsii* or the Status of Rolfs fungus after 70 Years. Technical bulletin no. 174. North Carolina agricultural experiment station, Raleigh, NC.

Belgiu, M., Drăguţ, L., 2016. Random forest in remote sensing: a review of applications and future directions. ISPRS J. Photogramm. Remote Sens. 114, 24–31. https://doi.org/10.1016/j.isprsjprs.2016.01.011.

Benesty, J., Chen, J., Huang, Y., Cohen, I., 2009. Pearson correlation coefficient. Noise Reduction in Speech Processing. Springer, Berlin, Heidelberg, pp. 1–4 https://doi.org/10.1007/978-3-642-00296-0_5.

Biswal, A.K., Ozias-Akins, P., Holbrook, C.C., 2024. Recent technological advancements for identifying and exploiting novel sources of pest and disease resistance for peanut improvement. Agronomy-Basel. 14 (12), 3071. https://doi.org/10.3390/agronomy14123071.

Branch, W.D., Csinos, A.S., 1987. Evaluation of peanut cultivars for resistance to field infection by *Sclerotium rolfsii*. Plant Dis. 71, 268–270. https://doi.org/10.1094/PD-71-0268.

Breiman, L., 1996. Bagging predictors. Mach. Learn. 24, 123–140. https://doi.org/10.1023/A:1018054314350.

Breiman, L., 2001. Random forests. Mach. Learn. 45, 5–32. https://doi.org/10.1023/A:1010933404324.

Brown, C.F., Brumby, S.P., Guzder-Williams, B., Birch, T., Hyde, S.B., Mazzariello, J., Czerwinski, W., Pasquarella, V.J., Haertel, R., Ilyushchenko, S., Schwehr, K., Weisse, M., Stolle, F., Hanson, C., Guinan, O., Moore, R., Tait, A.M., 2022. Dynamic world, near real-time global 10 m land use land cover mapping. Sci. Data. 9 (1), 251. https://doi.org/10.1038/s41597-022-01307-4.

Bylander, T., 2002. Estimating generalization error on two-class datasets using out-of-bag estimates. Mach. Learn. 48, 287–297. https://doi.org/10.1023/A:1013964023376.

Che, X., Zhang, H.K., Li, Z.B., Wang, Y., Sun, Q., Luo, D., Wang, H., 2024. Linearly interpolating missing values in time series helps little for land cover classification using recurrent or attention networks. ISPRS J. Photogramm. Remote Sens. 212, 73–95. https://doi.org/10.1016/j.isprsjprs.2024.04.021.

Chen, D., Shi, Y., Huang, W., Zhang, J., Wu, K., 2018a. Mapping wheat rust based on high spatial resolution satellite imagery. Comput. Electron. Agric. 152, 109–116. https://doi.org/10.1016/j.compag.2018.07.002.

Chen, H., Li, H., Liu, Z., Zhang, C., Zhang, S., Atkinson, P.M., 2023a. A novel greenness and water content composite index (GWCCI) for soybean mapping from single remotely sensed multispectral images. Remote Sens. Environ. 295, 113679. https://doi.org/10.1016/j.rse.2023.113679.

Chen, J., Chen, Z., Huang, R., You, H., Han, X., Yue, T., Zhou, G., 2023b. The effects of spatial resolution and resampling on the classification accuracy of wetland vegetation species and ground objects: a study based on high spatial resolution UAV images. Drones 7 (1), 61. https://doi.org/10.3390/drones7010061.

Chen, J.M., 1996. Evaluation of vegetation indices and a modified simple ratio for boreal applications. Can. J. Remote. Sens. 22 (3), 229–242. https://doi.org/10.1080/07038992.1996.10855178.

Chen, K.R., Ren, L., Xu, L., Chen, W., Liu, F., Fang, X.P., 2018b. Research progress on peanut southern stem rot caused by *Sclerotium rolfsii*. Chin. J. Oil Crop Sci. 40 (2), 302–308. https://doi.org/10.7505/j.issn.1007-9084.2018.02.018.

Chen, T., Guestrin, C., 2016. XGBoost: A scalable tree boosting system. Proceedings of the 22nd acm sigkdd International Conference on Knowledge Discovery and Data Mining, pp. 785–794 https://doi.org/10.1145/2939672.2939785.

Daughtry, C.S., Walthall, C.L., Kim, M.S., De Colstoun, E.B., McMurtrey, J.E., 2000. Estimating corn leaf chlorophyll concentration from leaf and canopy reflectance. Remote Sens. Environ. 74 (2), 229–239. https://doi.org/10.1016/S0034-4257(00)00113-9.

