



OPEN ACCESS

EDITED BY

Khurram Shahzad,
Lasbela University of Agriculture, Pakistan

REVIEWED BY

Belén Diezma,
Polytechnic University of Madrid, Spain
Yeison Alberto Garcés-Gómez,
Catholic University of Manizales, Colombia

*CORRESPONDENCE

Dailiang Peng

✉ pengdl@aircas.ac.cn

Changyong Dou

✉ doucy@aircas.ac.cn

RECEIVED 30 June 2025

ACCEPTED 29 September 2025

PUBLISHED 20 October 2025

CITATION

Zhang H, Peng D, Dou C, Lou Z, Zhang X,
Yu L, Song K, Zhang Y, Hu J, Zheng S, Lv Y,
Liu S, Zhang Y and Peng H (2025)
Enhancing early-season soybean
identification through optical and
SAR time-series integration.
Front. Plant Sci. 16:1656628.
doi: 10.3389/fpls.2025.1656628

COPYRIGHT

© 2025 Zhang, Peng, Dou, Lou, Zhang, Yu,
Song, Zhang, Hu, Zheng, Lv, Liu, Zhang and
Peng. This is an open-access article distributed
under the terms of the [Creative Commons
Attribution License \(CC BY\)](#). The use,
distribution or reproduction in other forums
is permitted, provided the original author(s)
and the copyright owner(s) are credited and
that the original publication in this journal is
cited, in accordance with accepted academic
practice. No use, distribution or reproduction
is permitted which does not comply with
these terms.

Enhancing early-season soybean identification through optical and SAR time-series integration

Hongchi Zhang^{1,2,3}, Dailiang Peng^{1,2,3*}, Changyong Dou^{1,2,3*},
Zihang Lou⁴, Xiaoyang Zhang⁵, Le Yu⁶, Kaishan Song⁷,
Yaqiong Zhang⁸, Jinkang Hu^{1,2,3}, Shijun Zheng⁶, Yulong Lv^{1,2,3},
Shengyi Liu^{1,9}, Yizhou Zhang^{1,2,3} and Hao Peng^{3,10}

¹Key Laboratory of Digital Earth Science, Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing, China, ²International Research Center of Big Data for Sustainable Development Goals, Beijing, China, ³College of Resource and Environment, University of Chinese Academy of Sciences, Beijing, China, ⁴Institute of Agricultural Remote Sensing and Information Technology Application, College of Environmental and Resource Sciences, Zhejiang University, Hangzhou, China, ⁵Geospatial Sciences Center of Excellence, Department of Geography Geospatial Sciences, South Dakota State University, Brookings, SD, United States, ⁶Department of Earth System Science, Tsinghua University, Beijing, China, ⁷State Key Laboratory of Black Soils Conservation and Utilization, Northeast Institute of Geography and Agroecology, Chinese Academy of Sciences, Changchun, China, ⁸Satellite Application Center for Ecology and Environment, Ministry of Ecology and Environment, Beijing, China, ⁹College of Forestry, Shenyang Agricultural University, Shenyang, Liaoning, China, ¹⁰State Key Laboratory of Desert and Oasis Ecology, Xinjiang Institute of Ecology and Geography, Chinese Academy of Sciences, Urumqi, Xinjiang, China

Soybean is an important grain and cash crop in China, and timely knowledge of its distribution is crucial for food security. However, traditional survey methods are time-consuming and limited in coverage. In contrast, satellite remote sensing enables large-scale, continuous, and cost-effective monitoring, providing reliable support for crop classification and yield forecasting. However, the high spectral similarity between soybean and maize during key phenological stages presents a major challenge for reliable classification. To address this, we propose a multi-source remote sensing approach that integrates Sentinel-1 SAR and Sentinel-2 optical time-series imagery. This method combines statistical descriptors, harmonic fitting parameters, phenological indicators, and radar-based features within a random forest classifier to achieve accurate soybean mapping. The study was conducted in the Jiusan Reclamation Area of Heilongjiang Province using satellite imagery from May to October 2019 for multi-source classification and temporal analysis. We systematically evaluated classification performance across different data sources and phenological stages and introduced the Earliest Identifiable Time (EIT) metric to assess temporal detection capabilities. Results show that the multi-source fusion approach outperforms single-source methods, achieving an overall accuracy (OA) of 96.85%, a Kappa coefficient of 0.9493, and an F1-score of 95.84% for soybean. Notably, SAR data significantly improved classification during the flowering stage—when optical imagery is often constrained—resulting in a maximum F1-score increase of 6.96%. Soybean classification accuracy increased rapidly with crop development, and the EIT was advanced to Day of Year (DOY) 210, approximately 20 days earlier than with

optical data alone. These findings demonstrate the effectiveness of multi-source remote sensing in enhancing both the accuracy and timeliness of crop classification under complex climatic conditions, offering valuable support for precise soybean mapping and in-season monitoring.

KEYWORDS

soybean mapping methods, remote sensing, Sentinel-1/2, time-series analysis, early-season crop identification

1 Introduction

Soybean (*Glycine max*) is one of the most important plant-based protein crops worldwide. Its high protein and oil content have made it a globally cultivated crop, playing a central role in both food and feed systems (Schmutz et al., 2010; Liu et al., 2017). In China, soybean serves as both a staple and an economic crop, yet domestic production meets only a small fraction of national demand (She et al., 2024). Accurate and timely information on the spatial distribution of soybean cultivation is essential for assessing planting scale, informing policy implementation, and supporting efforts to boost domestic production (Johnson and Mueller, 2021; Wei et al., 2023).

Remote sensing has emerged as a powerful tool for large-scale, dynamic, and continuous land surface monitoring. It enables efficient identification and spatiotemporal analysis of agricultural features (Weiss et al., 2020; Bian et al., 2023), meeting the precision requirements of modern smart agriculture while reducing labor and financial input. By providing timely and consistent crop information from space, satellite technology helps stabilize supply chains, supports better policy decisions, and ultimately strengthens national and global food security. However, in major soybean-producing regions of China, soybean, maize, and rice are the dominant crops cultivated within the same agricultural landscape. Among these, soybean and maize exhibit high phenological and spectral similarity, especially during the early pod-setting stage, making their discrimination via single-source optical data extremely challenging (Chen et al., 2023; You et al., 2023; Zhang et al., 2024). Meanwhile, as a typical paddy crop, rice differs from upland soybean and maize in water demand and growth environment, yet its spectra can overlap with soybean in specific phenological phases, like seedling stage with sparse canopy, under cloud-contaminated optical data. Thus, developing high-accuracy and efficient mapping methods that simultaneously distinguish soybean from maize and rice is critical for optimizing cropping structures, increasing yields, and ensuring national food security (Li et al., 2023).

The increasing availability of free satellite imagery from platforms such as Sentinel-1 and Sentinel-2 has transformed agricultural monitoring, enabling high-resolution observations in spatial, temporal, radiometric, and spectral dimensions (Roy et al.,

2014; Defourny et al., 2019; Khan et al., 2024). The integration of these data with advanced algorithms and cloud-based platforms like Google Earth Engine has made large-scale, field-level crop mapping more accessible and operational (Dong et al., 2016; Gorelick et al., 2017). Time-series remote sensing is particularly valuable for capturing dynamic crop phenology and reducing spectral confusion between similar crop types, such as maize and soybean, by leveraging the structured temporal signals embedded in satellite observations (You et al., 2021; Huang et al., 2024). Machine learning algorithms, especially Random Forest (RF), are widely used in crop classification due to their robustness, computational efficiency, and ability to handle high-dimensional, noisy data (Li et al., 2021; Song et al., 2021). RF performs well with time-series data and supports variable importance ranking, making it suitable for large-scale agricultural applications (You et al., 2021).

Although high-resolution optical imagery is effective for crop classification, especially at the end of the growing season, its utility is often limited in humid or rainy regions due to persistent cloud cover (Inglada et al., 2016; Jiao et al., 2022; Zhu et al., 2022; Maleki et al., 2024). Synthetic Aperture Radar (SAR), particularly from Sentinel-1, offers an all-weather, day-and-night imaging capability that ensures consistent data acquisition regardless of atmospheric conditions (Asam et al., 2022). SAR is sensitive to vegetation structure, biomass, and moisture content, complementing the spectral reflectance captured by optical sensors (Sun et al., 2019; Liao et al., 2020). For example, SAR can distinguish rice (with flooded fields in the early stage) from upland soybean/maize via specular reflection signals from water surfaces, while optical data excels at capturing chlorophyll-related spectral differences between soybean and maize. Numerous studies have demonstrated that combining Sentinel-1 and Sentinel-2 data significantly improves crop classification accuracy, often outperforming single-source methods (Gao et al., 2018; Moumni and Lahrouni, 2021; Li et al., 2022). This fusion leverages the strengths of both sensors: optical data effectively reflect vegetation activity and phenology, while SAR provides critical structural information and maintains coverage under cloud-obscured conditions (Qu et al., 2020; Huang et al., 2022; Xuan et al., 2023), leading to more robust characterization of crop dynamics.

Early-season crop type identification, mapping crops in early growth stages or before harvest, is crucial for supporting in-season

decision-making (Inglada et al., 2016; You and Dong, 2020; Shen et al., 2025). Early classification benefits applications such as yield prediction, irrigation scheduling, and agricultural insurance, as well as land leasing and commodity trading (Konduri et al., 2020; Gallo et al., 2023; Rezaei et al., 2023). SAR's all-weather observation capability is particularly valuable in the early season when optical imagery is frequently unavailable due to cloud cover (Wei et al., 2023; Zhou et al., 2024). Previous studies have shown that fusing SAR with optical data can achieve classification accuracy equivalent to that of optical-only data obtained a month later (Inglada et al., 2016), representing a significant advancement in early-season crop mapping.

Motivated by the increasing demand for accurate and timely information on soybean and recognizing the synergistic potential of multi-source time-series remote sensing, this study aims to enhance robust crop classification and early identification. Specifically, the objectives are: (1) to develop and evaluate an integrated multi-source time-series approach using Sentinel-1 SAR and Sentinel-2 optical data for accurate classification of major crops (rice, maize, and soybean), with a focus on improving soybean mapping; and (2) to investigate the temporal evolution of classification accuracy across the growing season, with the goal of determining the earliest identifiable time—defined as the point at which the soybean F1-score exceeds 0.9 (You and Dong, 2020). Additionally, we aim to quantify the incremental contribution of SAR data in improving classification accuracy and advancing the earliest identification window, thereby highlighting its value in supporting mid-season agricultural monitoring and decision-making.

2 Materials and methods

2.1 Study area

The Jiusan Reclamation Area is situated in the northwestern part of Heilongjiang Province, China, covering parts of Heihe, Suihua, and Qiqihar. Geographically, it spans from 122°24'E to 129°31'E longitude and from 45°30'N to 51°00'N latitude. The region features gently undulating terrain, forming a transitional zone between the Lesser Khingan Mountains and the Songnen Plain. Elevation ranges from 0 to 848 meters, with a general slope from northeast to southwest. The area experiences a mid-temperate, semi-humid continental monsoon climate, characterized by dry, windy springs and warm, humid summers. Annual precipitation ranges from 400 to 550mm, primarily concentrated between June and September. This climate regime—marked by the coincidence of heat and rainfall and large diurnal temperature variation—is highly conducive to crop growth and dry matter accumulation. The Jiusan Reclamation Area is a key agricultural production base in China, particularly for soybean, maize, and rice. Soybean has been cultivated in the region for decades with consistently stable yields, while maize acreage has expanded in recent years. Rice cultivation is concentrated in peripheral areas with sufficient water availability. According to agricultural statistics from Heilongjiang Province,

Jiusan ranks among the leading regions nationwide in terms of soybean planting area and total production, playing a vital role in ensuring national food security. The geographic location of the Jiusan Reclamation Area is shown in Figure 1.

