International Journal of Digital Earth

ISSN: 1753-8947 (Print) 1753-8955 (Online) Journal homepage: www.tandfonline.com/journals/tjde20

Identifying long-term winter wheat planting areas using decision tree-derived ensemble learning models and multi-source data

Siqi Yang , Caiyun Deng , Tianhe Xu , Ran Kang , Huiying Yin , Jian Guo , Li Zhang , Lulu Si & Hermann Josef Kaufmann

To cite this article: Siqi Yang , Caiyun Deng , Tianhe Xu , Ran Kang , Huiying Yin , Jian Guo , Li Zhang , Lulu Si & Hermann Josef Kaufmann (2026) Identifying long-term winter wheat planting areas using decision tree-derived ensemble learning models and multi-source data, International Journal of Digital Earth, 19:1, 2620858, DOI: 10.1080/17538947.2026.2620858

To link to this article: https://doi.org/10.1080/17538947.2026.2620858

INTERNATIONAL JOURNAL OF DIGITAL EARTH
2026, VOL. 19, NO. 1, 2620858
https://doi.org/10.1080/17538947.2026.2620858

RESEARCH ARTICLE

Identifying long-term winter wheat planting areas using decision tree-derived ensemble learning models and multi-source data

Siqi Yang[a,b], Caiyun Deng[a,b], Tianhe Xu[a,b], Ran Kang[a,b], Huiying Yin[a,b], Jian Guo[a,b], Li Zhang[a,b], Lulu Si[c] and Hermann Josef Kaufmann[d]

[a]School of Space Science and Technology, Shandong University, Weihai, People's Republic of China; [b]Shandong Key Laboratory of Optical Astronomy and Solar-Terrestrial Environment, School of Space Science and Physics, Shandong University, Weihai, People's Republic of China; [c]Institute of Water and Environmental Engineering (IIAMA), Universitat Politècnica de València, Valencia, Spain; [d]Institute of Space Sciences, Shandong University, Shandong, People's Republic of China

ABSTRACT

To provide high-quality data support for agricultural policy-making and grain subsidies, this study presents an efficient method for extracting winter wheat planting areas using Landsat series satellite imagery and monthly maximum normalized difference vegetation index (NDVI) stacks. Three ensemble decision-tree algorithms—Random Forest, XGBoost, and CatBoost—were compared. The best model achieved 91% cross-validation accuracy, with a strong statistical validation accuracy in terms of its municipal-scale validation $R^2 = 0.91$ (MAE = 49,650 hm^2; RMSE = 64,440 hm^2) and county-scale $R^2 = 0.84$ (MAE = 7125 hm^2, RMSE = 9875 hm^2). A spatiotemporal analysis revealed a notable decline in winter wheat area in Shandong Province, which was concentrated in its western and southern regions. The cultivation centre shifted westwards, and then northwards, and the landscape patterns transitioned from large, aggregated patches to small, dispersed patches. This trend of decreasing intensification level for the winter wheat planting regions poses challenges for achieving water-saving and intensive land management goals. These trends are influenced by climate, topography, and socioeconomic factors, as well as agricultural policies. The proposed method offers robust support for attaining large-scale crop monitoring and sustainable agricultural management.

ARTICLE HISTORY
Received 21 July 2025
Accepted 11 January 2026

KEYWORDS
Ensemble learning; Google Earth Engine; multisource data; NDVI time series; winter wheat extraction

1. Introduction

Winter wheat is a critical staple food crop both in China and globally, underpinning national and global food security (Dong et al. 2024). The precise monitoring of its spatiotemporal distribution therefore serves as a scientific prerequisite for reliable yield forecasting and evidence-based food policy design, which in turn directly impacts food security and sustainable agricultural development (Tian et al. 2021). Consequently, high-accuracy mapping of winter wheat cultivation is vital for addressing pressing challenges in food security and agricultural resource allocation.

Such mapping fundamentally hinges on the accurate delineation of the winter wheat planting areas, which requires the integration of both high spatiotemporal resolution data and advanced inversion algorithms. Remote sensing technology offers an efficient and accurate means of acquiring spatial distribution data over large regions, providing critical support for crop monitoring and yield forecasting (Potapov et al. 2022; Sadeh et al. 2021). In particular, multitemporal remote sensing data, with continuous spectral signatures, have been widely applied to enhance the accuracy of large-area crop growth monitoring and early yield prediction (Blickensdörfer et al. 2022; Sagan et al. 2021). For instance, MODIS data, with their advantages of high temporal resolution, have been widely used (Feng et al. 2021; Luo et al. 2023; Ren et al. 2021). However, the lower spatial resolution of MODIS data poses challenges in maintaining interpretation accuracy. Conversely, high-spatial resolution remote sensing images have been increasingly

CONTACT Caiyun Deng Dengcaiyun@sdu.edu.cn School of Space Science and Technology, Shandong University, Weihai, 264209, People's Republic of China

Supplemental data for this article can be accessed online at https://doi.org/10.1080/17538947.2026.2620858.

Article highlights

- Ensemble models (RF, XGBoost, and CatBoost) accurately mapped winter wheat planting areas.
- An optimized sampling strategy effectively reduced the reliance on annual field data collection schemes.
- A decreasing intensification level was detected for winter wheat planting regions.
- The spatiotemporal dynamics and key factors influencing wheat cultivation were captured.

employed for winter wheat area extraction in previous studies. For example, Yuan et al. (2022) proposed an object-oriented approach combined with deep learning using Sentinel-2A satellite imagery to extract winter wheat areas. However, despite their advantages in delineating small and fragmented fields, the application of high-resolution images in large-scale and long term winter wheat monitoring is often constrained by its substantial computational burden, pronounced intra-class spectral variation, and inherently short data record. In this context, advanced inversion algorithms, particularly machine learning (ML) and deep learning (DL), are critical for extracting reliable information from complex and heterogeneous data sources. They are finding growing application in the mapping of large-scale winter wheat acreage due to their superior accuracy over traditional methods (Fan et al. 2020, 2024). Although DL demonstrates great potential for complex feature extraction, ML is often a more effective solution for large-scale crop mapping due to its lower demand for training data, greater computational efficiency, and higher model interpretability. Building on the advantages of ML, this study therefore focuses on developing an ensemble learning framework specifically designed to enhance temporal stability and generalisation for mapping winter wheat areas over a 25-year period.