Deng, Y., Wu, C., Li, M., Chen, R., 2015. RNDSI: a ratio normalized difference soil index for remote sensing of urban/suburban environments. Int. J. Appl. Earth Obs. Geoinf. 39, 40–48. https://doi.org/10.1016/j.jag.2015.02.010.

Di, Y., Dong, J., Zhu, F., Fu, P., 2022. A robust but straightforward phenology-based ginger mapping algorithm by using unique phenology features, and time-series sentinel-2 images. Comput. Electron. Agric. 198, 107066. https://doi.org/10.1016/j.compag.2022.107066.

Dietterich, T.G., 2000. Ensemble Methods in Machine Learning. International workshop on multiple classifier systems, pp. 1–15 https://doi.org/10.1007/3-540-45014-9_1.

Dong, J., Xiao, X., 2016. Evolution of regional to global paddy rice mapping methods: a review. ISPRS J. Photogramm. Remote Sens. 119, 214–227. https://doi.org/10.1016/j.isprsjprs.2016.05.010.

Dong, J., Xiao, X., Menarguez, M.A., Zhang, G., Qin, Y., Thau, D., Biradar, C., Moore, B., 2016. Mapping paddy rice planting area in northeastern Asia with Landsat 8 images, phenology-based algorithm and google earth engine. Remote Sens. Environ. 185, 142–154. https://doi.org/10.1016/j.rse.2016.02.016.

Du, X., Li, Q., Shang, J., Liu, J., Qian, B., Jing, Q., Dong, T., Fan, D., Wang, H., Zhao, L., Lieff, S., Davies, T., 2019. Detecting advanced stages of winter wheat yellow rust and aphid infection using RapidEye data in North China plain. GISci. Remote Sens. 56 (7), 1093–1113. https://doi.org/10.1080/15481603.2019.1613804.

Foody, G.M., 2002. Status of land cover classification accuracy assessment. Remote Sens. Environ. 80 (1), 185–201. https://doi.org/10.1016/S0034-4257(01)00295-4.

Frampton, W.J., Dash, J., Watmough, G., Milton, E.J., 2013. Evaluating the capabilities of sentinel-2 for quantitative estimation of biophysical variables in vegetation. ISPRS J. Photogramm. Remote Sens. 82, 83–92. https://doi.org/10.1016/j.isprsjprs.2013.04.007.

Friedman, J.H., 2001. Greedy function approximation: a gradient boosting machine. Ann. Stat. 29 (5), 1189–1232. https://doi.org/10.1214/aos/1013203451.

Gao, X., Chi, H., Huang, J., Han, Y., Li, Y., Ling, F., 2024. Comparison of cloud-mask algorithms and machine-learning methods using sentinel-2 imagery for mapping paddy rice in Jianghan plain. Remote Sens. 16 (7), 1305. https://doi.org/10.3390/rs16071305.

Geurts, P., Ernst, D., Wehenkel, L., 2006. Extremely randomized trees. Mach. Learn. 63, 3–42. https://doi.org/10.1007/s10994-006-6226-1.

Gillespie, A.R., Kahle, A.B., Walker, R.E., 1987. Color enhancement of highly correlated images. II. Channel ratio and "chromaticity" transformation techniques. Remote Sens. Environ. 22 (3), 343–365. https://doi.org/10.1016/0034-4257(87)90088-5.

Gitelson, A., Merzlyak, M.N., 1994. Spectral reflectance changes associated with autumn senescence of *Aesculus hippocastanum* L. and *Acer platanoides* L. leaves. Spectral features and relation to chlorophyll estimation. J. Plant Physiol. 143 (3), 286–292. https://doi.org/10.1016/S0176-1617(11)81633-0.

Gitelson, A.A., Kaufman, Y.J., Merzlyak, M.N., 1996. Use of a green channel in remote sensing of global vegetation from EOS-MODIS. Remote Sens. Environ. 58 (3), 289–298. https://doi.org/10.1016/S0034-4257(96)00072-7.

Gitelson, A.A., Viña, A., Arkebauer, T.J., Rundquist, D.C., Keydan, G., Leavitt, B., 2003. Remote estimation of leaf area index and green leaf biomass in maize canopies. Geophys. Res. Lett. 30 (5), 1248. https://doi.org/10.1029/2002GL016450.

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., Moore, R., 2017. Google earth engine: planetary-scale geospatial analysis for everyone. Remote Sens. Environ. 202, 18–27. https://doi.org/10.1016/j.rse.2017.06.031.

Guo, A., Huang, W., Ye, H., Dong, Y., Ma, H., Ren, Y., Ruan, C., 2020. Identification of wheat yellow rust using spectral and texture features of hyperspectral images. Remote Sens. 12 (9), 1419. https://doi.org/10.3390/rs12091419.