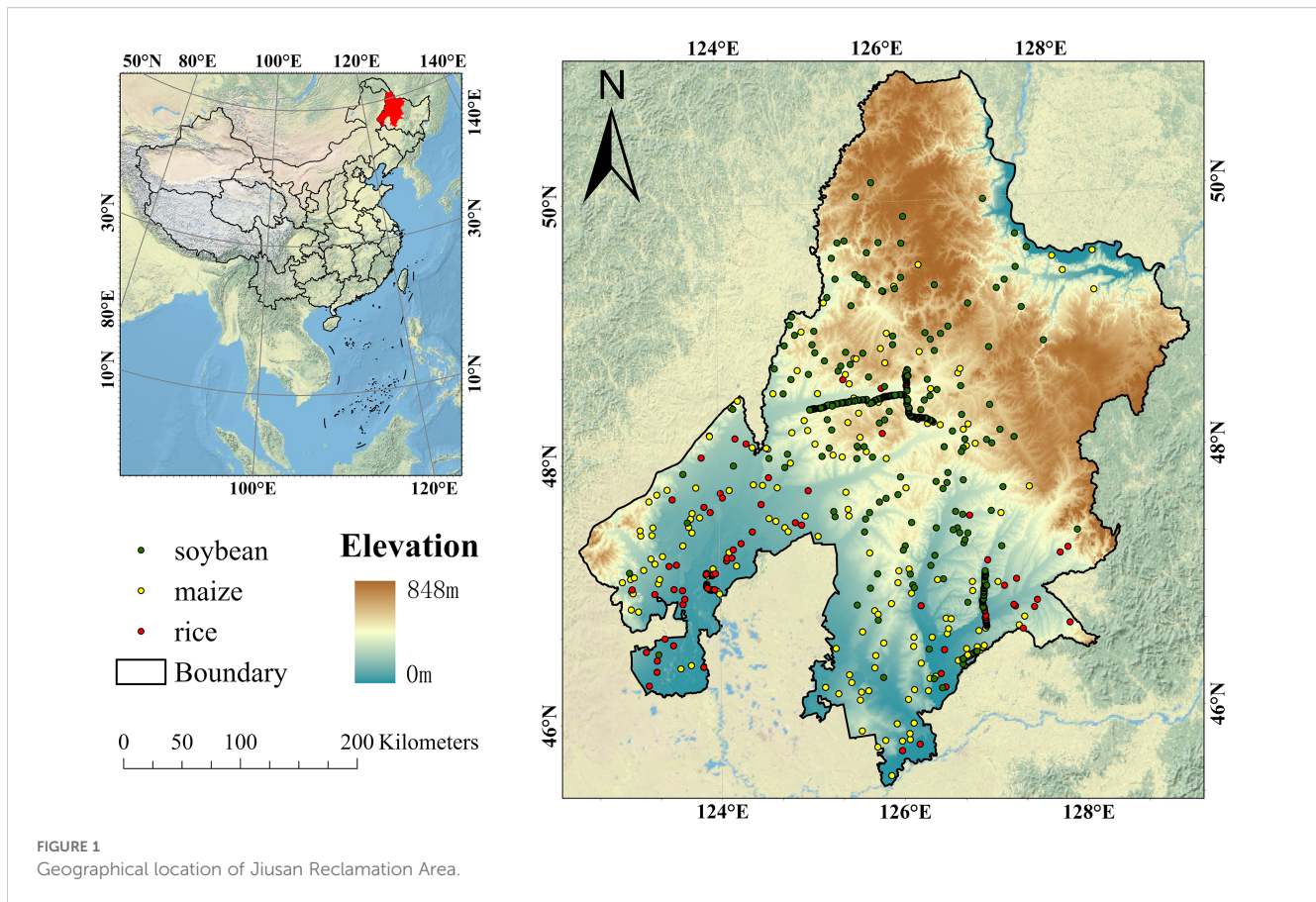
2.2 Datasets

2.2.1 Sentinel-2 imagery and preprocessing

The Sentinel-2A and Sentinel-2B satellites, operated by the European Space Agency (ESA), are equipped with the MultiSpectral Instrument (MSI), which captures imagery at spatial resolutions ranging from 10 to 60 meters and provides a revisit frequency of five days. The MSI records data across 13 spectral bands, covering the visible to shortwave infrared (SWIR) regions (Immitzer et al., 2016). In this study, we uniformly resampled all Sentinel-2 bands to 10 meters spatial resolution for consistency in multi-source data fusion and crop classification—specifically, using 10-meter resolution bands (Blue, Green, Red, NIR) directly, and resampling 20-meter bands (Red Edge 2, SWIR1, SWIR2) to 10 meters via bilinear interpolation. We employed Level-2A surface reflectance (SR) products to identify soybean cultivation in the Jiusan Reclamation Area. These products have been publicly accessible via the Google Earth Engine (GEE) data catalog since 2019. To mitigate cloud contamination, the QA60 band—a bitmask containing cloud and cirrus detection information—was used to mask out opaque clouds and cirrus. All available Sentinel-2 images acquired between May and October 2019 with cloud cover below 50% were considered. Multiple satellite orbits were included to ensure complete spatial coverage of the study area. The number of observations varied across locations, with overlapping orbital paths resulting in denser temporal coverage in certain regions. In some cases, multiple acquisitions were made on the same day at different times, potentially introducing spectral variability. Furthermore, uneven image availability due to cloud cover and masking criteria led to irregular temporal intervals between usable observations. To generate a consistent time series for each pixel, 10-day median composites were created using all valid observations within each period (Amani et al., 2020). For regions lacking high-quality data due to persistent cloudiness or other issues, linear interpolation was applied using adjacent time steps to fill temporal gaps.

2.2.2 Sentinel-1 imagery and preprocessing

The Sentinel-1A and Sentinel-1B satellites are equipped with C-band dual-polarization synthetic aperture radar (SAR) sensors, capable of acquiring imagery with spatial resolutions ranging from 5 to 40 meters and a revisit interval of 12 days. Sentinel-1 supports four imaging modes: Stripmap (SM), Interferometric Wide Swath (IW), Extra Wide Swath (EW), and Wave Mode (WV). Among these, the IW mode is commonly used for land applications and provides dual-polarization data, specifically vertical transmit–vertical receive (VV) and vertical transmit–horizontal receive (VH) polarizations. In this study, we used Sentinel-1 Ground Range Detected (GRD) products acquired in IW mode, which have a spatial resolution of 10 meters (azimuth) ×



20 meters (range); these were further resampled to 10 meters spatial resolution (consistent with Sentinel-2) using nearest-neighbor interpolation to enable pixel-level multi-source data fusion. The preprocessing workflow included the following steps: (1) thermal noise removal; (2) radiometric calibration; (3) terrain correction using Shuttle Radar Topography Mission (SRTM) or ASTER digital elevation models (DEMs); and (4) conversion of terrain-corrected backscatter coefficients into decibel (dB) values. To further suppress speckle noise inherent in SAR imagery, a Refined Lee filter with a 7×7 moving window was applied (Yommy et al., 2015). To ensure temporal alignment with Sentinel-2 optical data, 10-day composite images were generated by calculating the median value of all valid Sentinel-1 observations within each corresponding period.

2.3 Auxiliary data

2.3.1 Crop samples

This study collected a total of 1,424 ground sample points for soybean, maize, and rice in the study area for the year 2019, including 621 soybean samples, 588 maize samples, and 215 rice samples. These samples were collected with a mobile GIS device and subsequently checked against high-resolution Google Earth imagery and two seasonal Sentinel-2 RGB composites. Samples with mislabeling or located on roads and field boundaries were removed. The dataset was randomly divided into training and

validation sets with a ratio of 8:2. A cropland mask was applied to exclude non-cropland pixels and prevent their inclusion in the crop classification process. Both the crop sample data and the cropland mask used in this study were derived from the 2019 Cropland Data Layer (CDL2019) for Northeast China (You et al., 2021).

2.3.2 Statistical yearbook data

To validate the estimated soybean planting areas, official statistical data for 2019 soybean cultivation areas in Heihe, Suihua, and Qiqihar were obtained from the 2020 Heilongjiang Statistical Yearbook. For area accuracy assessment, the classified 10m resolution soybean maps were aggregated within Google Earth Engine using the pixelArea function to derive the total soybean area. The absolute difference between the remote sensing-derived soybean area and the official statistical yearbook records was then calculated to quantify area estimation error.

2.4 Methodology

The research framework of this study is illustrated in Figure 2 and consists of four main components: (1) Data Preprocessing: Sentinel-1 and Sentinel-2 time-series imagery was acquired and processed to remove outliers and cloud contamination, resulting in a clean and temporally consistent dataset. Smooth, continuous time

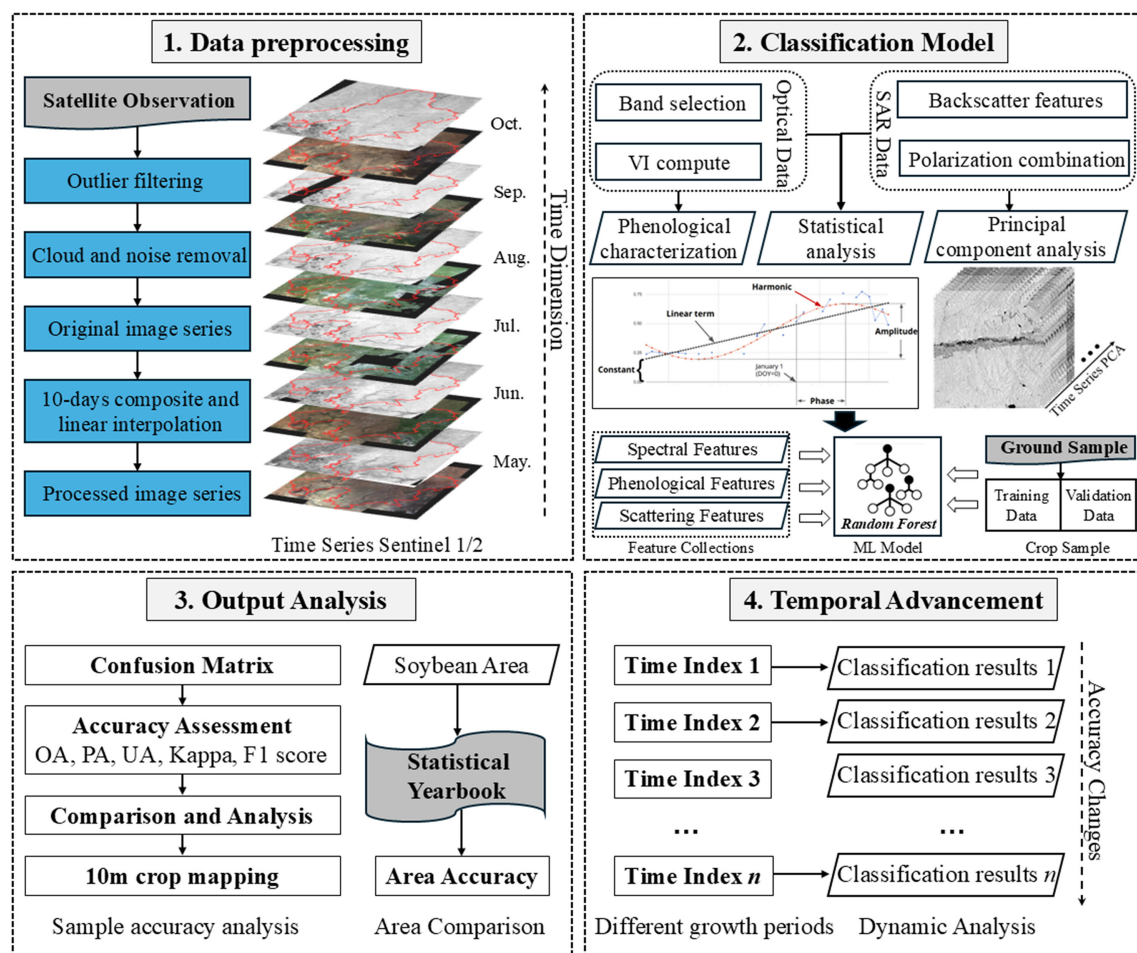


FIGURE 2
Research framework.

series were constructed through 10-day median compositing and linear interpolation. (2) Classification Model Development: Features were extracted from both optical and SAR data, including spectral bands, vegetation indices, backscatter coefficients, and polarization metrics. These were integrated with phenological indicators derived from the time series of EVI calculated from Sentinel-2 imagery, which capture the timing of crop growth onset, senescence, and season length. In addition, statistical descriptors and principal component analysis (PCA) outputs of the Sentinel-1 backscatter time series (VV, VH, and related indices) were included to reduce redundancy while retaining key temporal variations. Together, these features formed a comprehensive dataset for classification. A Random Forest classifier was then trained using ground-truth samples for crop type identification. (3) Result Analysis and Accuracy Assessment: Classification performance was evaluated using confusion matrices and five standard metrics. Overall accuracy (OA) reflects the proportion of correctly classified samples. Producer's accuracy (PA) measures omission errors, while user's accuracy (UA) measures commission errors. The Kappa coefficient adjusts OA by accounting for chance agreement. The F1-score, as the harmonic mean of PA and UA, provides a balanced measure particularly

useful when class sizes are uneven. Together, these metrics give a comprehensive evaluation of classification performance. Soybean planting area was estimated based on a 10-meter resolution classification map and compared with official statistical yearbook data to validate area estimation accuracy. (4) Time-Series Dynamic Analysis: To assess how classification accuracy evolved with crop growth, we designed a temporal progression experiment. Starting from May 1 (early sowing period), image sequences were gradually extended in 10-day steps until October 8 (harvest), resulting in 14 datasets. For each sequence, classification was performed and evaluated using the same independent validation set. This procedure allowed us to track temporal changes in accuracy and to determine the Earliest Identifiable Time (EIT), defined as the first date when the soybean F1-score exceeded 0.9.