Additionally, the distinctive spectral changes in plants during different growth periods are key to identifying crop distributions. Time-series vegetation indices derived from satellite data have been widely employed in winter wheat planting area extraction, as they effectively capture phenological signatures, thereby offering high accuracy and broad applicability (Liu et al. 2020). Among them, the multitemporal NDVI has proven particularly effective at identifying winter wheat and achieving superior extraction precision (Li and Lei 2021). It has been widely used to extract winter wheat planting areas at different spatial and temporal scales (Dong et al. 2020; Wang et al. 2022). These methodologies provide the basis for deriving high-precision crop distribution information. Nevertheless, it is still challenging to extract long-term, high-spatial-resolution, and high-precision winter wheat planting distribution areas, especially in regions with complex terrains and diverse planting structures.

In this study, we integrated multisource remote sensing data with three advanced machine learning techniques to develop an efficient and accurate method for extracting winter wheat planting areas. The main contributions are as follows. (1) We constructed a long-term time-series dataset by intergrating Landsat series data spanning multiple operational periods with MODIS satellite data. By plotting the maximum NDVI curve for both winter wheat and non-wheat sampling points across growth stages, we enhanced the sensitivity for winter wheat identification and improving segmentation accuracy. (2) We evaluated the performance of different machine learning methods and identified the optimal one for extracting winter wheat planting areas. The proposed method effectively improves precision and provides a scalable approach for regional-scale applications. (3) We produced a high-precision winter wheat planting area dataset for Shandong Province from 2000 to 2024 at a 30-metre spatial resolution. This dataset provides a valuable foundation for winter wheat health monitoring, yield forecasting, agricultural resource management, and disaster impact assessment. This study is of significant importance for safeguarding national food security and formulating science-driven agricultural policies.

2. Materials and methods

2.1. Study area

To better capture the phenological information of winter wheat and extract representative features, we selected Shandong Province as the study area, as it is one of the primary wheat-producing regions in China (Figure 1). Shandong is not only an economically significant province with a large population but also a major agricultural hub, ranking among the top producers of grain, vegetables, fruits, and other key

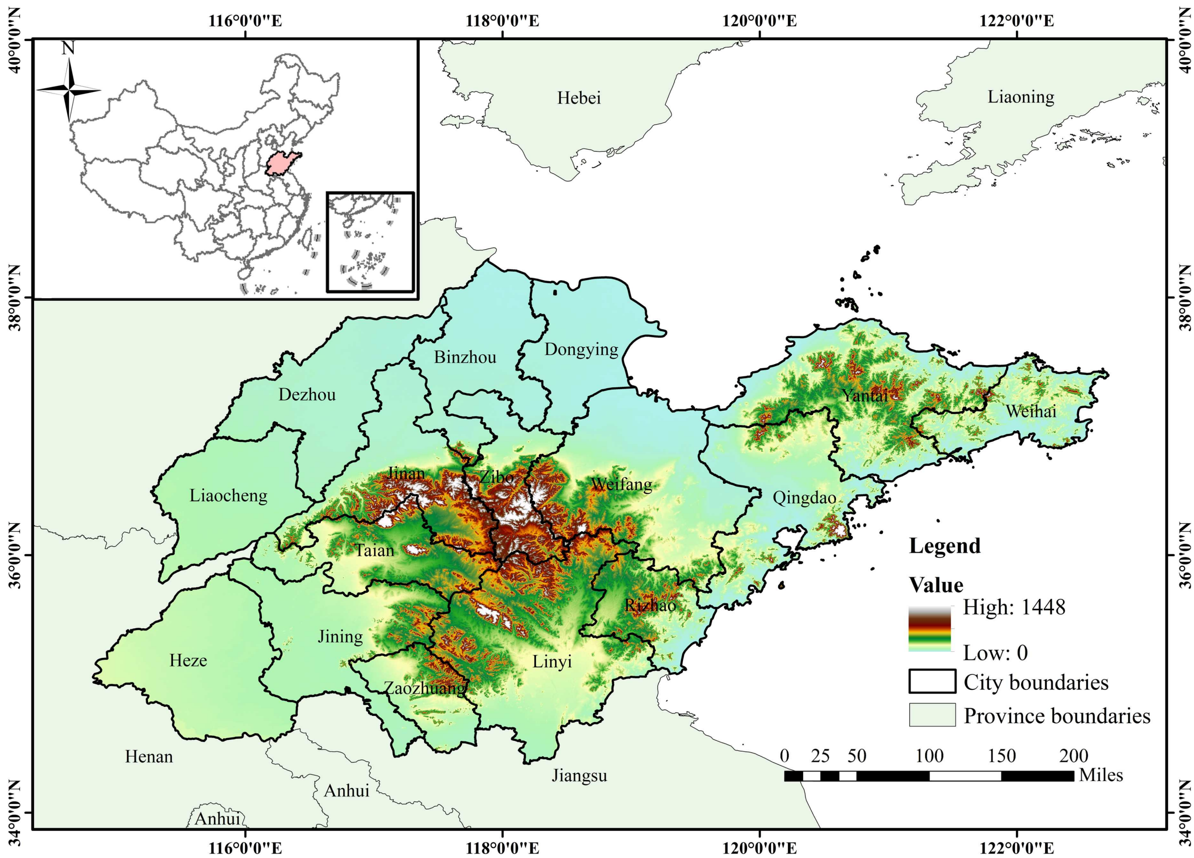

Figure 1. Geographical location of the study area.

agricultural products in the country. Wheat, in particular, is a major staple crop and food source in Shandong Province. In 2023, the wheat planting area in the province reached 40,088.53 km^2, accounting for 17.38% of China's total wheat acreage, ranking it as the second-largest wheat-producing province (http://www.moa.gov.cn). Shandong Province is located in the eastern coastal part of China, downstream of the Yellow River, and has a strategic position bordered by the Yellow Sea and the Bohai Sea. The province spans approximately 420 km from north to south and 700 km from east to west, covering a total area of 157,100 km^2. It consists of 16 cities and 137 counties. Its landforms consist primarily of hills and plains, making it well-suited for large-scale grain crop cultivation. The region experiences a warm temperate monsoon climate, with mild temperatures and distinct seasons (Hou et al. 2021; Zhang et al. 2021). Rainfall decreases from 850 mm along the southeast coast to 550 mm inland in the northwest, with nearly 80% of the precipitation occurring from June to September. Floods and droughts remain significant risks to crops. The primary cropping systems in the study area are winter wheat and summer corn, with winter wheat being the main crop harvested in the summer. Winter wheat is sown from September to October and harvested from late May to June of the following year (Wang, Shi, and Wen 2023). Due to the region's varied terrain and climatic conditions, the spatial distribution of winter wheat cultivation in Shandong exhibits significant heterogeneity (Feng et al. 2024; Wu et al. 2019). Given the widespread importance of wheat cultivation in the province, accurately identifying the planting area and distribution information of winter wheat is conducive to regional high-quality agricultural development.