Guo, J., Jin, Y., Ye, H., Huang, W., Zhao, J., Cui, B., Liu, F., Deng, J., 2022. Recognition of areca leaf yellow disease based on planetscope satellite imagery. Agronomy-Basel. 12 (1), 14. https://doi.org/10.3390/agronomy12010014.

Guo, W., Sun, H., Qiao, H., Zhang, H., Zhou, L., Dong, P., Song, X., 2023. Spectral detection of peanut southern blight severity based on continuous wavelet transform and machine learning. Agriculture-Basel. 13 (8), 1504. https://doi.org/10.3390/agriculture13081504.

Guo, W., Gong, Z., Gao, C., Yue, J., Fu, Y., Sun, H., Zhang, H., Zhou, L., 2024. An accurate monitoring method of peanut southern blight using unmanned aerial vehicle remote sensing. Precis. Agric. 25 (4), 1857–1876. https://doi.org/10.1007/s11119-024-10137-w.

Haralick, R.M., Shanmugam, K., Dinstein, I., 1973. Textural features for image classification. IEEE Trans. Syst. Man Cybern. SMC-3(6), 610–621. https://doi.org/10.1109/TSMC.1973.4309314.

Hardisky, M.A., Klemas, V., Smart, R., 1983. The influence of soil salinity, growth form, and leaf moisture on the spectral radiance of *Spartina alterniflora* canopies. Photogramm. Eng. Remote. Sens. 49, 77–83.

Hu, Y., An, R., Wang, B., Xing, F., Ju, F., 2020. Shape adaptive neighborhood information-based semi-supervised learning for hyperspectral image classification. Remote Sens. 12 (18), 2976. https://doi.org/10.3390/rs12182976.

Huang, R., Xiao, Y., Li, S., Li, J., Weng, W., Shao, Q., Zhang, J., Zhang, Y., Yang, L., Huang, C., Sun, W., Liu, W., Jin, H., Huang, J., 2025. A novel framework for dynamic and quantitative mapping of damage severity due to compound drought-heatwave impacts on tea plantations, integrating sentinel-2 and UAV images. Comput. Electron. Agric. 228, 109688. https://doi.org/10.1016/j.compag.2024.109688.

Huete, A., Didan, K., Miura, T., Rodriguez, E.P., Gao, X., Ferreira, L.G., 2002. Overview of the radiometric and biophysical performance of the MODIS vegetation indices. Remote Sens. Environ. 83 (1–2), 195–213. https://doi.org/10.1016/S0034-4257(02)00096-2.

Joshi, P., Sandhu, K.S., Dhillon, G.S., Chen, J., Bohara, K., 2024. Detection and monitoring wheat diseases using unmanned aerial vehicles (UAVs). Comput. Electron. Agric. 224, 109158. https://doi.org/10.1016/j.compag.2024.109158.

Karasiak, N., Dejoux, J.F., Monteil, C., Sheeren, D., 2022. Spatial dependence between training and test sets: another pitfall of classification accuracy assessment in remote sensing. Mach. Learn. 111 (7), 2715–2740. https://doi.org/10.1007/s10994-021-05972-1.

Kim, M., Park, S., Anderson, C., Stensaas, G.L., 2022. System characterization report on planet's superdove (No. 2021–1030-F). US Geological Survey https://doi.org/10.3133/ofr20211030F.

Kononenko, I., 1994. Estimating attributes: Analysis and extensions of RELIEF. European conference on machine learning, pp. 171–182 https://doi.org/10.1007/3-540-57868-4_57.

Kouadio, L., El Jarroudi, M., Belabess, Z., Laasli, S.E., Roni, M.Z.K., Amine, I.D.I., Mokhtari, N., Mokrini, F., Junk, J., Lahlali, R., 2023. A review on UAV-based applications for plant disease detection and monitoring. Remote Sens. 15 (17), 4273. https://doi.org/10.3390/rs15174273.

Li, H., Huang, J., Zhang, C., Ning, X., Zhang, S., Atkinson, P.M., 2024. An efficient and generalisable approach for mapping paddy rice fields based on their unique spectra during the transplanting period leveraging the CIE colour space. Remote Sens. Environ. 313, 114381. https://doi.org/10.1016/j.rse.2024.114381.

Li, M., Li, S., Zhang, Z., Du, P., Zhou, L., Dong, W., 2021. Sensitivity to carboxin and population diversity of *Sclerotium rolfsii* from peanut in Henan province. J. Henan Agric. Sci. 50 (5), 64–73. https://doi.org/10.15933/j.cnki.1004-3268.2021.05.009.

Li, P., Wang, Y., Wang, C., Tian, L., Lin, M., Xu, S., Zhu, C., 2025a. A comparison of recent global time-series land cover products. Remote Sens. 17 (8), 1417. https://doi.org/10.3390/rs17081417.