2.4.1 Parameter selection

To improve crop classification performance, this study selected a set of parameters for feature extraction based on the physiological and biochemical characteristics of soybean growth. Two categories of spectral data were employed to classify soybean and other crop types: (1) reflectance values from six spectral bands, and (2) values from five vegetation indices (Table 1). The six spectral bands

TABLE 1 Optical vegetation index.

Indices	Formulation*
EVI	$EVI = 2.5 \times \frac{\rho_{NIR} - \rho_{red}}{\rho_{NIR} + 6 \times \rho_{red} - 7.5 \times \rho_{blue} + 1}$
LSWI	$LSWI = \frac{\rho_{NIR} - \rho_{SWIR1}}{\rho_{NIR} + \rho_{SWIR1}}$
REPI	$REPI = 705 + 35 \times \frac{(\rho_{red} + \rho_{RE3})/2 - \rho_{RE1}}{\rho_{RE2} - \rho_{RE1}}$
RENDVI	$RENDVI = \frac{\rho_{NIR} - \rho_{RE2}}{\rho_{NIR} + \rho_{RE2}}$
NDSVI	$NDPI = \frac{\rho_{SWIR1} - \rho_{Red}}{\rho_{SWIR1} + \rho_{Red}}$

*pblue, pgreen, pred, pre1, pre2, pre3, pnir, and pswir1 represent the surface reflectance of Sentinel-2 MSI bands, corresponding to Band 2 (blue, 496.6 nm (S2A)/492.1 nm (S2B)), Band 3 (green, 560 nm (S2A)/559 nm (S2B)), Band 4 (red, 664.5 nm (S2A)/665 nm (S2B)), Band 5 (Red Edge 1, 703.9 nm (S2A)/703.8 nm (S2B)), Band 6 (Red Edge 2, 740.2 nm (S2A)/739.1 nm (S2B)), Band 7 (Red Edge 3, 782.5 nm (S2A)/779.7 nm (S2B)), Band 8A (NIR, 864.8 nm (S2A)/864 nm (S2B)), and Band 11 (SWIR1, 1613.7 nm (S2A)/1610.4 nm (S2B)), respectively.

included Red (492.4 nm), Green (559.8 nm), Blue (664.6 nm), Red Edge 2 (RE2, 740 nm), Shortwave Infrared 1 (SWIR1, 1610 nm), and Shortwave Infrared 2 (SWIR2, 2190 nm). Previous studies have highlighted the potential of SWIR1, SWIR2, and RE2 bands in effectively distinguishing between maize and soybean (Luo et al., 2021; Chen et al., 2023). In addition, six widely used spectral indices were calculated: Enhanced Vegetation Index (EVI), Land Surface Water Index (LSWI), Red Edge Position Index (REPI), Red Edge Normalized Difference Vegetation Index (RENDVI), and Normalized Difference Senescence Vegetation Index (NDSVI). The EVI time series is commonly used to extract temporal and phenological features for various crops. LSWI, which is sensitive to both leaf and soil moisture, is particularly effective for distinguishing rice from upland crops such as maize and soybean. REPI and RENDVI, which utilize Sentinel-2 red edge bands, are well-suited for estimating canopy chlorophyll content and nitrogen levels.

To improve soybean identification accuracy under cloudy conditions, this study selected five key parameters from Sentinel-1 VV/VH dual-polarized SAR imagery to capture crop structural and

TABLE 2 SAR parameter index.

Parameters	Proxies	Description
Backscattering Ratio	$\sigma_{VH}^0, \sigma_{VV}^0$	Throughout the soybean growth period, alterations in the growth status and density of soybean leaves, stems, and pods can have substantial effects on the backscattering ratio (Pageot et al., 2020).
Cross-Polarization Ratio	$\frac{\sigma_{VH}^0 * \sigma_{VV}^0}{\sigma_{VV}^0 + \sigma_{VH}^0}$	Fluctuations over time in this index reflect changes in moisture content and structure that are associated with phenological development (Khabbazan et al., 2019).
RVI	$RVI = \frac{4 \times \sigma_{VH}^0}{\sigma_{VH}^0 + \sigma_{VV}^0}$	RVI can characterize both crop biomass and the LAI (Chang et al., 2018).

physical characteristics across different growth stages (Table 2). These SAR features include the backscatter coefficients of the VV and VH polarization channels (σ_{VV}^0 and σ_{VH}^0), their combined forms, and the dual-polarization Radar Vegetation Index (RVI) derived from these channels. The σ_{VV}^0 and σ_{VH}^0 coefficients reflect the radar backscattering response from vertical and horizontal plant structures, respectively (Pageot et al., 2020). VV polarization is more sensitive to vertical canopy characteristics and is thus closely related to soybean biomass and canopy density, while VH polarization responds more strongly to variations in surface roughness and structural complexity, making it useful for detecting changes during mid to late phenological stages. RVI, a normalized index computed from the VV and VH channels, effectively characterizes crop canopy vigor and growth dynamics (Mandal et al., 2020). Additionally, the Cross-Polarization Ratio (CPR), defined as the ratio of the product to the sum of σ_{VV}^0 and σ_{VH}^0 , enhances sensitivity to backscatter intensity during key phenological events such as the pod-filling stage (Velooso et al., 2017). Collectively, these SAR-derived parameters provide essential physical and temporal insights that support robust soybean classification in multi-source remote sensing applications.

2.4.2 Feature extraction

Based on the spectral curve dynamics of crops and incorporating phenology and spectral index information, this study extracted statistical features, peak growth period features, harmonic fitting features, phenological features, and SAR features to enhance the discrimination between soybeans and non-soybean crops, as summarized in Table 3.

1. Statistical Features and Peak Growth Period Features

Statistical features include the maximum, minimum, variance, and the 15th, 50th, and 90th percentiles of 11 spectral parameters throughout the entire growth season. The maximum and minimum values represent the upper and lower bounds of spectral variation, which differ significantly among crops and help distinguish soybeans from others. Variance reveals the fluctuation in spectral curves, reflecting the stability of the growth state. Percentiles provide spectral distribution information during early, middle, and late growth stages. All these statistics were calculated at the pixel level across the full time series, rather than aggregated by plot, which preserves intra-field variability and enables the detection of mixed conditions such as intercropping. Peak growth period features are extracted based on the timing of peak values in the EVI and LSWI indices. The corresponding “greenest” and “wettest” images are composited for these peak periods, from which the 11 spectral parameters are extracted as features representing the crop’s growth peak. This period typically coincides with soybean pod development, when its spectral behavior most notably differs from other crops, making these features especially critical for classification.

2. Harmonic Fitting Features

Harmonic fitting is an effective time series modeling method to extract periodic features. By treating the time series as a periodic function and applying Fourier transformation, the original curve can be reconstructed using several sine and cosine components. The function is expressed as:

$$f(t) = a_0 + \sum_{k=1}^n [a_k \cos(2\pi k\omega t) + b_k \sin(2\pi k\omega t)]$$

where $f(t)$ represents the fitted vegetation index value at time t , a_0 is the constant term, a_k and b_k are the coefficients for cosine and sine terms respectively. n is the order of the harmonic fitting, and ω is the frequency, set to 1.5. t denotes the position of the current day of year (DOY) within the year, expressed as a decimal between 0 and 1.

3. Phenological Features

Phenological features directly reflect the timing of crop developmental stages from sowing to maturity and serve as important remote sensing parameters for representing crop growth processes. Different crops exhibit distinct spectral variation timings due to differences in planting dates and growth cycles, making phenological features effective classification criteria. The phenological features extracted include: (i) cumulative growing degree days for the start (SOS), end (EOS), and length (LOS) of the growing season; (ii) EVI values at SOS and EOS; (iii) cumulative EVI during the growing season as a proxy for accumulated biomass. These are extracted using a threshold method: the EVI time series is sorted and the median value is selected as the threshold. The first date when EVI exceeds this threshold marks SOS, the last date marks EOS, and LOS is the difference between the two. This approach suits single-cropping systems (e.g., soybean, maize, rice), where EVI typically shows a clear increase and decline pattern.

4. SAR Features

For the four SAR feature parameters, we leveraged the SAR image time series to extract key crop characteristics, including statistical and principal component features. The statistical features are consistent with those used for optical data, covering maximum, minimum, variance, and the 15th, 50th, and 90th percentiles of five SAR parameters. These statistics help convey the average level and temporal variation within crop-specific time series curves. Additionally, principal component analysis (PCA) was performed on the Sentinel-1 time series in the temporal domain, with the first three principal components selected as the SAR principal component features.

2.4.3 Random forest

A Random Forest (RF) classifier was employed to identify soybean cultivation. As a non-parametric machine learning algorithm, RF offers strong fault tolerance and has been widely adopted in crop classification and mapping studies due to its robustness, accuracy, and efficiency. The algorithm is also well supported by the Google Earth Engine (GEE) platform, making it convenient for large-scale implementation. In this study, the RF model was configured using two key parameters: (1) numberOfTrees, which defines the number of decision trees in the ensemble. A larger value generally improves classification accuracy but increases computation time linearly; this parameter was set to 100. (2) minLeafPopulation, which specifies the minimum number of samples required at each leaf node. To reduce the risk of overfitting, it was set to 10. In addition to classification, the RF algorithm provides a measure of feature importance, enabling effective feature ranking and selection. Feature importance was assessed using the Mean Decrease in Impurity (MDI), which quantifies the reduction in node impurity contributed by each feature across all decision trees. Features contributing more to impurity reduction are considered more informative. Classification performance was evaluated using an independent test dataset. Standard accuracy metrics were reported, including overall accuracy (OA), producer’s accuracy (PA), user’s accuracy (UA), the Kappa coefficient, F1 score, and area-based accuracy.

$$OA = \frac{N_s + N_o}{n}$$

$$PA = \frac{N_s}{C_s} \quad UA = \frac{N_s}{T_s}$$

$$F1 \text{ score} = \frac{UA \times PA}{UA + PA} \times 2$$

$$Area \text{ Accuracy} = 1 - \frac{|Area_{gt} - Area_{rs}|}{Area_{gt}}$$

TABLE 3 Summary of classification features.