2.2. Data source and processing

Three main types of data were used: multi-source remote sensing data (Landsat 5\7\8 and MODIS products), sampling data, and related auxiliary data. This study employs multi-source remote sensing data and machine learning techniques, generates seasonal maximum NDVI composite images and extracts

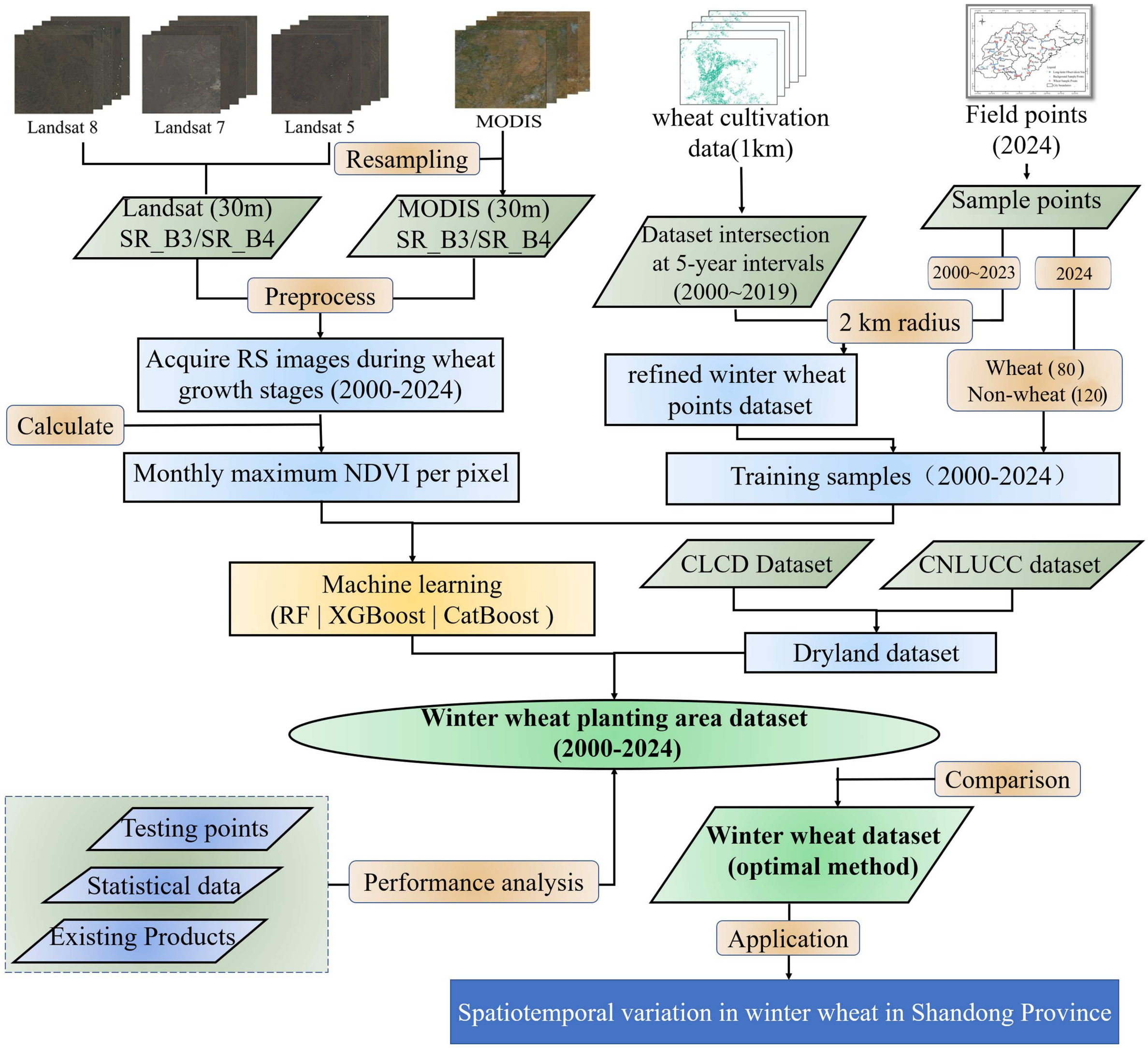

Figure 2. The technical flowchart for extracting winter wheat.

Table 1. Satellite parameters for the Landsat series and the Terra satellite.

Data type	Usage period	Sensor	Spatial resolution	Revisit cycle
Landsat 5	March 1, 1984 – June 5, 2013	Multispectral Scanner (MSS) and Thematic Mapper (TM)	Multispectral bands of TM sensor: 30 m, thermal infra-red bands: 120 m	Global coverage every 16 days
Landsat 7	April 15, 1999 – May 31, 2003	Enhanced Thematic Mapper Plus (ETM+)	Panchromatic band: 15 m, other bands: 30 m	Global coverage every 16 days
Landsat 8	February 11, 2013 - Present	Operational Land Imager (OLI) and Thermal Infra-red Sensor (TIRS)	OLI multispectral bands: 30 m, TIRS thermal infra-red bands: 100 m	Global coverage every 16 days
Terra	December 18, 1999 - Present	MODIS	250 m, 500 m, 1000 m	Global coverage every 1-2 days

the winter wheat planting area using the Google Earth Engine (GEE) platform. The details of the corresponding data sources, data processing, and methods are shown in Section 2.2 and Section 2.3. The flow chart applied in the study for extracting winter wheat is shown in Figure 2.

2.2.1. Multi-source remote sensing data

In this study, Landsat surface reflectance (SR) products were used for long-term spectral characteristics extraction of winter wheat (Table 1) (USGS). Considering the limitations of satellite service lifespans, this study utilised Landsat imagery from different satellites based on the availability of data for each year. All data were atmospherically corrected by the U.S. Geological Survey (USGS), ensuring image quality and accuracy. NDVI values were derived from the near-infra-red (SR_B4) and red (SR_B3) bands. To minimise cloud and shadow interference, the maximum NDVI value for each month was selected for subsequent seasonal variation analysis. Due to sensor malfunctions in early Landsat satellites, we employed an improved Spatial and Temporal Adaptive Reflectance Fusion Model (STARFM) to generate fused images that effectively compensate for the missing observations through performing localised cross-sensor calibration between MODIS and Landsat data (Li et al. 2020; N. A. S. A.). MODIS NDVI was calculated using the near-infra-red (sur_refl_b02) and red (sur_refl_b01) bands, with images resampled to 30 metres to match Landsat resolution. In particular, Landsat 5 and 7 data were primarily used for 2000–2003, Landsat 5 primarily for 2004–2013, and Landsat 8 primarily for 2014–2024. MODIS data supplemented the analysis in years with incomplete Landsat coverage (2001, 2012, 2013). Data were processed for the period February–June (2000–2024). Monthly NDVI calculations were performed for all remote sensing images within the study area, and the maximum NDVI value for each month was extracted based on a pixel-wise comparison. The calculation of the maximum NDVI value effectively reduced the interference of clouds and shadows on data quality, thus providing the best representation of vegetation cover for the region in that month.