Li, S., Dong, P., Zhang, H., Xu, X., Shi, L., Sun, T., Qiao, H., Yue, J., Guo, W., 2025b. Integrating ecological niche and epidemiological models to predict wheat fusarium head blight using remote sensing and meteorological data. Comput. Electron. Agric. 234, 110255. https://doi.org/10.1016/j.compag.2025.110255.

Lillesand, T., Kiefer, R.W., Chipman, J., 2015. Remote Sensing and Image Interpretation. seventh ed. John Wiley & Sons, Hoboken, NJ.

Liu, D., Mao, X., Zhang, G., He, L., Wang, L., Zhang, F., Wang, Q., Zhou, L., 2024a. Antifungal activity and mechanism of physcion against *Sclerotium rolfsii*, the causal agent of peanut southern blight. J. Agric. Food Chem. 72 (28), 15601–15612. https://doi.org/10.1021/acs.jafc.4c02519.

Liu, M., He, W., Zhang, H., 2024b. WPS: a whole phenology-based spectral feature selection method for mapping winter crop from time-series images. ISPRS J. Photogramm. Remote Sens. 210, 141–159. https://doi.org/10.1016/j.isprsjprs.2024.03.005.

Liu, Y., Liu, M., Liu, G., Sun, H., An, L., Zhao, R., Tang, W., Zhao, F., Yan, X., Ma, Y., Li, M., 2024c. Estimating canopy chlorophyll content of powdery mildew stressed winter wheat by different spatial resolutions of UAV-imagery. Comput. Electron. Agric. 227, 109621. https://doi.org/10.1016/j.compag.2024.109621.

Lu, L., Luo, J., Xin, Y., Duan, H., Sun, Z., Qiu, Y., Xiao, Q., 2022. How can UAV contribute in satellite-based *Phragmites australis* aboveground biomass estimating? Int. J. Appl. Earth Obs. Geoinf. 114, 103024. https://doi.org/10.1016/j.jag.2022.103024.

Ma, H., Jing, Y., Huang, W., Shi, Y., Dong, Y., Zhang, J., Liu, L., 2018. Integrating early growth information to monitor winter wheat powdery mildew using multi-temporal landsat-8 imagery. Sensors-Basel. 18 (10), 3290. https://doi.org/10.3390/s18103290.

Ma, H., Huang, W., Jing, Y., Yang, C., Han, L., Dong, Y., Ye, H., Shi, Y., Zheng, Q., Liu, L., Ruan, C., 2019. Integrating growth and environmental parameters to discriminate powdery mildew and aphid of winter wheat using bi-temporal landsat-8 imagery. Remote Sens. 11 (7), 846. https://doi.org/10.3390/rs11070846.

Ma, H., Zhang, J., Huang, W., Ruan, C., Chen, D., Zhang, H., Zhou, X., Gui, Z., 2024a. Monitoring yellow rust progression during spring critical wheat growth periods using multi-temporal sentinel-2 imagery. Pest Manag. Sci. 80 (12), 6082–6095. https://doi.org/10.1002/ps.8336.

Ma, R., Zhang, N., Zhang, X., Bai, T., Yuan, X., Bao, H., He, D., Sun, W., He, Y., 2024b. Cotton verticillium wilt monitoring based on UAV multispectral-visible multi-source feature fusion. Comput. Electron. Agric. 217, 108628. https://doi.org/10.1016/j.compag.2024.108628.

Maestri, D., 2024. Groundnut and tree nuts: a comprehensive review on their lipid components, phytochemicals, and nutraceutical properties. Crit. Rev. Food Sci. Nutr. 64 (21), 7426–7450. https://doi.org/10.1080/10408398.2023.2185202.

Mao, W., Wang, Y., Wang, Y., 2003. Real-time detection of between-row weeds using machine vision. 2003 ASAE Annual Meeting. https://doi.org/10.13031/2013.15381 031004

Merzlyak, M.N., Gitelson, A.A., Chivkunova, O.B., Rakitin, V.Y., 1999. Non-destructive optical detection of pigment changes during leaf senescence and fruit ripening. Physiol. Plant. 106 (1), 135–141. https://doi.org/10.1034/j.1399-3054.1999.106119.x.

Najem, S., Baghdadi, N., Bazzi, H., Zribi, M., 2024. Incidence angle normalization of c-band radar backscattering coefficient over agricultural surfaces using dynamic cosine method. Remote Sens. 16 (20), 3838. https://doi.org/10.3390/rs16203838.