Feature type	Feature name	Processing method	Quantity
Vegetation Indices Time Series	EVI, LSWI, RENDVI, REPI, NDSVI	min, max, std, and 15/50/90th percentile	5×6
Visible and Red-Edge Band Time Series	B2, B3, B4, B6 (490nm – 740nm)		4×6
Shortwave Infrared Band Time Series	B11, B12 (1610nm – 2200nm)		2×6
Phenological Features	SOS EOS LOS	Median method	3×1
EVI Time Series Features	(EVI) Phase and Amplitude	Harmonic fitting	1×2
Accumulated Biomass Features	EVI	Accumulation	1×1
Statistical Features	σ_{VH}^0 , σ_{VV}^0 , CPR, RVI	max, min, mean, stdv, 15/50/90th percentile	7×4
Principal Component Features	σ_{VH}^0 , σ_{VV}^0 , CPR, RVI	Principal Component Analysis	3×4

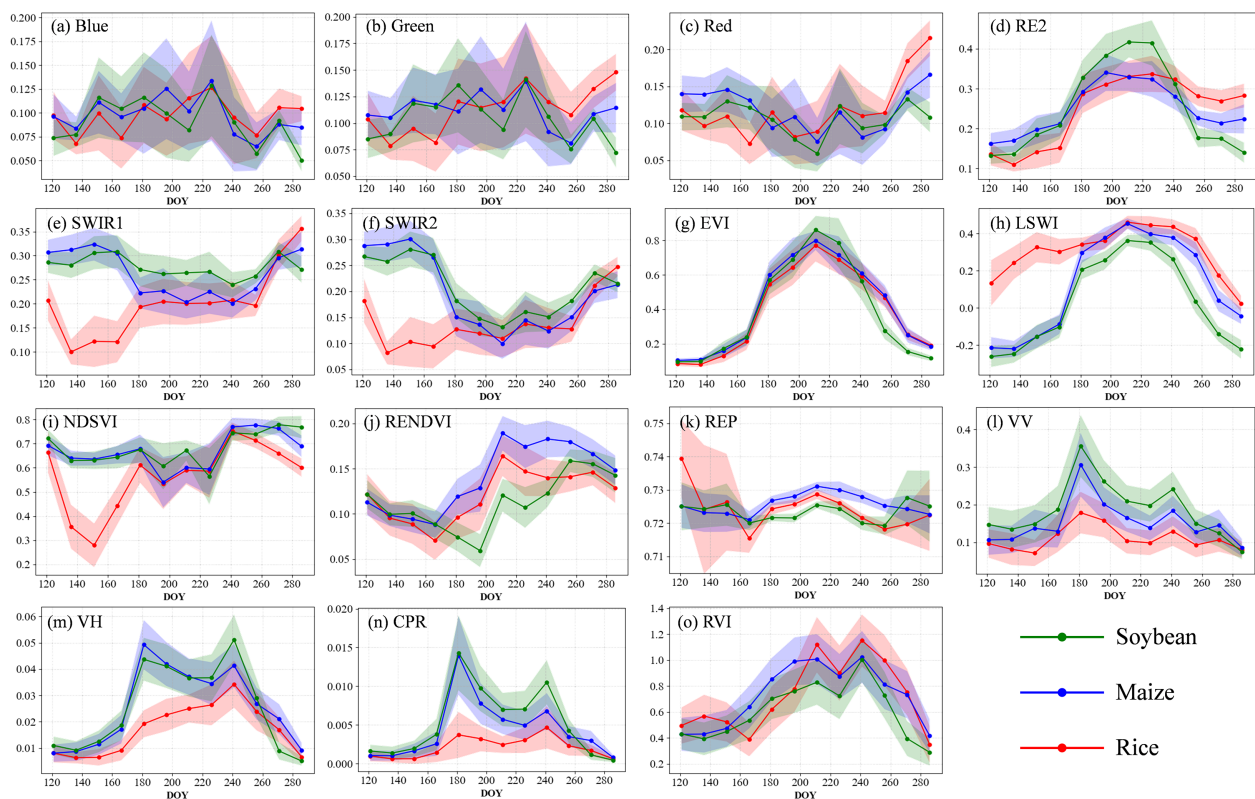


FIGURE 3

Time series curves of spectral reflectance (a–f), vegetation indices (g–k), and radar-based parameters (l–o) for soybean and other major crops. The curves are generated from all training samples within the study area. The x-axis represents the day of year (DOY), while the y-axis shows the values of spectral reflectance and vegetation indices. Solid lines indicate the mean values, and shaded areas represent one standard deviation.

Where N_S is the number of correctly classified soybean samples; C_S is the total number of samples classified as soybean; T_S is the total number of soybean validation samples; N_O is the number of correctly classified non-soybean samples; n is the total number of all validation samples; $Area_{gt}$ represents the actual soybean planting area obtained from statistical yearbook data, while $Area_{rs}$ denotes the soybean planting area estimated through remote sensing classification.

3 Results and discussion

3.1 Spectral reflectance and backscattering characteristics of crops

Figure 3 illustrates the time series curves of key spectral bands, vegetation indices, and SAR backscatter parameters used for crop classification, highlighting the seasonal dynamics of soybean, maize, and rice throughout the growing period. In the visible bands (Blue, Green, and Red), high standard deviations and substantial overlap among the three crops limit their discriminative power. Specifically, soybean and maize maintain relatively stable reflectance before DOY 220, followed by a rapid decline reaching a minimum around DOY 260. In contrast, rice exhibits a distinct pattern characterized by an initial decrease, a subsequent rise, and a sharp drop. Red band

reflectance shows an inverted U-shaped trajectory, consistent across all crops, reflecting chlorophyll dynamics during the growth cycle. In the red-edge band (RE2), reflectance increases steadily during early growth and peaks around DOY 220. Soybean displays the highest RE2 values prior to this point but declines rapidly thereafter due to senescence, eventually falling below those of maize and rice. The SWIR1 and SWIR2 bands, which are sensitive to water content, show significantly lower reflectance for rice during the inundation phase (DOY 120–200), distinguishing it from upland crops. Between DOY 200 and 260, soybean reflectance in these bands remains consistently higher than maize, offering some potential for discrimination. For vegetation indices, both EVI and NDVI exhibit patterns similar to RE2, capturing the physiological transition from rapid vegetative growth to senescence. However, their time series also show substantial overlaps across crops. LSWI, which reflects canopy and surface water content, remains consistently high and stable for rice throughout the season, providing a reliable indicator for its identification. NDSVI shows a pronounced rise-then-fall trend in rice, with greater temporal variability than in soybean and maize, effectively capturing the phenological rhythm of rice development. REPI, which correlates with cumulative biomass, is higher in maize during the late season, indicating a red-edge shift associated with increased chlorophyll content. RENDVI values are notably higher for maize than for soybean and rice between DOY

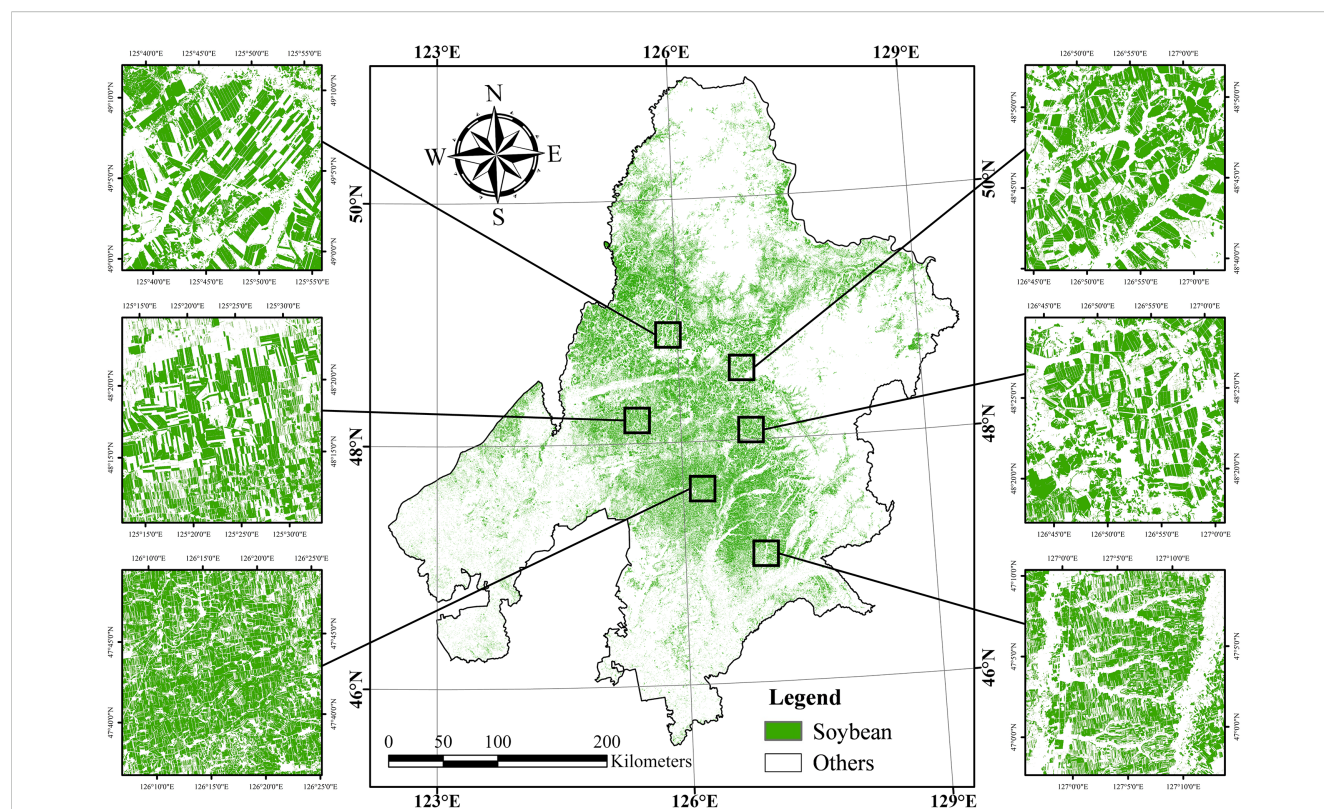


FIGURE 4
Identification and mapping of soybean planting areas in Jiusan Reclamation Area.

200 and 260, offering improved separability, particularly between soybean and maize.

SAR-derived features (VV and VH backscatter) are influenced by both vegetation structure and moisture content and follow a general rise–fall pattern across all crops. Between DOY 120 and 140, rice exhibits a sharp decline in backscatter due to specular reflection from water surfaces, maintaining low values throughout the season, which distinctly separates it from upland crops. In the VV channel, soybean and maize backscatter gradually increase from DOY 120 to 200 as canopy moisture accumulates, then stabilize before declining sharply after DOY 260. In the VH channel, soybean shows consistently higher backscatter than maize between DOY 200 and 260, indicating a more complex canopy structure. The Radar Vegetation Index (RVI) peaks around DOY 200, with rice maintaining significantly lower values than soybean and maize, enabling strong crop discrimination. Soybean's RVI also exhibits less temporal variability than maize, reflecting greater growth stability. The Cross-Polarization Ratio (CPR) further enhances classification accuracy; during the peak growth phase (DOY 220–240), soybean displays higher CPR values than both maize and rice, making it one of the most informative SAR-derived features for distinguishing soybean.

3.2 Soybean classification accuracy based on different data sources

Using multi-source remote sensing data, crops in the study area were classified, and soybean planting areas were extracted. The

classification achieved high accuracy, with an overall accuracy (OA) of 96.85%, a Kappa coefficient of 0.9493, and soybean-specific metrics of 98.30% for producer accuracy (PA), 93.51% for user accuracy (UA), 95.84% for F1-score, and 96.57% for area-based accuracy. The spatial distribution of soybean cultivation is illustrated in Figure 4.

Soybean planting was primarily concentrated between 46°N and 48°N, within a hilly transitional zone between the southern edge of the Lesser Khingan Mountains and the Songnen Plain—an area characterized by favorable agroecological conditions. The total soybean area in the study region reached 2.62 million hectares. Among the administrative divisions, Heihe recorded the largest planting area, with approximately 1.27 million hectares (48.33% of the total), mainly distributed across Nenjiang, Wudalianchi, and Beian, with smaller areas in the northern county of Sunwu. Qiqihar ranked second, with 0.83 million hectares (31.84%), primarily concentrated in eastern and northern counties including Nehe, Keshan, Kedong, Baiquan, and Yian, and with scattered fields in Gannan County in the west. Suihua accounted for 0.52 million hectares (19.83%), mainly located in the northern counties of Hailun, Suiling, and Wangkui. Spatially, soybean cultivation exhibited a relatively clustered distribution, though the field parcels were generally fragmented, small, and discontinuous. This pattern is influenced by natural topography, river networks, and the predominance of smallholder farming. In northern Qiqihar, soybean is often intercropped with maize, and fields tend to be narrow and elongated (approximately 60m × 1200m), with closely