2.2.2. Training and testing sample data

The winter wheat training and testing samples for 2024 were systematically collected through field investigations conducted by our research team across 16 prefecture-level cities in Shandong Province. Field sampling was executed between May 31 and June 8, 2024, resulting in a comprehensive dataset consisting of 76 wheat sample points and over 900 background reference points (Supplementary Figure S1). The geographic coordinates for each sampling location were precisely recorded using handheld GPS receivers, with longitudes and latitudes measured to seven decimal places, ensuring submeter positional accuracy. This level of precision ensured both the spatial accuracy and data reliability necessary for subsequent agricultural remote sensing analyses. Finally, the winter wheat training dataset for 2024 consisted of 60 winter wheat points and 90 randomly selected background points, while the testing dataset for 2024 included 16 winter wheat points and 30 randomly selected background points to evaluate the classification performance.

The winter wheat training samples for 2000–2023 were derived from a lower spatial resolution crop dataset, specifically the '1km-grid crop harvesting area dataset for three main crops of China from 2000 to 2019' (https://www.nesdc.org.cn/). The 1km-grid crop harvesting area dataset utilises GLASS LAI data, combined with the inflection point and threshold methods, to extract key phenological information. A comparison with county-level statistical data shows that the average coefficient of determination (R^2) for wheat distribution is 0.85, with a root mean square error (RMSE) of 10.02 thousand hectares (Luo et al. 2020). To optimise sample size selection, we conducted statistical validation by randomly selecting four years of imagery and testing model performance with 200 versus 300 samples per category. The results indicated that beyond 200 samples, the R^2 showed no significant improvement (fluctuation < 0.02), whereas the minimum increase in the MAE exceeded 10%. Therefore, we ultimately chose 80 winter wheat points that were closer to the actual measured points, as well as 120 non-wheat points. The sampling process in this study was as follows. It was initially assumed that the locations of land use would remain relatively stable within each five-year period, given the consistency of land policies in China. The winter wheat distribution points dataset, derived from the 1 km-grid crop dataset, was further subdivided into four periods: 2000–2004, 2005–2009, 2010–2014, and 2015–2019. Based on this, a refined winter wheat points dataset was constructed by creating intersection datasets using the five-year datasets for each period, and then extracting the intersecting points within a 2 km radius of the 2024 field-sampled points. Similarly, a refined non-winter wheat points dataset was constructed following the same procedure as the winter wheat points dataset. Ultimately, winter wheat training sample datasets for 2000–2019 were generated, with each dataset containing 80 high-accuracy wheat sampling points and 120 non-wheat sampling points per year. This approach significantly enhanced the spatial resolution and classification accuracy, providing a robust foundation for subsequent analyses. Winter wheat training sample datasets for 2020–2023 were based on the 1-km grid dataset from 2019, and sample points within a 2 km radius of the field-measured point data from 2024 were extracted, including 80 winter wheat points and 120 randomly selected background points. The cross-validation dataset for each year from 2000 to 2023 was created by partitioning the annual sample dataset using a hold-out method, where approximately two-thirds of the samples were allocated to the training set and the remaining one-third to the test set. Additionally, user validation testing data for the same period was acquired from official statistics bulletins (http://tjj.shandong.gov.cn/).

2.2.3. Mask data

The mask data for extracting the winter wheat distribution were reconstructed using the 'China Land Cover Dataset (CLCD)' provided by Wuhan University (Yang and Huang 2021), and the 'China's Multiperiod Land Use Land Cover Remote Sensing Monitoring Dataset (CNLUCC)' supplied by the Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences (Xu et al. 2018). The CNLUCC dataset uses a two-tier classification system, and the dryland classification data were resampled to their enhance spatial resolution from the original 1 km to 30 m. In contrast, the cropland data acquired from the CLCD, which already had a spatial resolution of 30 m, were utilised directly. To improve the accuracy of the land use data, the intersection of these two datasets was selected as the primary land use dataset for this study. The extracted winter wheat areas were further refined by applying a mask based on this dataset, effectively excluding the interference from other land use types.

2.2.4. Validation data

The extraction methods were verified using multilevel statistical data. City-level winter wheat planting area data for 16 cities were obtained from the Shandong Statistical Yearbook (2000–2023) published by the Shandong Provincial Bureau of Statistics (Statistics, S. P. B. o 2023). County-level data from 2006–2007 and 2022–2023 were acquired from the National Bureau of Statistics of China (https://data.cnki.net/). These datasets are widely recognised for their reliability and accuracy, making them the most precise sources of agricultural statistics.

To compare the performance of the proposed method with that of the competing approaches, three representative datasets were selected. The ChinaWheat10 dataset (CNW10) (Yang et al. 2023), with 10 m spatial resolution covering 2018–2024, was developed through training data generation and model transfer techniques using integrated Sentinel-2 and Sentinel-1 imagery combined with the Winter Crop Index. The Winter Wheat Planting Distribution over China (WWDC) (Dong et al. 2020), with 30 m spatial resolution spanning 2001–2024, was generated through the time-weighted dynamic time warping method applied to Landsat and Sentinel imagery. Additionally, the harvesting area dataset for China's staple crops (IHAC) with 1 km resolution from 2000–2019 was used in the comparative analysis (Luo et al. 2020).

2.3. Methodology

2.3.1. Methods for generating a long-term NDVI of winter wheat stack

The normalised difference vegetation index (NDVI) is a widely used remote sensing metric for assessing vegetation growth and ecosystem health statuses. It is calculated using the reflectance difference between the near-infra-red (NIR) and red (Red) spectral bands, as expressed in the following equation:

$$NDVI = \frac{NIR - RED}{NIR + RED} \quad (1)$$

where NDVI values range from −1 to +1, and higher values indicate healthier vegetation.

Winter wheat follows a distinct growth cycle, sown between September and October and harvested in June (Liu et al. 2023). Seasonal NDVI variations reflect different winter wheat growth stages, from dormancy to maturation. NDVI remains low during the overwintering period (mid-December to February) and rises in March as regreening begins, peaking in April and May during jointing and heading. A decline in June indicates maturation and approaching harvest (Qu, Li, and Zhang 2021; Roznik, Boyd, and Porth 2022; Vannoppen and Gobin 2021). While garlic exhibits phenologically similar NDVI trajectories to winter wheat (Ryu, Oh, and Cho 2021), the NDVI of it are lower than that of winter wheat. Additionally, their cultivation patterns differ significantly. Winter wheat dominates Shandong's flat drylands and paddy fields as a primary staple crop, whereas garlic cultivation demonstrates strong spatial aggregation within scattered, intercropped areas. According to multi-year data from the Shandong Provincial Department of Agriculture and Rural Affairs, garlic occupied merely 2.6865 million mu (3.8% of winter wheat area) in 2024. Crucially, incorporating arable land masks (Zhang et al. 2018; Zhang, Qiu, and Qin 2019) combined with soil-type exclusion effectively isolates winter wheat fields, as garlic's fragmented distribution exerts negligible impact on classification accuracy after these filtering processes.