Nakalembe, C., Becker-Reshef, I., Bonifacio, R., Hu, G., Humber, M.L., Justice, C.J., Keniston, J., Mwangi, K., Rembold, F., Shukla, S., Urbano, F., Whitcraft, A.K., Li, Y., Zappacosta, M., Jarvis, I., Sanchez, A., 2021. A review of satellite-based global agricultural monitoring systems available for Africa. Glob. Food Secur. 29, 100543. https://doi.org/10.1016/j.gfs.2021.100543.

Neupane, K., Baysal-Gurel, F., 2021. Automatic identification and monitoring of plant diseases using unmanned aerial vehicles: a review. Remote Sens. 13 (19), 3841. https://doi.org/10.3390/rs13193841.

Pasquarella, V.J., Brown, C.F., Czerwinski, W., Rucklidge, W.J., 2023. Comprehensive quality assessment of optical satellite imagery using weakly supervised video learning. Proceedings of the IEEE/CVF Conference on Computer Vision and PATTERN recognition, pp. 2125–2135 https://doi.org/10.1109/CVPRW59228.2023.00206.

Peng, Y., Qiu, B., Tang, Z., Xu, W., Yang, P., Wu, W., Chen, X., Zhu, X., Zhu, P., Zhang, X., Wang, X., Zhang, C., Wang, L., Li, M., Liang, J., Huang, Y., Cheng, F., Chen, J., Wu, F., Jian, Z., Li, Z., 2024. Where is tea grown in the world: a robust mapping framework for agroforestry crop with knowledge graph and sentinels images. Remote Sens. Environ. 303, 114016. https://doi.org/10.1016/j.rse.2024.114016.

Pu, R., Gong, P., Yu, Q., 2008. Comparative analysis of EO-1 ALI and hyperion, and landsat ETM+ data for mapping forest crown closure and leaf area index. Sensors-Basel. 8 (6), 3744–3766. https://doi.org/10.3390/s8063744.

Punja, Z.K., 1985. The biology, ecology, and control of *Sclerotium rolfsii*. Annu. Rev. Phytopathol. 23 (1), 97–127. https://doi.org/10.1146/annurev.py.23.090185.000525.

Qiu, B., Luo, Y., Tang, Z., Chen, C., Lu, D., Huang, H., Chen, Y., Chen, N., Xu, W., 2017. Winter wheat mapping combining variations before and after estimated heading dates. ISPRS J. Photogramm. Remote Sens. 123, 35–46. https://doi.org/10.1016/j.isprsjprs.2016.09.016.

Rahali, L., Praticò, S., Lanucara, S., Modica, G., 2025. CubeSat constellations: new era for precision agriculture? Comput. Electron. Agric. 230, 109764. https://doi.org/10.1016/j.compag.2024.109764.

Rideout, S.L., Brenneman, T.B., Culbreath, A.K., Langston Jr., D.B., 2008. Evaluation of weather-based spray advisories for improved control of peanut stem rot. Plant Dis. 92 (3), 392–400. https://doi.org/10.1094/pdis-92-3-0392.

Robnik-Šikonja, M., Kononenko, I., 2003. Theoretical and empirical analysis of ReliefF and RReliefF. Mach. Learn. 53, 23–69. https://doi.org/10.1023/A:1025667309714.

Rodríguez-Puerta, F., Perroy, R.L., Barrera, C., Price, J.P., García-Pascual, B., 2024. Five-year evaluation of sentinel-2 cloud-free mosaic generation under varied cloud cover conditions in Hawaii. Remote Sens. 16 (24), 4791. https://doi.org/10.3390/rs16244791.

Rommel, E., Giese, L., Fricke, K., Kathöfer, F., Heuner, M., Mölter, T., Deffert, P., Asgari, M., Näthe, P., Dzunic, F., Rock, G., Bongartz, J., Burkart, A., Quick, I., Schröder, U., Baschek, B., 2022. Very high-resolution imagery and machine learning for detailed mapping of riparian vegetation and substrate types. Remote Sens. 14 (4), 954. https://doi.org/10.3390/rs14040954.

Rondeaux, G., Steven, M., Baret, F., 1996. Optimization of soil-adjusted vegetation indices. Remote Sens. Environ. 55 (2), 95–107. https://doi.org/10.1016/0034-4257(95)00186-7.

Roujean, J.L., Breon, F.M., 1995. Estimating PAR absorbed by vegetation from bidirectional reflectance measurements. Remote Sens. Environ. 51 (3), 375–384. https://doi.org/10.1016/0034-4257(94)00114-3.

Sanjel, S., Colee, J., Barocco, R.L., Dufault, N.S., Tillman, B.L., Punja, Z.K., Seepaul, R., Small, I.M., 2024. Environmental factors influencing stem rot development in peanut: predictors and action thresholds for disease management. Phytopathology 114 (2), 393–404. https://doi.org/10.1094/phyto-05-23-0164-r.