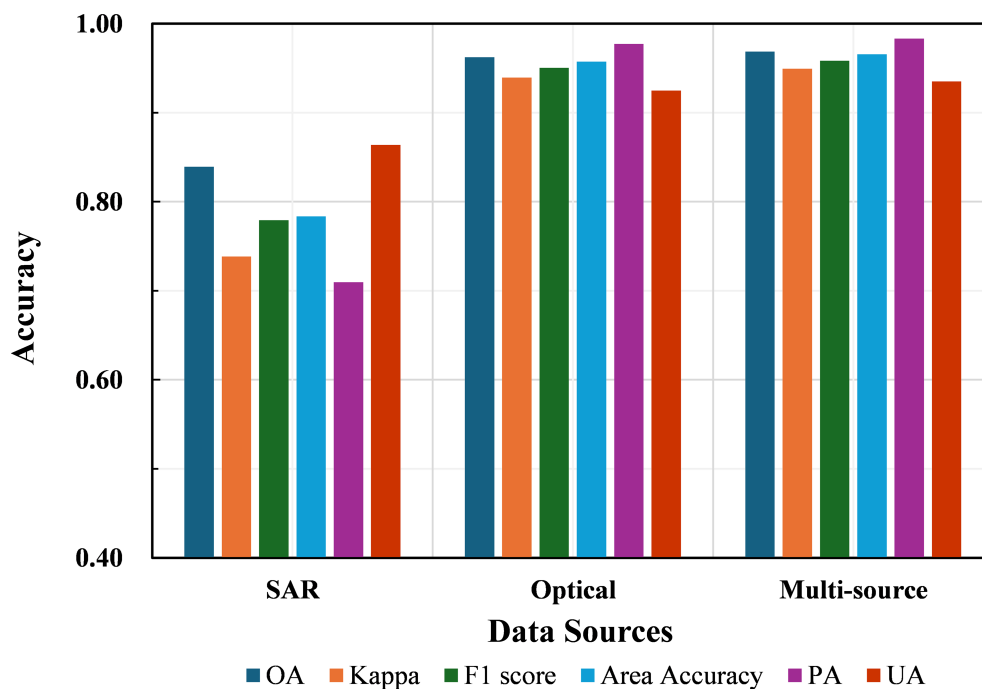


FIGURE 5
Accuracy results of different data sources.

spaced rows. In contrast, northwestern Qiqihar and southwestern Heihe feature more regular and contiguous fields, typically rectangular and reaching up to $800\text{m} \times 2400\text{m}$. In the mountainous areas of central and northern Heihe, irregular field shapes are common due to the complex terrain. At the junction of Heihe, Qiqihar, and Suihua, a state-operated farming zone and the core area of intensive soybean cultivation. Fields in this region are generally square-shaped (approximately $400\text{m} \times 400\text{m}$), clearly delineated by internal road networks, with minimal intercropping. The planting system here is more large-scale and centralized, reflecting a more modern and industrialized agricultural structure.

In terms of classification performance using individual data sources (Figure 5), the overall accuracy (OA) and Kappa coefficient based solely on SAR data reached 83.92% and 0.7384, respectively. Although these values are lower than those achieved using optical data, SAR still exhibits a certain level of capability in crop discrimination. In contrast, the optical-only approach delivered significantly higher performance, with OA and Kappa values of 96.23% and 0.9394, respectively—second only to the multi-source fusion method—highlighting its strong ability to distinguish between crop types. When integrating both optical and SAR data, the multi-source fusion approach achieved the highest accuracy, with an OA of 96.85% and a Kappa coefficient of 0.9494. Compared with the optical-only approach, this represents an improvement of 0.62% in OA and 0.0099 in Kappa; compared with the SAR-only method, the improvement is more substantial—14.93% in OA and 0.2010 in Kappa. These results suggest that multi-source fusion can further enhance classification performance, even when starting

from a high baseline, though the incremental gains over optical data alone are relatively modest.

From the perspective of soybean identification (Figure 5), the three data source strategies consistently followed the performance hierarchy: multi-source > optical > SAR. Evaluation metrics reveal that the SAR-only approach yielded notably higher user accuracy (UA) than producer accuracy (PA) for soybean, indicating a tendency toward underestimation, with the identified soybean area totaling only 1.985 million hectares. In contrast, both the optical and multi-source approaches exhibited higher PA than UA, suggesting more balanced classification outcomes. The F1 scores for soybean identification were 77.91% for SAR, 95.03% for optical, and 95.84% for the multi-source approach. Area-based accuracy followed a similar trend: 78.37% (SAR), 95.74% (optical), and 96.57% (multi-source). Compared to optical data, the multi-source fusion method achieved improvements of 0.81% in F1 score and 0.83% in area accuracy; when compared to SAR, these improvements rose to 17.93% and 18.20%, respectively. Overall, while multi-source data only slightly outperforms optical data, it substantially enhances performance over SAR, underscoring its value in robust and precise soybean identification.

Figure 6 illustrates crop classification results for four selected local scenes, comparing outputs derived from optical data alone, SAR data alone, and a fusion of both data sources. Each classification map is accompanied by a corresponding reference map and high-resolution Google Earth imagery for validation. The optical-only classification already demonstrates strong performance, with clearly defined crop boundaries and complete field delineation. Even in regions where soybean and maize are

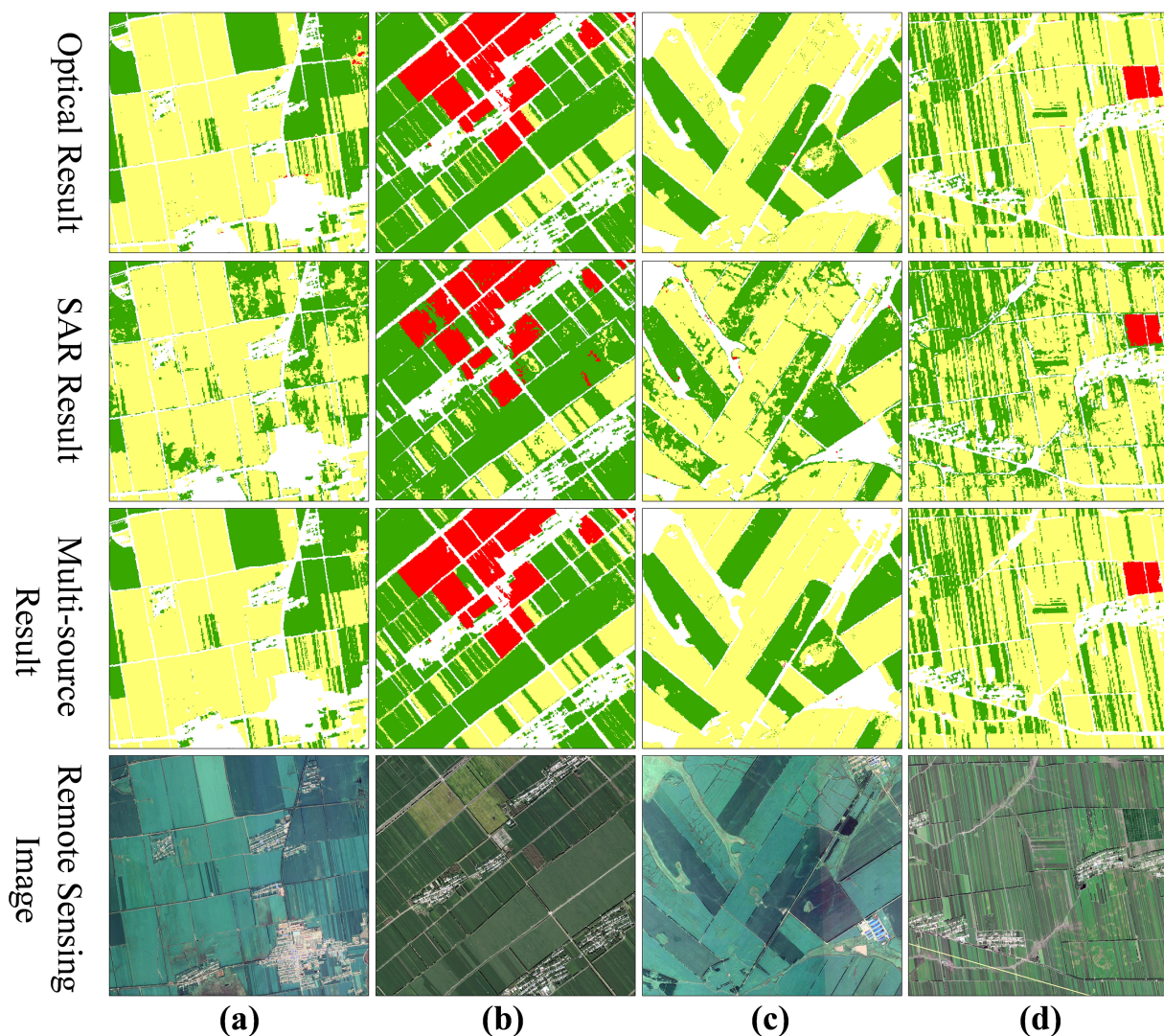


FIGURE 6

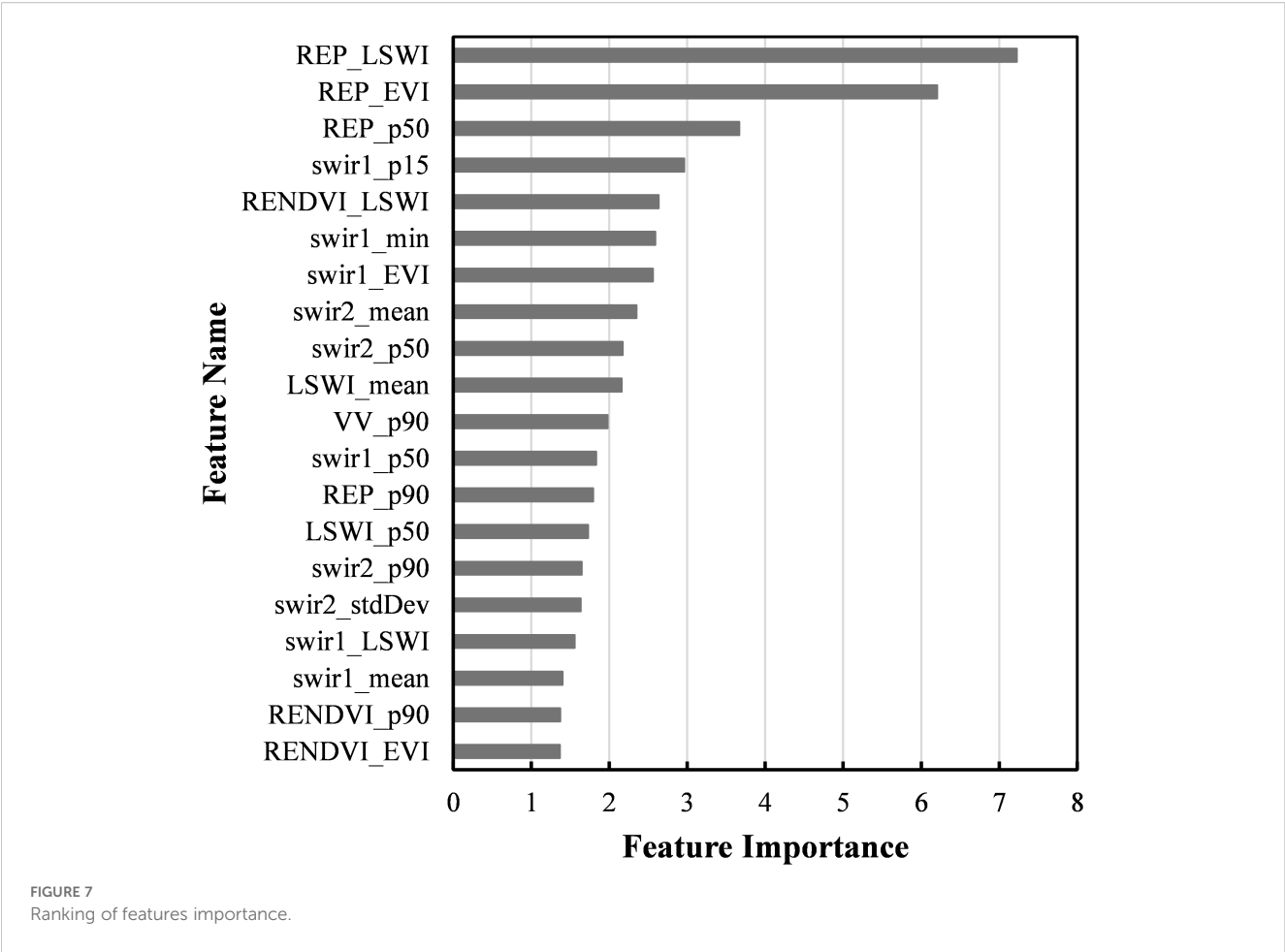
Comparison of local classification results under different data sources. (a–d) show four representative local scenes selected from different parts of the Jiusan Reclamation Area. For each scene, four maps are presented: the optical-only classification, the SAR-only classification, the multi-source classification, and the corresponding reference map derived from high-resolution Google Earth imagery. In the classification maps, soybean, maize, and rice are shown in green, yellow, and red, respectively.

intercropped, the model successfully achieves fine-scale differentiation between the two crops. In contrast, the SAR-only classification delivers more limited results, capturing only the general spatial patterns of crop distribution, but exhibiting considerable speckle noise and frequent misclassifications. The highest classification accuracy is attained using the multi-source fusion approach. Although the visual outcomes are broadly similar to those from the optical-only method, a detailed comparison reveals that the optical results still contain localized misclassification patches. These are significantly reduced in the multi-source results, leading to more precise delineation of crop field boundaries that align more closely with high-resolution reference imagery.