To construct a long-term NDVI stack for winter wheat in Shandong Province, this study utilised multitemporal satellite imagery from Landsat and MODIS, processed within the Google Earth Engine (GEE) platform. The maximum NDVI value of every sampling point within each month was extracted to capture seasonal dynamics. The preprocessing workflow included spatial cropping to the study area, resampling of MODIS data using bilinear interpolation for spatial consistency, and cloud masking to remove contaminated pixels. Radiometric and atmospheric corrections were applied to enhance data quality. The constructed long-term NDVI stack enables an assessment of winter wheat growth trends at five-year intervals, facilitating the identification of temporal variations and potential influencing factors. Figure 3 presents the multi-temporal NDVI stack of winter wheat derived from remote sensing imagery.

2.3.2. Machine learning algorithms for winter wheat extraction

This study employed three widely used decision tree-based ensemble learning algorithms —Random Forest (RF), Extreme Gradient Boosting (XGBoost), and Categorical Boosting (CatBoost)— to classify winter wheat planting areas based on a long-term NDVI dataset (2000–2024).

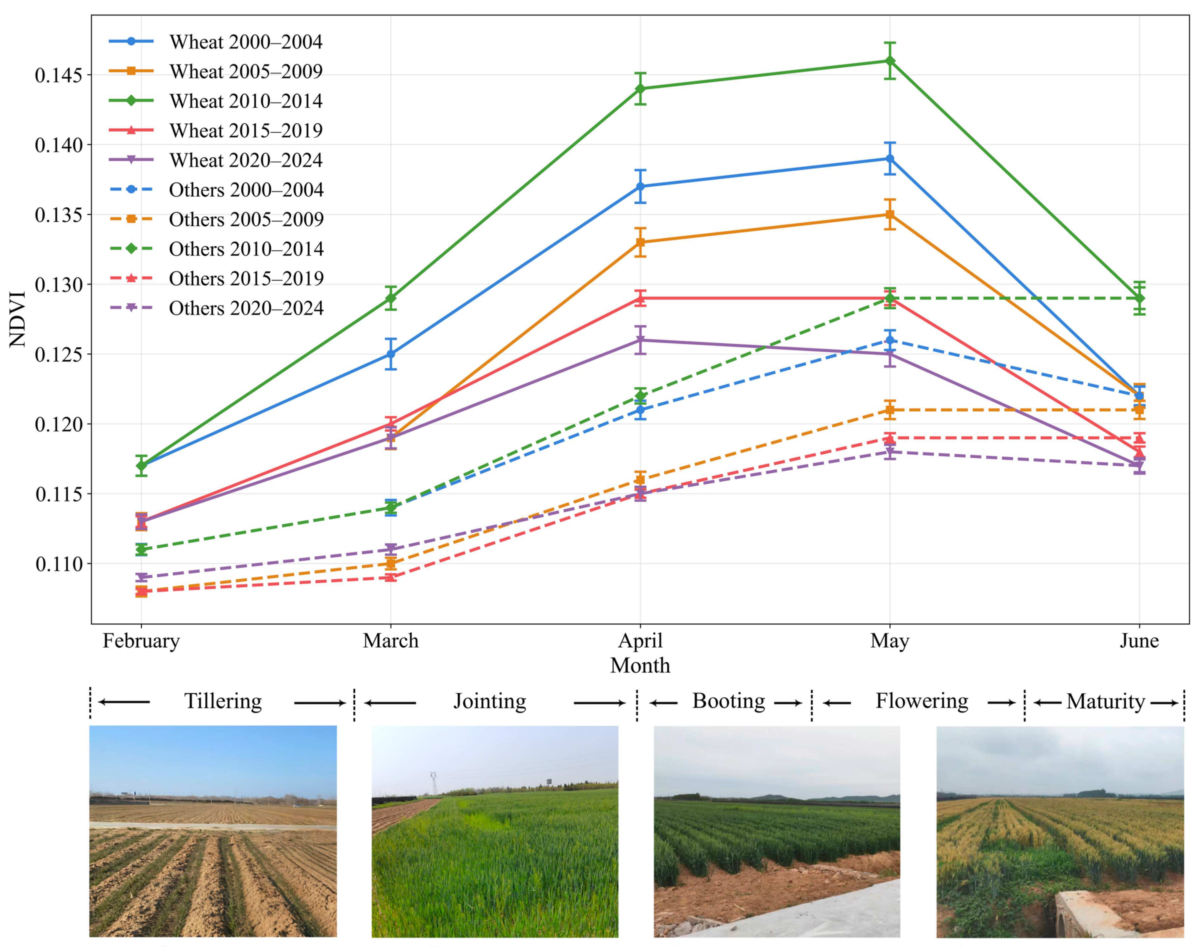

Figure 3. Average values for sampling points of monthly maximum NDVI in Shandong Province at five-year Intervals.

RF is an ensemble learning method that constructs multiple decision trees using bootstrap sampling at both data and feature levels. This randomisation introduces diversity among the trees, reducing correlation and lowering model variance, effectively mitigating overfitting. RF efficiently handles high-dimensional data, automatically identifies feature interactions, and supports parallel computation, making it well-suited for large-scale datasets (Asadollah, Jodar-Abellan, and Pardo 2024; Wang et al. 2022). The predictive formula of RF can be expressed as follows:

$$y_{RF} = \frac{1}{T}\sum_{t=1}^{T} f_t(x) \tag{2}$$

where y_{RF} represents the predicted output, T is the total number of decision trees, and $f_t(x)$ is the prediction result from the t^{th} decision tree.

XGBoost is an optimised implementation of Gradient Boosted Decision Trees (GBDT), designed for high efficiency and scalability (Bentéjac, Csörgő, and Martínez-Muñoz 2021). It refines predictions iteratively using second-order Taylor expansion and incorporates a regularisation term in the loss function to prevent overfitting, enhancing model generalisation (Zhang et al. 2021). XGBoost supports column sampling, missing value handling, and parallel computation, making it effective for large datasets (Zhao et al. 2021). The objective function of XGBoost can be expressed as:

$$\begin{aligned} y_{XG} &= \sum_{i=1}^{n} l(y_i, \hat{y}_i^{(t-1)} + f_t(x_i)) + \Omega(f_t) + C, \\ l(y_i, \hat{y}_i) &= -y_i \log(\hat{y}_i) - (1 - y_i)\log(1 - \hat{y}_i) \end{aligned} \tag{3}$$

where $\hat{y}_i^{(t-1)}$ is the predicted value from the $t-1^{th}$ iteration, $f_t(x_i)$ is the prediction from the t^{th} tree, $\Omega(f_t)$ is the regularisation term for the t^{th} tree, and the term C refers to the regularisation terms from the previous $t-1$ iterations, which are considered constants in the current iteration.