Santoso, H., Gunawan, T., Jatmiko, R.H., Darmosarkoro, W., Minasny, B., 2011. Mapping and identifying basal stem rot disease in oil palms in North Sumatra with QuickBird imagery. Precis. Agric. 12, 233–248. https://doi.org/10.1007/s11119-010-9172-7.

Sconyers, L.E., Brenneman, T.B., Stevenson, K.L., Mullinix, B.G., 2005. Effects of plant spacing, inoculation date, and peanut cultivar on epidemics of peanut stem rot and tomato spotted wilt. Plant Dis. 89 (9), 969–974. https://doi.org/10.1094/PD-89-0969.

Shi, C., Wang, L., 2014. Incorporating spatial information in spectral unmixing: a review. Remote Sens. Environ. 149, 70–87. https://doi.org/10.1016/j.rse.2014.03.034.

Sishodia, R.P., Ray, R.L., Singh, S.K., 2020. Applications of remote sensing in precision agriculture: a review. Remote Sens. 12 (19), 3136. https://doi.org/10.3390/rs12193136.

Snoek, J., Larochelle, H., Adams, R.P., 2012. Practical Bayesian optimization of machine learning algorithms. Proceedings of the 26th International Conference on Neural Information Processing Systems, pp. 2951–2959 https://doi.org/10.5555/2999325.2999464.

Somers, B., Asner, G.P., 2013. Multi-temporal hyperspectral mixture analysis and feature selection for invasive species mapping in rainforests. Remote Sens. Environ. 136, 14–27. https://doi.org/10.1016/j.rse.2013.04.006.

Song, X., Yang, C., Wu, M., Zhao, C., Yang, G., Hoffmann, W.C., Huang, W., 2017. Evaluation of sentinel-2A satellite imagery for mapping cotton root rot. Remote Sens. 9 (9), 906. https://doi.org/10.3390/rs9090906.

Su, J., Liu, C., Coombes, M., Hu, X., Wang, C., Xu, X., Li, Q., Guo, L., Chen, W.H., 2018. Wheat yellow rust monitoring by learning from multispectral UAV aerial imagery. Comput. Electron. Agric. 155, 157–166. https://doi.org/10.1016/j.compag.2018.10.017.

Sun, H., Zhou, L., Shu, M., Zhang, J., Feng, Z., Feng, H., Song, X., Yue, J., Guo, W., 2024. Estimation of peanut southern blight severity in hyperspectral data using the synthetic minority oversampling technique and fractional-order differentiation. Agriculture-Basel. 14 (3), 476. https://doi.org/10.3390/agriculture14030476.

Tamiminia, H., Salehi, B., Mahdianpari, M., Quackenbush, L., Adeli, S., Brisco, B., 2020. Google earth engine for geo-big data applications: a meta-analysis and systematic review. ISPRS J. Photogramm. Remote Sens. 164, 152–170. https://doi.org/10.1016/j.isprsjprs.2020.04.001.

Tian, Y., Zhao, F., Meng, R., Sun, R., Zhang, Y., Shen, Y., Wang, B., Liu, J., Li, M., 2025. A vision foundation model-based method for large-scale forest disturbance mapping using time series sentinel-1 SAR data. Remote Sens. Environ. 325, 114775. https://doi.org/10.1016/j.rse.2025.114775.

Tucker, C.J., 1979. Red and photographic infrared linear combinations for monitoring vegetation. Remote Sens. Environ. 8 (2), 127–150. https://doi.org/10.1016/0034-4257(79)90013-0.

Urbanowicz, R.J., Meeker, M., La Cava, W., Olson, R.S., Moore, J.H., 2018. Relief-based feature selection: introduction and review. J. Biomed. Inform. 85, 189–203. https://doi.org/10.1016/j.jbi.2018.07.014.

Variath, M.T., Janila, P., 2017. Economic and academic importance of peanut. In: Varshney, R., Pandey, M., Puppala, N. (Eds.), The Peanut Genome. Springer, Cham, pp. 7–26 https://doi.org/10.1007/978-3-319-63935-2_2.

Velastegui-Montoya, A., Montalván-Burbano, N., Carrión-Mero, P., Rivera-Torres, H., Sadeck, L., Adami, M., 2023. Google earth engine: a global analysis and future trends. Remote Sens. 15 (14), 3675. https://doi.org/10.3390/rs15143675.

Venter, Z.S., Barton, D.N., Chakraborty, T., Simensen, T., Singh, G., 2022. Global 10 m land use land cover datasets: a comparison of dynamic world, world cover and esri land cover. Remote Sens. 14 (16), 4101. https://doi.org/10.3390/rs14164101.