Feature selection was conducted using the Mean Decrease in Impurity (MDI) metric from the Random Forest classifier to assess the importance of each variable. Based on their importance scores, all

features were ranked, and a subset comprising the top-ranked 20 features was identified as optimal (Figure 7). Table 4 compares the classification performance of the full feature set and the selected subset. The results indicate that the subset yields classification accuracy largely comparable to that of the full set. The most notable difference is a roughly 1% decrease in user accuracy (UA) for soybean in the reduced feature set, resulting in a more pronounced overestimation of soybean area and a corresponding 3.4% decline in area-based accuracy. Minor reductions were also observed in overall accuracy (OA), Kappa coefficient, and soybean F1 score, each by approximately 0.3%. Notably, this reduction in feature dimensionality achieved an 87% decrease in data volume. Therefore, the selected 20-feature subset offers a favorable trade-off, maintaining high classification accuracy while significantly improving computational efficiency.

The suffixes in the feature names denote specific types of information: “_EVI” and “_LSWI” represent feature values



extracted during the respective peak periods of the EVI and LSWI indices; “_p15,” “_p50,” and “_p90” correspond to the 15th, 50th (median), and 90th percentiles; while “_min,” “_max,” “_mean,” and “_stdDev” denote common statistical measures, including minimum, maximum, mean, and standard deviation. The feature importance ranking indicates that REPI-related features occupy the highest positions, particularly those derived from the wettest and greenest composites, as well as median values. This highlights their critical role in crop classification, especially for soybean identification, and aligns well with the results from temporal curve analysis. Features from the shortwave infrared bands (SWIR1 and SWIR2) also demonstrate strong discriminative power, comprising 10 of the top 20 features. Additionally, LSWI and RENDVI features appear frequently, contributing to the effective identification of rice and to the differentiation between soybean and maize, respectively. Among SAR-derived features, only the 90th percentile of VV polarization is included in the optimal subset, suggesting the dominance of optical features within the

multi-source dataset. This finding is consistent with previous studies reporting that Sentinel-1 and Sentinel-2 data contribute approximately 27% and 73%, respectively, to overall crop classification accuracy. Moreover, the frequent presence of growth peak features and 90th percentile values reflects that crops exhibit their greatest spectral separability during peak growth stages.

3.3 Changes in soybean recognition accuracy with crop growth and temporal progression

To examine how soybean recognition accuracy varies throughout different stages of crop development, this study simulates classification performance over time by defining a series of image sequences with progressively extended temporal coverage. As illustrated in [Figure 8](#), the start date is fixed at May 1st, representing the early sowing period, while the end date is

TABLE 4 Accuracy of full feature set and selected feature subset.

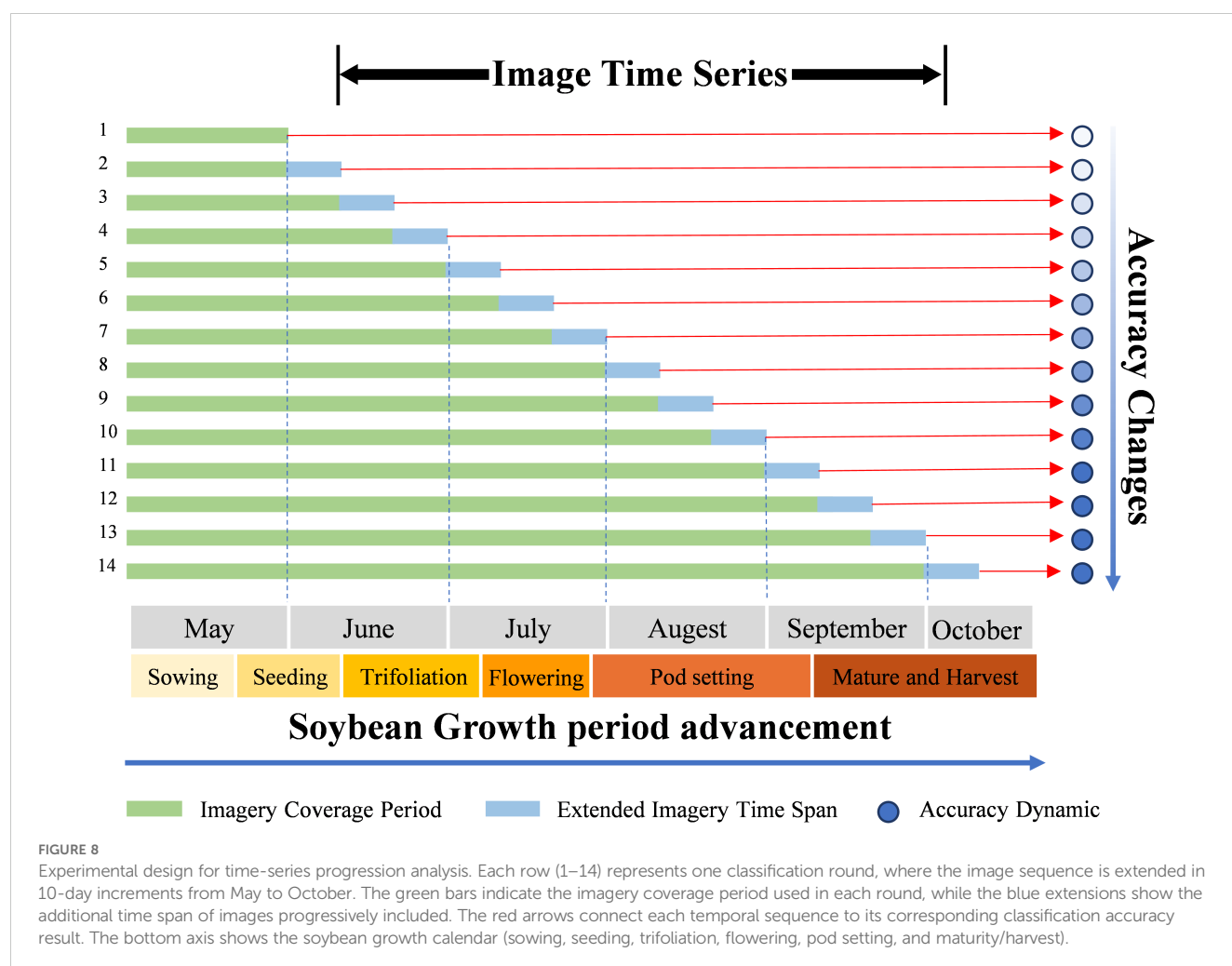
Feature set	OA	Kappa	Soybean PA	Soybean UA	Soybean F1 score	Soybean area accuracy
Full Feature Set	96.85%	0.9493	98.30%	93.51%	95.84%	96.57%
Selected Feature Set	96.63%	0.9460	98.86%	92.55%	95.60%	93.17%

incrementally extended in 10-day intervals until October 8th, corresponding to crop maturity and harvest. This results in a total of 14 experimental rounds. In each round, Sentinel-1 and Sentinel-2 imagery within the specified timeframe is used to construct a time series, which is then analyzed using the proposed classification method to identify soybean cultivation and assess recognition accuracy. Due to limited data availability in early stages, the first experimental round covers the period from May 1st to May 31st. As the season progresses, the image sequence becomes increasingly complete, leading to gradual improvements in recognition accuracy. At a certain point, accuracy reaches a practical threshold that allows for early-season mapping. To determine this threshold—referred to as the earliest identifiable time (EIT)—this study adopts the criterion proposed by You and Dong (2020), whereby the F1 score for soybean classification first exceeds 0.9. Moreover, this experiment compares the temporal evolution of classification accuracy across three data input strategies—optical only, SAR only, and multi-source fusion—highlighting the advantages of multi-source integration at various phenological stages. This dynamic assessment of accuracy under both crop growth and time-series progression underscores the value

of multi-source data in improving both the precision and timeliness of soybean identification.

As shown in Figure 9, soybean recognition accuracy is closely linked to the crop's phenological development. As growth progresses, both overall accuracy (OA) and the F1 score steadily increase. When segmented by phenological stages, average classification performance improves accordingly: during the sowing period (DOY ≤ 150), OA reaches 78.27% and F1 score is 74.94%; during the three-leaf stage (DOY 160–180), OA rises to 82.07%, and F1 to 78.22%; in the flowering period (DOY 190–210), OA improves to 89.23%, and F1 to 86.32%; during the pod-setting stage (DOY 220–240), OA climbs to 93.80%, and F1 reaches 92.09%; and by the grain-filling and maturity stage (DOY 250–280), OA reaches 94.72%, with an F1 score of 93.23%. Accuracy gains become less pronounced in the later stages.

The most substantial improvements in accuracy occur during the flowering and pod-setting stages. These phases correspond to vigorous vegetative growth, canopy closure, and enhanced physiological differentiation, making crop-specific spectral and structural features more detectable. This pattern aligns with the feature importance analysis, which indicates a higher concentration



of influential features during these periods. In early stages—such as sowing and the three-leaf phase—soybean plants remain short with sparse canopy cover, and imagery primarily captures bare soil, resulting in lower recognition accuracy. By the three-leaf stage, young stems emerge but the canopy remains underdeveloped, producing modest accuracy gains (OA and F1 increase by approximately 3.80% and 3.28%, respectively). The flowering period marks a critical point for yield formation, as branching intensifies and canopy cover becomes complete. Both optical and SAR features are more pronounced, leading to sharper accuracy gains: OA and F1 improve by 7.16% and 8.10% over the previous stage. During pod-setting, continued physiological development and pod enlargement further enhance separability, yielding an additional OA increase of 4.57% and F1 increase of 5.77%. In the grain-filling and maturity stages, soybean senescence leads to spectral stabilization, resulting in only slight accuracy improvements—0.92% in OA and 1.14% in F1.

When comparing classification performance across different data sources throughout crop growth, multi-source fusion consistently achieves the highest accuracy, demonstrating superior robustness and adaptability. Accuracy improves gradually in the early season and increases sharply after flowering, plateauing around DOY 230 (mid-pod-setting). Optical data performs similarly to multi-source data across most stages and remains the primary source for remote sensing classification. However, SAR-only data, despite generally lower accuracy, performs particularly well during critical stages—especially under conditions of poor optical image quality due to frequent cloud cover. Notably, around DOY 190, SAR-based accuracy temporarily surpasses that of optical data, highlighting its potential for independent classification in cloudy conditions. Figure 9a further quantifies SAR's complementary role. During the sowing and three-leaf stages, SAR contributes modest OA improvements (~2–3%) due to limited optical information. As soybeans enter flowering, SAR's added value increases significantly, peaking at a 6.69% OA improvement on DOY 190. This contribution declines during pod-setting and grain-filling stages, as optical features become more distinctive and the marginal benefit of SAR diminishes. Analysis of the F1 score trends in Figure 9b reveals consistent patterns across data sources: slow improvement in early stages, rapid gains from flowering to pod-setting, and eventual stabilization. Multi-source fusion consistently yields the highest F1 score, followed by optical, with SAR alone performing lowest. Around DOY 200 (mid-flowering stage), F1 scores for optical and multi-source data surge as soybean canopy and pod characteristics become prominent. Compared to the previous round, the F1 score for optical data increases by 12.88%, and for multi-source data by 7.59%. SAR's F1 score reaches 77.45% on DOY 190 and gradually peaks at 84.99% by DOY 280. These trends reinforce the supplementary value of SAR. From sowing to the three-leaf stage, SAR contributes approximately 3% to the F1 score. During the flowering stage, this contribution increases to a peak of 6.96%, before declining in the later stages as optical information becomes dominant.