CatBoost, another GBDT-based algorithm, is optimised for categorical feature processing (Uribeetxebarria, Castellón, and Aizpurua 2023; Zhai et al. 2023). It employs ordered boosting to reduce gradient bias and prediction shifts, improving accuracy and generalisation. CatBoost minimises the need for manual preprocessing, performs well on large-scale categorical datasets, and is optimised for memory efficiency. Its predictive model follows:

$$y_{Cat} = \sum_{t=1}^{T} \alpha_t \cdot \sum_{j=1}^{J_t} \beta_{tj} \cdot I(x \in R_{tj}) \tag{4}$$

where y_{Cat} is the output label, T is the total number of trees, J_t is the number of leaf nodes in the t^{th} tree (consistent with oblivious tree structure, R_{tj} denotes the j^{th} leaf region of tree t), α_t is the step size for the t^{th} tree, β_{tj} is the optimal leaf weight for the j^{th} leaf of the t^{th} tree, computed via minimising the loss on the leaf's samples, and $I(x \in R_{tj})$ is the indicator function, which represents whether the input feature x belongs to the j^{th} category of the t^{th} tree.

For model optimisation in this study, CatBoost parameters were fine-tuned using GridSearchCV. The trained classification model was applied to NDVI raster images, generating high-precision classification results.

2.3.3. Accuracy assessment metrics

In this study, the accuracy of machine learning models was evaluated by comparing the estimated winter wheat planting areas with the statistical yearbook records at both the city and county levels. Annual accuracy metrics, including the coefficient of determination (R^2), root mean square error (RMSE), and mean absolute error (MAE), are used to assess predictive performance and reliability. R^2 quantifies the proportion of variance in observed values explained by the model, ranging from 0 to 1, with higher values indicating better performance:

$$R^2 = 1 - \frac{RSS}{TSS} = 1 - \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2} \tag{5}$$

where y_i and $\hat{y}_i$ denote actual and predicted values, respectively, n is the number of samples, and $\bar{y}$ is the mean of the true values.

RMSE evaluates prediction accuracy by measuring the square root of the mean squared errors, making it sensitive to large deviations:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{6}$$

MAE, defined as the mean absolute difference between predicted and actual values, provides a robust error measure less sensitive to outliers:

$$MAE = \frac{1}{n}\cdot\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{7}$$

While RMSE emphasises larger errors, MAE offers a balanced assessment, making their combined use beneficial for evaluating model performance.

2.3.4. Spatiotemporal analysis of winter wheat

To investigate the spatiotemporal dynamics of winter wheat cultivation in Shandong Province, the standard deviation ellipse and centroid migration model were employed in this study. The standard deviation ellipse, a robust spatial statistical tool, characterises the spatial pattern and temporal dynamics of wheat planting distribution through key parameters including the major axis, minor axis, and orientation angle. Specifically, the major axis delineates the primary directional trend of cultivation, the minor axis corresponds to the secondary distributional axis, and the ellipse area quantifies the spatial extent of planting coverage. Additionally, spatial statistical analyses were utilised to quantify temporal variations in winter wheat cultivation area. These analyses were conducted using the ArcGIS 10.2 platform and Origin 2024.

2.3.5. Intensive analysis of winter wheat

The landscape intensive index serves as an indicator of structural and functional changes within ecosystems. We introduced this index to investigate the intensification level of winter wheat cultivation areas in Shandong Province. Patch density (PD) quantifies the number of patches per unit area, with higher values indicating more severe fragmentation. The largest patch index (LPI) measures the dominance of the largest patch within the landscape, where elevated values suggest stronger control by a single dominant patch. The aggregation index (AI) evaluates the spatial clustering of homogeneous patches, with higher values reflecting greater patch connectivity and aggregation. We employed Fragstats 4.2 software to calculate the aforementioned landscape indices across Shandong Province for the period 2000–2024.

3. Results

3.1. Comparison among ensemble learning methods for winter wheat area extraction

The performance of three ensemble learning algorithms—RF, XGBoost, and CatBoost— were compared using a city-level accuracy validation over the period from 2000-2023. The results showed that all three models exhibited strong fitting capabilities, with the average R^2 reaching 0.9 during this period (Figure 4). Notably, the CatBoost model outperformed the other two algorithms, achieving a slightly higher average R^2 of 0.91. In addition, the CatBoost model showed lower prediction errors, with average MAE and average RMSE values of 49.65×10^3 hm^2 and 64.44×10^3 hm^2, respectively, which were both significantly smaller than those of RF and XGBoost ($p < 0.05$). Overall, the CatBoost model emerges as the most promising approach for extracting winter wheat planting areas, offering significant advantages in terms of both MAE and RMSE.

3.2. Performance analysis of the optimal ensemble learning method

3.2.1. Cross-validation

The CatBoost model was further utilised to accurately delineate winter wheat planting areas in Shandong Province. A comprehensive cross-validation analysis was conducted using a testing dataset containing 46