Wang, J., Xiao, X., Liu, L., Wu, X., Qin, Y., Steiner, J.L., 2020. Mapping sugarcane plantation dynamics in Guangxi, China, by time series sentinel-1, sentinel-2 and landsat images. Remote Sens. Environ. 247, 111951. https://doi.org/10.1016/j.rse.2020.111951.

Wang, Q., Tang, Y., Ge, Y., Xie, H., Tong, X., Atkinson, P.M., 2023a. A comprehensive review of spatial-temporal-spectral information reconstruction techniques. Sci. Remote Sens. 8, 100102. https://doi.org/10.1016/j.srs.2023.100102.

Wang, X., Fang, S., Yang, Y., Du, J., Wu, H., 2023b. A new method for crop type mapping at the regional scale using multi-source and multi-temporal sentinel imagery. Remote Sens. 15 (9), 2466. https://doi.org/10.3390/rs15092466.

Wang, Y., Sun, Y., Cao, X., Wang, Y., Zhang, W., Cheng, X., 2023c. A review of regional and global scale land use/land cover (LULC) mapping products generated from satellite remote sensing. ISPRS J. Photogramm. Remote Sens. 206, 311–334. https://doi.org/10.1016/j.isprsjprs.2023.11.014.

Wen, T., Liu, J., Fu, Y., Yue, J., Li, Y., Guo, W., 2024. Development and evaluation of a new spectral index to detect peanut southern blight disease using canopy hyperspectral reflectance. Horticulturae. 10 (2), 128. https://doi.org/10.3390/horticulturae10020128.

Woodward, J.E., Brenneman, T.B., Kemerait Jr., R.C., 2013. Chemical control of peanut diseases: Targeting leaves, stems, roots, and pods with foliar-applied fungicides. In: Nita, M. (Ed.), Fungicides – Showcases of Integrated Plant Disease Management from around the World. IntechOpen, Rijeka, Croatia, pp. 55–76 https://doi.org/10.5772/51064.

Xiao, X.M., Boles, S., Liu, J.Y., Zhuang, D.F., Frolking, S., Li, C.S., Salas, W., Moore, B., 2005. Mapping paddy rice agriculture in southern China using multi-temporal MODIS images. Remote Sens. Environ. 95 (4), 480–492. https://doi.org/10.1016/j.rse.2004.12.009.

Xiao, Y., Dong, Y., Huang, W., Liu, L., Ma, H., 2021. Wheat fusarium head blight detection using UAV-based spectral and texture features in optimal window size. Remote Sens. 13 (13), 2437. https://doi.org/10.3390/rs13132437.

Xie, C., Huang, C.H., Vallad, G.E., 2014. Mycelial compatibility and pathogenic diversity among *Sclerotium rolfsii* isolates in the southern United States. Plant Dis. 98 (12), 1685–1694. https://doi.org/10.1094/pdis-08-13-0861-re.

Xu, B., Meng, R., Chen, G., Liang, L., Lv, Z., Zhou, L., Sun, R., Zhao, F., Yang, W., 2023. Improved weed mapping in corn fields by combining UAV-based spectral, textural, structural, and thermal measurements. Pest Manag. Sci. 79 (7), 2591–2602. https://doi.org/10.1002/ps.7443.

Xu, H., Su, G., Li, C., Deng, W., 2025. The differences, advantages, and disadvantages of various image compositing methods on the Google Earth Engine platform: an exploration. Geo-spat. Inf. Sci., 1–28 https://doi.org/10.1080/10095020.2025.2542534.

Xu, P., Tsendbazar, N.E., Herold, M., de Bruin, S., Koopmans, M., Birch, T., Carter, S., Fritz, S., Lesiv, M., Mazur, E., Pickens, A., Potapov, P., Stolle, F., Tyukavina, A., Van De Kerchove, R., Zanaga, D., 2024. Comparative validation of recent 10 m-resolution global land cover maps. Remote Sens. Environ. 311, 114316. https://doi.org/10.1016/j.rse.2024.114316.

Xun, L., Zhang, J., Cao, D., Yang, S., Yao, F., 2021. A novel cotton mapping index combining sentinel-1 SAR and sentinel-2 multispectral imagery. ISPRS J. Photogramm. Remote Sens. 181, 148–166. https://doi.org/10.1016/j.isprsjprs.2021.08.021.

Yang, G., Li, X., Liu, P., Yao, X., Zhu, Y., Cao, W., Cheng, T., 2023. Automated in-season mapping of winter wheat in China with training data generation and model transfer. ISPRS J. Photogramm. Remote Sens. 202, 422–438. https://doi.org/10.1016/j.isprsjprs.2023.07.004.