The EIT for soybean using multi-source data is DOY 210 (early pod-setting stage), when canopy structures become clearly distinguishable and spectral features are well defined. At this point, the F1 scores for multi-source, optical, and SAR data are 91.80%, 87.57%, and 78.24%, respectively. Compared to the optical-only strategy, multi-source fusion advances the EIT by approximately 20 days, with a 4.23% improvement in F1 score. In contrast, optical data reaches an F1 score of 90% at DOY 230 (late pod-setting to early maturity), while SAR data never achieves the EIT threshold throughout the season.

In summary, incorporating SAR data significantly advances the EIT by approximately 20 days, enhancing the timeliness of soybean mapping. This earlier detection capability provides valuable lead time for downstream applications such as yield forecasting, pest and disease monitoring, seasonal planning, and agricultural decision-making.

3.4 Socio-economic implications of early and accurate crop identification

Early and accurate crop identification provides significant economic and social benefits across multiple scales. At the national and regional levels, timely information on crop distribution and growth conditions enhances yield forecasting and monitoring, offering governments and markets more reliable production and supply expectations to support food security decisions and emergency responses such as reserve management and relief distribution (Hoefsloot et al., 2012; Setiyono et al., 2014). At the farm level, detailed crop maps enable more efficient allocation of inputs such as fertilizers, seeds, and irrigation, while guiding key management decisions on sowing, irrigation, and pest control, thereby improving resource-use efficiency, reducing yield losses, and stabilizing farmers' incomes (Wu et al., 2023). Beyond farm management, remote sensing-derived indicators of crop cover and dynamics are increasingly applied in agricultural insurance, where they function as claim triggers and inputs for premium pricing, improving both the efficiency and sustainability of insurance systems. When combined with crop modeling, these data further strengthen the transparency and credibility of multi-risk assessments (Basso et al., 2013; Liu et al., 2025). From a broader macroeconomic perspective, reliable early assessments help reduce market uncertainty, stabilize supply chains, and inform price forecasting. Such high-quality information has been shown to mitigate global market shocks (Tanaka et al., 2023) and strengthens national food security strategies by providing spatially and temporally refined data for timely and targeted policy responses (Zhang et al., 2025).

4 Conclusion

This study addresses the persistent challenges of spectral confusion and data incompleteness in soybean remote sensing

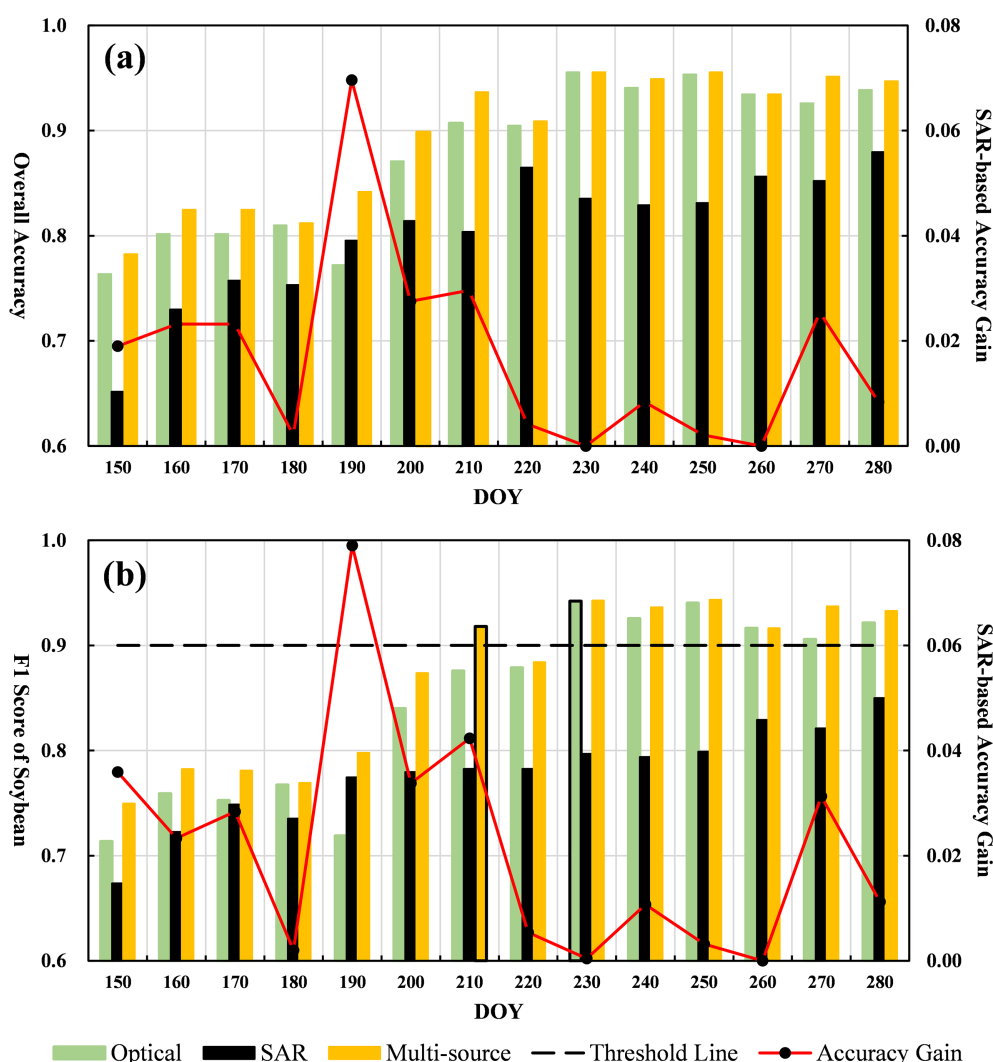


FIGURE 9

Temporal progression results: (a) Bar chart of OA for three data sources with red line indicating SAR contribution; (b) Bar chart of soybean F1 scores with red line for SAR contribution and black dashed line representing the F1 = 0.9 threshold.

classification by proposing a multi-source approach that integrates Sentinel-1 SAR and Sentinel-2 optical time-series imagery. Through systematic evaluation, we demonstrate that multi-source data fusion significantly enhances both the accuracy and timeliness of soybean identification. The integration of optical and SAR data yielded the highest classification performance, achieving an overall accuracy (OA) of 96.85%, a Kappa coefficient of 0.9493, a soybean F1-score of 95.84%, and an area accuracy of 96.57%. While the improvement over optical data alone was relatively modest—raising the F1-score and area accuracy by 0.81% and 0.83%, respectively—the improvement over SAR-only classification was substantial. This highlights the strong baseline performance of optical data and the complementary role of SAR, particularly under cloud-contaminated or spectrally ambiguous conditions, where SAR data significantly contributes to classification robustness. Soybean classification accuracy improved progressively with crop development. During the sowing stage, weak spectral signals led to low recognition

accuracy. As soybeans entered the flowering and pod-setting stages, canopy closure and physiological changes enhanced both spectral and structural separability, resulting in rapid gains in classification accuracy. The F1-score increased from 78.22% in July to 92.09% in August. Notably, SAR data provided the greatest incremental value during the flowering and pod-setting stages, with accuracy gains of up to 6.96%. In contrast, during the grain-filling and maturity stages, the contribution of SAR diminished as optical signals became more distinctive and stable. Furthermore, multi-source data advanced the EIT for soybean to DOY 210, approximately 20 days earlier than when using optical data alone. This temporal advancement is crucial for early-season applications such as yield forecasting, pest and disease monitoring, and harvest scheduling.

In summary, the integration of multi-source time-series remote sensing and the targeted extraction of phenological and key-stage features constitute an effective strategy for high-precision and early

soybean identification. The methodology proposed in this study provides both theoretical insights and practical tools for large-scale agricultural monitoring.

Data availability statement

Publicly available datasets were analyzed in this study. This data can be found here: The Sentinel-1 and Sentinel-2 imagery used in this study is publicly available from the Copernicus Open Access Hub: <https://scihub.copernicus.eu/> Preprocessing and time-series analysis were performed via Google Earth Engine (<https://earthengine.google.com/>) using freely accessible Sentinel datasets. Repository: Copernicus Open Access Hub Accession number(s): Not applicable (data are accessed by geographic coordinates and date range).

Author contributions

HZ: Writing – original draft, Methodology, Formal Analysis, Visualization, Data curation, Conceptualization. DP: Validation, Project administration, Supervision, Methodology, Writing – review & editing, Conceptualization, Funding acquisition. CD: Supervision, Writing – review & editing, Conceptualization. ZL: Software, Methodology, Validation, Writing – review & editing. XZ: Visualization, Writing – review & editing, Software. LY: Supervision, Writing – review & editing. KS: Writing – review & editing, Resources. YaZ: Resources, Writing – review & editing. JH: Writing – review & editing. SZ: Writing – review & editing. YL: Writing – review & editing. SL: Writing – review & editing. YiZ: Writing – review & editing. HP: Writing – review & editing.

References

- Amani, M., Kakooei, M., Moghimi, A., Ghorbanian, A., Ranjgar, B., Mahdavi, S., et al. (2020). Application of google earth engine cloud computing platform, sentinel imagery, and neural networks for crop mapping in Canada. *Remote Sens.* 12, 3561. doi: 10.3390/rs12213561
- Asam, S., Gessner, U., Almengor González, R., Wenzl, M., Kriese, J., and Kuenzer, C. (2022). Mapping crop types of Germany by combining temporal statistical metrics of sentinel-1 and sentinel-2 time series with LPIS data. *Remote Sens.* 14, 2981. doi: 10.3390/rs14132981
- Basso, B., Cammarano, D., and Carfagna, E. (2013). *Review of crop yield forecasting methods and early warning systems*, (Rome, Italy: Food and Agriculture Organization of the United Nations (FAO)). Available online at: http://www.fao.org/fileadmin/templates/ess/documents/meetings_and_workshops/GS_SAC_2013/Improving_methods_for_crops_estimates/Crop_Yield_Forecasting_Methods_and_Early_Warning_Systems_Lit_review.pdf (Accessed October 07, 2025).
- Bian, Y., Li, L., and Jing, W. (2023). CACPU-Net: Channel attention U-net constrained by point features for crop type mapping. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.1030595
- Chang, J. G., Shoshany, M., and Oh, Y. (2018). Polarimetric radar vegetation index for biomass estimation in desert fringe ecosystems. *IEEE Trans. Geosci. Remote Sens.* 56, 7102–7108. doi: 10.1109/TGRS.2018.2848285
- Chen, H., Li, H., Liu, Z., Zhang, C., Zhang, S., and Atkinson, P. M. (2023). A novel Greenness and Water Content Composite Index (GWCCI) for soybean mapping from single remotely sensed multispectral images. *Remote Sens. Environ.* 295, 113679. doi: 10.1016/j.rse.2023.113679
- Defourny, P., Bontemps, S., Bellemans, N., Cara, C., Dedieu, G., Guzzonato, E., et al. (2019). Near real-time agriculture monitoring at national scale at parcel resolution: Performance assessment of the Sen2-Agri automated system in various cropping systems around the world. *Remote Sens. Environ.* 221, 551–568. doi: 10.1016/j.rse.2018.11.007
- Dong, J., Xiao, X., Menarguez, M. A., Zhang, G., Qin, Y., Thau, D., et al. (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. doi: 10.1016/j.rse.2016.02.016
- Gallo, I., Ranghetti, L., Landro, N., La Grassa, R., and Boschetti, M. (2023). In-season and dynamic crop mapping using 3D convolution neural networks and sentinel-2 time series. *ISPRS J. Photogrammetry Remote Sens.* 195, 335–352. doi: 10.1016/j.isprsjprs.2022.12.005
- Gao, H., Wang, C., Wang, G., Zhu, J., Tang, Y., Shen, P., et al. (2018). A crop classification method integrating GF-3 PolSAR and sentinel-2A optical data in the Dongting lake basin. *Sensors* 18, 3139. doi: 10.3390/s18093139
- Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., and Moore, R. (2017). Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sens. Environ.* 202, 18–27. doi: 10.1016/j.rse.2017.06.031
- Hoefsloot, P., Ines, A., Dam, J., Duveiller, G., Kayitakire, F., and Hansen, J. (2012). *Combining crop models and remote sensing for yield prediction - concepts, applications and challenges for heterogeneous, smallholder environments*. (Luxembourg: Publications Office of the European Union). Available online at: <https://data.europa.eu/doi/10.2788/72447>.