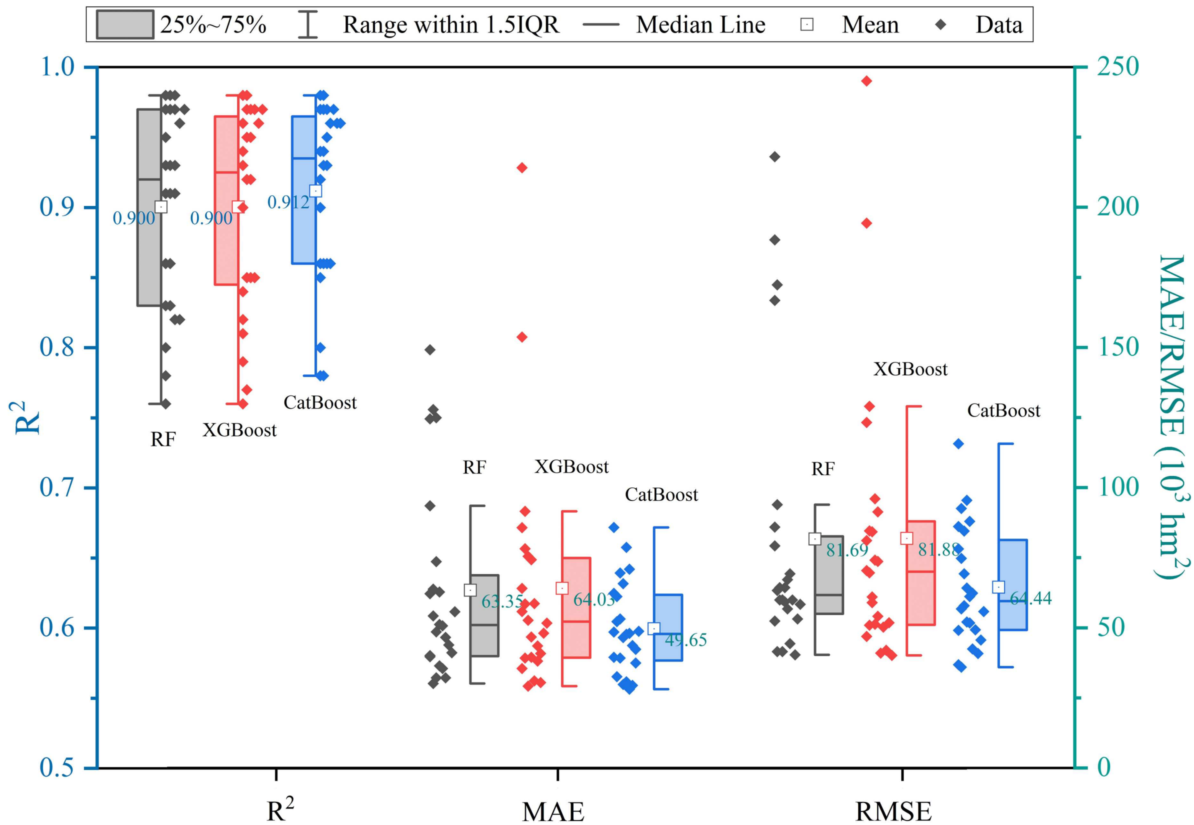

Figure 4. Comparison of accuracy results for three algorithms from 2000 to 2023.

samples per year from 2000 to 2024 (Figure 5). The results demonstrated high cross-validation accuracy, with values ranging from 0.85 to 0.98 and an average value of 0.913 across the 2000–2024 period. These findings indicate that the CatBoost model is highly effective and reliable for precisely mapping winter wheat cultivation areas.

3.2.2. City-level accuracy validation

City-level accuracy validation results for the period 2000–2023 were shown in Figure 6. The R^2 values ranged from 0.78 to 0.98, with most exceeding 0.86, indicating strong model performance for winter wheat extraction. Nevertheless, the model demonstrated robust performance overall, with an R^2 of 0.91, a MAE of 49.65×10^3 hm^2, and a RMSE of 64.44×10^3 hm^2. These results underscore the model's stability and suitability for large-scale, dynamic monitoring of winter wheat planting areas, even in years with limited data resolution.

3.2.3. County-level accuracy validation

The county-level results demonstrated strong model performance. Using Landsat 5 as the primary data source prior to 2014, the R^2 values of 0.79 in 2006 and 0.78 in 2007. The MAE was 7.6×10^3 hm^2 in 2006 and 8.9×10^3 hm^2 in 2007, while the RMSE was 10.8×10^3 hm^2 and 12.4×10^3 hm^2 respectively. With Landsat 8 adopted as the primary data source after 2014, R^2 values reached 0.91 in 2022 and 0.89 in 2023, with corresponding RMSE values of 8.4×10^3 hm^2 and 7.9×10^3 hm^2 (Figure 7). Compared to the municipal level, the R^2 at the county level slightly decreased, suggesting some performance loss when applying the model at a finer spatial scale. This may be due to the smaller sample size or increased spatial heterogeneity at the county level. However, both RMSE and MAE were lower at the county level, suggesting that finer-scale modelling places greater demands on the algorithm's stability and generalisation. The slopes of the fitted lines, except for 2007 (1.13), were all near the 1 ± 0.1 range, indicating that the model's predicted values demonstrate higher consistency with the statistical data within the overall

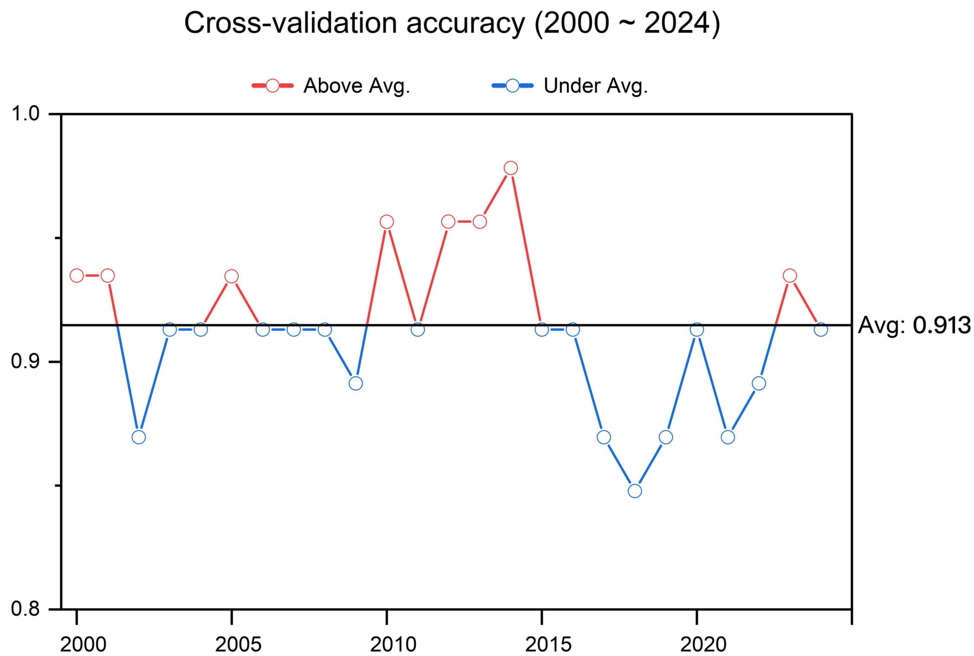

Figure 5. Cross-validation accuracy of winter wheat mapping using CatBoost (2000–2024).

error margin. Despite this, the county-level results remain within an acceptable range, demonstrating the model's strong regional transferability.