Ye, H., Huang, W., Huang, S., Cui, B., Dong, Y., Guo, A., Ren, Y., Jin, Y., 2020. Recognition of banana fusarium wilt based on UAV remote sensing. Remote Sens. 12 (6), 938. https://doi.org/10.3390/rs12060938.

Yuan, L., Zhang, J., Zhao, J., Huang, L., Yang, X., Wang, J., 2013. Mapping of powdery mildew using multi-spectral HJ-CCD image in Beijing suburban area. Optik. 124 (21), 4734–4738. https://doi.org/10.1016/j.ijleo.2013.01.103.

Yuan, L., Zhang, J., Shi, Y., Nie, C., Wei, L., Wang, J., 2014. Damage mapping of powdery mildew in winter wheat with high-resolution satellite image. Remote Sens. 6 (5), 3611–3623. https://doi.org/10.3390/rs6053611.

Yuan, L., Bao, Z., Zhang, H., Zhang, Y., Liang, X., 2017. Habitat monitoring to evaluate crop disease and pest distributions based on multi-source satellite remote sensing imagery. Optik. 145, 66–73. https://doi.org/10.1016/j.ijleo.2017.06.071.

Yuan, L., Yu, Q., Xiang, L., Zeng, F., Dong, J., Xu, O., Zhang, J., 2025. Integrating UAV and high-resolution satellite remote sensing for multi-scale rice disease monitoring. Comput. Electron. Agric. 234, 110287. https://doi.org/10.1016/j.compag.2025.110287.

Zhang, C., Marzougui, A., Sankaran, S., 2020. High-resolution satellite imagery applications in crop phenotyping: An overview. Comput. Electron. Agric. 175, 105584. https://doi.org/10.1016/j.compag.2020.105584.

Zhang, C., Kerner, H., Wang, S., Hao, P., Li, Z., Hunt, K.A., Abernethy, J., Zhao, H., Gao, F., Di, L., Guo, C., Liu, Z., Yang, Z., Mueller, R., Boryan, C., Chen, Q., Beeson, P.C., Zhang, H.K., Shen, Y., 2025. Remote sensing for crop mapping: a perspective on current and future crop-specific land cover data products. Remote Sens. Environ. 330, 114995. https://doi.org/10.1016/j.rse.2025.114995.

Zhang, H., Wang, L., Tian, T., Yin, J., 2021. A review of unmanned aerial vehicle low-altitude remote sensing (UAV-LARS) use in agricultural monitoring in China. Remote Sens. 13 (6), 1221. https://doi.org/10.3390/rs13061221.

Zhang, H., He, J., Chen, S., Zhan, Y., Bai, Y., Qin, Y., 2023. Comparing three methods of selecting training samples in supervised classification of multispectral remote sensing images. Sensors-Basel. 23 (20), 8530. https://doi.org/10.3390/s23208530.

Zhang, J., Pu, R., Yuan, L., Wang, J., Huang, W., Yang, G., 2014. Monitoring powdery mildew of winter wheat by using moderate resolution multi-temporal satellite imagery. PLoS One. 9 (4), e93107. https://doi.org/10.1371/journal.pone.0093107.

Zhao, G., Zhang, Y., Lan, Y., Deng, J., Zhang, Q., Zhang, Z., Li, Z., Liu, L., Huang, X., Ma, J., 2023. Application progress of UAV-LARS in identification of crop diseases and pests. Agronomy-Basel. 13 (9), 2232. https://doi.org/10.3390/agronomy13092232.

Zhao, Q., Yu, L., Li, X., Peng, D., Zhang, Y., Gong, P., 2021. Progress and trends in the application of google earth and google earth engine. Remote Sens. 13 (18), 3778. https://doi.org/10.3390/rs13183778.

Zheng, Q., Huang, W., Cui, X., Shi, Y., Liu, L., 2018. New spectral index for detecting wheat yellow rust using sentinel-2 multispectral imagery. Sensors-Basel. 18 (3), 868. https://doi.org/10.3390/s18030868.

Zheng, Z., Yuan, J., Yao, W., Kwan, P., Yao, H., Liu, Q., Guo, L., 2024. Fusion of UAV-acquired visible images and multispectral data by applying machine-learning methods in crop classification. Agronomy-Basel. 14 (11), 2670. https://doi.org/10.3390/agronomy14112670.

Zhou, Z.H., 2012. Ensemble Methods: Foundations and Algorithms. first ed. CRC Press, Boca Raton, FL https://doi.org/10.1201/b12207.

Zhu, W., Feng, Z., Dai, S., Zhang, P., Wei, X., 2022. Using UAV multispectral remote sensing with appropriate spatial resolution and machine learning to monitor wheat scab. Agriculture-Basel. 12 (11), 1785. https://doi.org/10.3390/agriculture12111785.