Funding

The author(s) declare financial support was received for the research and/or publication of this article. This work was supported by the National Natural Science Foundation of China (Grant No. 42471372), and the Science and Disruptive Technology Program, AIRCAS (Grant No. 2024-AIRCAS-SDPT-15).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declare that no Generative AI was used in the creation of this manuscript. Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

- Huang, Y., Qiu, B., Chen, C., Zhu, X., Wu, W., Jiang, F., et al. (2022). Automated soybean mapping based on canopy water content and chlorophyll content using Sentinel-2 images. *Int. J. Appl. Earth Observation Geoinformation* 109, 102801. doi: 10.1016/j.jag.2022.102801
- Huang, X., Vrieling, A., Dou, Y., Belgiu, M., and Nelson, A. (2024). A robust method for mapping soybean by phenological aligning of Sentinel-2 time series. *ISPRS J. Photogrammetry Remote Sens.* 218, 1–18. doi: 10.1016/j.isprsjprs.2024.10.015
- Immitzer, M., Vuolo, F., and Atzberger, C. (2016). First experience with sentinel-2 data for crop and tree species classifications in central Europe. *Remote Sens.* 8, 166. doi: 10.3390/rs8030166
- Inglada, J., Vincent, A., Arias, M., and Marais-Sicre, C. (2016). Improved early crop type identification by joint use of high temporal resolution SAR and optical image time series. *Remote Sens.* 8, 362. doi: 10.3390/rs8050362
- Jiao, X., McNairn, H., Yekkehkhany, B., Dingle Robertson, L., and Ihuoma, S. (2022). Integrating Sentinel-1 SAR and Sentinel-2 optical imagery with a crop structure dynamics model to track crop condition. *Int. J. Remote Sens.* 43, 6509–6537. doi: 10.1080/01431161.2022.2142077
- Johnson, D. M., and Mueller, R. (2021). Pre- and within-season crop type classification trained with archival land cover information. *Remote Sens. Environ.* 264, 112576. doi: 10.1016/j.rse.2021.112576
- Khabbazan, S., Vermunt, P., Steele-Dunne, S., Ratering Arntz, L., Marinetti, C., van der Valk, D., et al. (2019). Crop monitoring using sentinel-1 data: A case study from the Netherlands. *Remote Sens.* 11, 1887. doi: 10.3390/rs11161887
- Khan, W., Minallah, N., Sher, M., Khan, M. A., Rehman, A. U., Al-Ansari, T., et al. (2024). Advancing crop classification in smallholder agriculture: A multifaceted approach combining frequency-domain image co-registration, transformer-based parcel segmentation, and Bi-LSTM for crop classification. *PLoS One* 19, e0299350. doi: 10.1371/journal.pone.0299350
- Konduri, V. S., Kumar, J., Hargrove, W. W., Hoffman, F. M., and Ganguly, A. R. (2020). Mapping crops within the growing season across the United States. *Remote Sens. Environ.* 251, 112048. doi: 10.1016/j.rse.2020.112048
- Li, C., Chen, W., Wang, Y., Wang, Y., Ma, C., Li, Y., et al. (2022). Mapping winter wheat with optical and SAR images based on google earth engine in Henan province, China. *Remote Sens.* 14, 284. doi: 10.3390/rs14020284
- Li, H., Song, X.-P., Hansen, M. C., Becker-Reshef, I., Adusei, B., Pickering, J., et al. (2023). Development of a 10-m resolution maize and soybean map over China: Matching satellite-based crop classification with sample-based area estimation. *Remote Sens. Environ.* 294, 113623. doi: 10.1016/j.rse.2023.113623
- Li, X., Yu, L., Peng, D., and Gong, P. (2021). A large-scale, long time-series, (1984–2020) of soybean mapping with phenological features: Heilongjiang Province as a test case. *Int. J. Remote Sens.* 42, 7332–7356. doi: 10.1080/01431161.2021.1957177
- Liao, C., Wang, J., Xie, Q., Baz, A. A., Huang, X., Shang, J., et al. (2020). Synergistic use of multi-temporal RADARSAT-2 and VENUS data for crop classification based on 1D convolutional neural network. *Remote Sens.* 12, 832. doi: 10.3390/rs12050832
- Liu, Y., Ford Ramsey, A., and Zhou, Z. (2025). “Model and forecast combination for predictive yield distributions in crop insurance,” in *Quantitative Risk Management in Agricultural Business*. Eds. H. Assa, P. Liu and S. Wang (Springer Nature Switzerland, Cham), 197–216. doi: 10.1007/978-3-031-80574-5_8
- Liu, X., Rahman, T., Song, C., Su, B., Yang, F., Yong, T., et al. (2017). Changes in light environment, morphology, growth and yield of soybean in maize-soybean intercropping systems. *Field Crops Res.* 200, 38–46. doi: 10.1016/j.fcr.2016.10.003
- Luo, C., Liu, H., Lu, L., Liu, Z., Kong, F., and Zhang, X. (2021). Monthly composites from Sentinel-1 and Sentinel-2 images for regional major crop mapping with Google Earth Engine. *J. Integr. Agric.* 20, 1944–1957. doi: 10.1016/S2095-3119(20)63329-9
- Maleki, S., Baghdadi, N., Bazzi, H., Dantas, C. F., Ienco, D., Nasrallah, Y., et al. (2024). Machine learning-based summer crops mapping using sentinel-1 and sentinel-2 images. *Remote Sens.* 16, 4548. doi: 10.3390/rs16234548
- Mandal, D., Kumar, V., Ratha, D., Dey, S., Bhattacharya, A., Lopez-Sanchez, J. M., et al. (2020). Dual polarimetric radar vegetation index for crop growth monitoring using sentinel-1 SAR data. *Remote Sens. Environ.* 247, 111954. doi: 10.1016/j.rse.2020.111954
- Moumni, A., and Lahrouni, A. (2021). Machine learning-based classification for crop-type mapping using the fusion of high-resolution satellite imagery in a semiarid area. *Scientifica* 2021, 1–20. doi: 10.1155/2021/8810279
- Pageot, Y., Baup, F., Inglada, J., Baghdadi, N., and Demarez, V. (2020). Detection of irrigated and rainfed crops in temperate areas using sentinel-1 and sentinel-2 time series. *Remote Sens.* 12, 3044. doi: 10.3390/rs12183044
- Qu, Y., Zhao, W., Yuan, Z., and Chen, J. (2020). Crop mapping from sentinel-1 polarimetric time-series with a deep neural network. *Remote Sens.* 12, 2493. doi: 10.3390/rs12152493
- Rezaei, E. E., Webber, H., Asseng, S., Boote, K., Durand, J. L., Ewert, F., et al. (2023). Climate change impacts on crop yields. *Nat. Rev. Earth Environ.* 4, 831–846. doi: 10.1038/s43017-023-00491-0
- Roy, D. P., Wulder, M. A., Loveland, T. R., C.E., W., Allen, R. G., Anderson, M. C., et al. (2014). Landsat-8: Science and product vision for terrestrial global change research. *Remote Sens. Environ.* 145, 154–172. doi: 10.1016/j.rse.2014.02.001
- Schmutz, J., Cannon, S. B., Schlueter, J., Ma, J., Mitros, T., Nelson, W., et al. (2010). Genome sequence of the palaeopolyploid soybean. *Nature* 463, 178–183. doi: 10.1038/nature08670
- Setiyono, T., Nelson, A., and Holec, F. (2014). “Remote sensing based crop yield monitoring and forecasting,” in *Crop monitoring for improved food security*. (Rome, Italy: Food and Agriculture Organization of the United Nations (FAO)). Available online at: <https://openknowledge.fao.org/handle/20.500.14283/i4273e> (Accessed October 07, 2025).
- She, B., Hu, J., Huang, L., Zhu, M., and Yin, Q. (2024). Mapping soybean planting areas in regions with complex planting structures using machine learning models and Chinese GF-6 WFV data. *Agriculture* 14, 231. doi: 10.3390/agriculture14020231
- Shen, Y., Zhang, X., Tran, K. H., Ye, Y., Gao, S., Liu, Y., et al. (2025). Near real-time corn and soybean mapping at field-scale by blending crop phenometrics with growth magnitude from multiple temporal and spatial satellite observations. *Remote Sens. Environ.* 318, 114605. doi: 10.1016/j.rse.2025.114605
- Song, X.-P., Huang, W., Hansen, M. C., and Potapov, P. (2021). An evaluation of Landsat, Sentinel-2, Sentinel-1 and MODIS data for crop type mapping. *Sci. Remote Sens.* 3, 100018. doi: 10.1016/j.srs.2021.100018
- Sun, C., Bian, Y., Zhou, T., and Pan, J. (2019). Using of multi-source and multi-temporal remote sensing data improves crop-type mapping in the subtropical agriculture region. *Sensors* 19, 2401. doi: 10.3390/s19102401
- Tanaka, T., Sun, L., Becker-Reshef, I., Song, X.-P., and Puricelli, E. (2023). Satellite forecasting of crop harvest can trigger a cross-hemispheric production response and improve global food security. *Commun. Earth Environ.* 4, 334. doi: 10.1038/s43247-023-00992-2
- Veloso, A., Mermoz, S., Bouvet, A., Le Toan, T., Planells, M., Dejoux, J.-F., et al. (2017). Understanding the temporal behavior of crops using Sentinel-1 and Sentinel-2-like data for agricultural applications. *Remote Sens. Environ.* 199, 415–426. doi: 10.1016/j.rse.2017.07.015
- Wei, P., Ye, H., Qiao, S., Liu, R., Nie, C., Zhang, B., et al. (2023). Early crop mapping based on sentinel-2 time-series data and the random forest algorithm. *Remote Sens.* 15, 3212. doi: 10.3390/rs15133212
- Weiss, M., Jacob, F., and Duveiller, G. (2020). Remote sensing for agricultural applications: A meta-review. *Remote Sens. Environ.* 236, 111402. doi: 10.1016/j.rse.2019.111402
- Wu, B., Zhang, M., Zeng, H., Tian, F., Potgieter, A. B., Qin, X., et al. (2023). Challenges and opportunities in remote sensing-based crop monitoring: a review. *Natl. Sci. Rev.* 10, nwac290. doi: 10.1093/nsr/nwac290
- Xuan, F., Dong, Y., Li, J., Li, X., Su, W., Huang, X., et al. (2023). Mapping crop type in Northeast China during 2013–2021 using automatic sampling and tile-based image classification. *Int. J. Appl. Earth Observation Geoinformation* 117, 103178. doi: 10.1016/j.jag.2022.103178
- Yommy, A. S., Liu, R., and Wu, A. S. (2015). “SAR image despeckling using refined lee filter,” in *2015 7th International Conference on Intelligent Human-Machine Systems and Cybernetics* (IEEE, Hangzhou, China), 260–265. doi: 10.1109/IHMSC.2015.236
- You, N., and Dong, J. (2020). Examining earliest identifiable timing of crops using all available Sentinel 1/2 imagery and Google Earth Engine. *ISPRS J. Photogrammetry Remote Sens.* 161, 109–123. doi: 10.1016/j.isprsjprs.2020.01.001
- You, N., Dong, J., Huang, J., Du, G., Zhang, G., He, Y., et al. (2021). The 10-m crop type maps in Northeast China during 2017–2019. *Sci. Data* 8, 41. doi: 10.1038/s41597-021-00827-9
- You, N., Dong, J., Li, J., Huang, J., and Jin, Z. (2023). Rapid early-season maize mapping without crop labels. *Remote Sens. Environ.* 290, 113496. doi: 10.1016/j.rse.2023.113496
- Zhang, H., Lou, Z., Peng, D., Zhang, B., Luo, W., Huang, J., et al. (2024). Mapping annual 10-m soybean cropland with spatiotemporal sample migration. *Sci. Data* 11, 439. doi: 10.1038/s41597-024-03273-5
- Zhang, M., Xu, X., Ou, J., Zhang, Z., Chen, F., Shi, L., et al. (2025). Mapping 1-km soybean yield across China from 2001 to 2020 based on ensemble learning. *Sci. Data* 12, 408. doi: 10.1038/s41597-025-04738-x
- Zhou, X., Wang, J., Shan, B., and He, Y. (2024). Early-season crop classification based on local window attention transformer with time-series RCM and sentinel-1. *Remote Sens.* 16, 1376. doi: 10.3390/rs16081376
- Zhu, M., She, B., Huang, L., Zhang, D., Xu, H., and Yang, X. (2022). Identification of soybean based on Sentinel-1/2 SAR and MSI imagery under a complex planting structure. *Ecol. Inf.* 72, 101825. doi: 10.1016/j.ecoinf.2022.101825