3.2.4. Comparison with other productions

The performance of the CatBoost model was also evaluated against three widely used datasets—CNW10, WWDC, and IHAC—using city-level validation metrics with the accuracy detailed in Table 2. The CatBoost model achieved a higher overall average accuracy than the WWDC during the period 2015-2023, under a fair comparison as both are based primarily on Landsat 8 data. The model also surpassed the IHAC in accuracy during their overlapping period during (2015–2019). When compared against the Sentinel-2-based CNW10 dataset for their overlapping period (2019–2023), the accuracy of the CatBoost model was only slightly lower than that of the high-accuracy, Sentinel-2-based CNW10 dataset. Overall, the CatBoost model demonstrated robust performance in winter wheat mapping. These results confirm its potential as a valuable tool for enhancing the precision and reliability of agricultural remote sensing monitoring.

3.3. Spatiotemporal variation in winter wheat in Shandong Province

3.3.1. Spatial patterns in winter wheat during 2000–2024

The winter wheat distribution in Shandong Province from 2000 to 2024, extracted using the CatBoost model, was shown in Figure 8. The results showed significant regional patterns, with key cultivation areas in the western and southern regions, including Binzhou, Dezhou, Liaocheng, Heze, Jining, and Linyi, which were crucial for national grain production (Figure 8). This distribution may be influenced by Shandong's topography, climate, soil, and policies: the western, flat areas support mechanised farming, while the central mountainous regions have lower planting density, and the eastern coastal areas are constrained by land salinization and climatic factors.

A marked regional variation in wheat cultivation across Shandong Province was identified using the centroid model and standard deviation ellipse analysis (Figure 9). Specifically, the northwestern and southwestern regions exhibit high-density wheat planting, with larger standard deviation ellipses indicating broader spatial distribution and larger scale. In 2000, the ellipse was relatively elongated along a southwest-northeast axis, reflecting a wide distribution of wheat cultivation in these directions. By 2024,

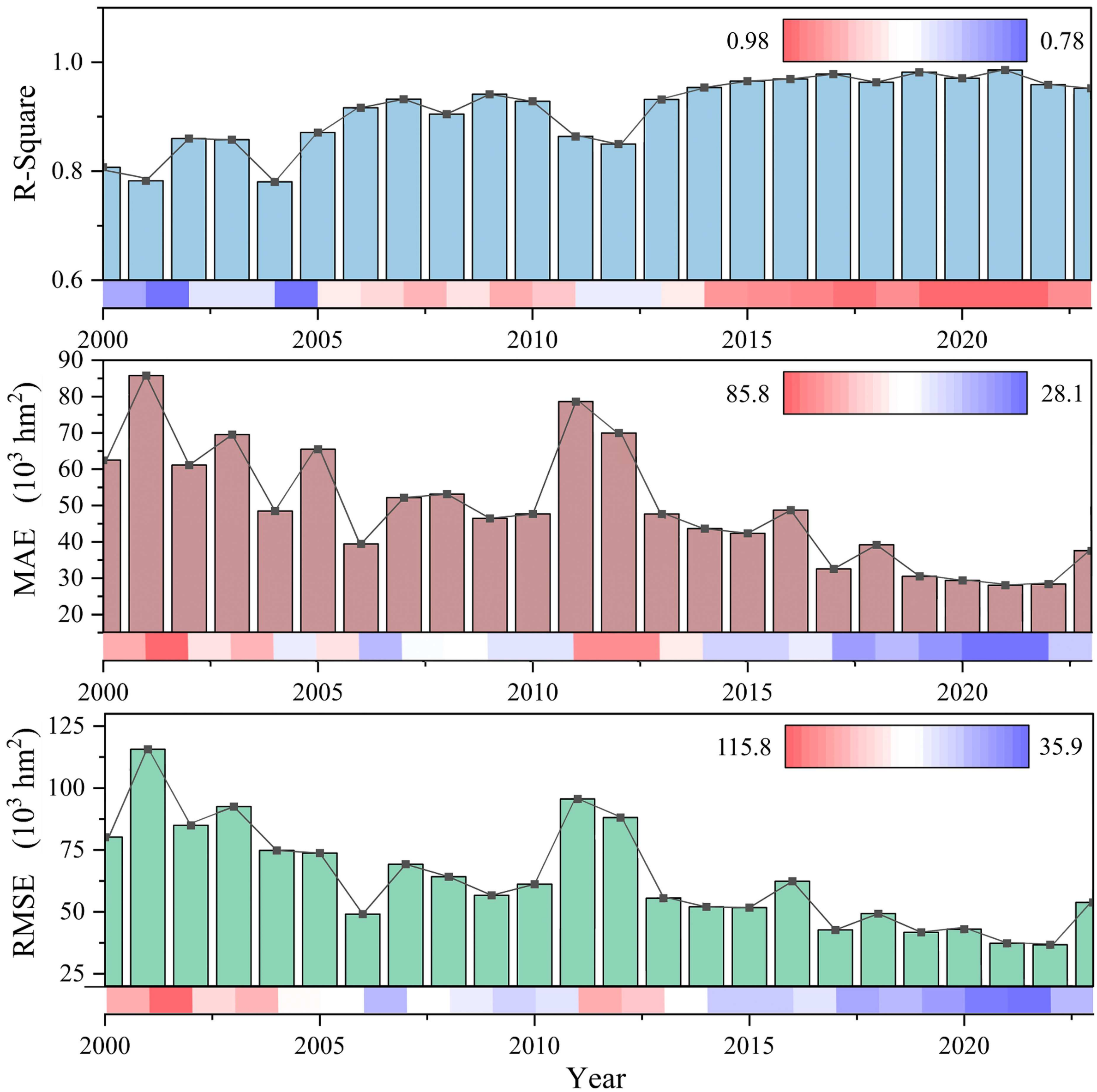

Figure 6. City-Level winter wheat mapping accuracy using CatBoost (2000–2023).

the ellipse area contracted, and the minor axis lengthened, signifying a more concentrated spatial distribution of wheat cultivation (Figure 9b). The centroid model analysis further revealed a gradual shift in the centre of wheat cultivation from 2000 to 2024 that the centre initially shifted southwestward, followed by a westward shift, and ultimately a northeastward shift (Figure 9c). Overall, the centre movement route of winter wheat remained largely confined within the boundaries of Taian city.

3.3.2. Temporal change in winter wheat during 2000–2024

The temporal changes in winter wheat planting areas from 2000 to 2024, derived through the CatBoost model, revealed significant spatiotemporal heterogeneity at both provincial and municipal scales (Figure 10). At the provincial level, a pronounced downward trajectory was observed, with the total winter wheat planting area decreasing by 20.4% over the 25 years (Figure 10a). By 2024, the planting area had declined to 2828.43×10^3 km^2, compared to the baseline value in 2000. Trend analysis confirmed this decline was not monotonic but exhibited contrasting phases across distinct periods. At the city level, the winter wheat planting area ranged from 20.1×10^3 km^2 to 485.7×10^3 km^2 in 2024, with Heze showing the largest area and Weihai the smallest (Figure 10b). Temporal trends in planting area varied considerably

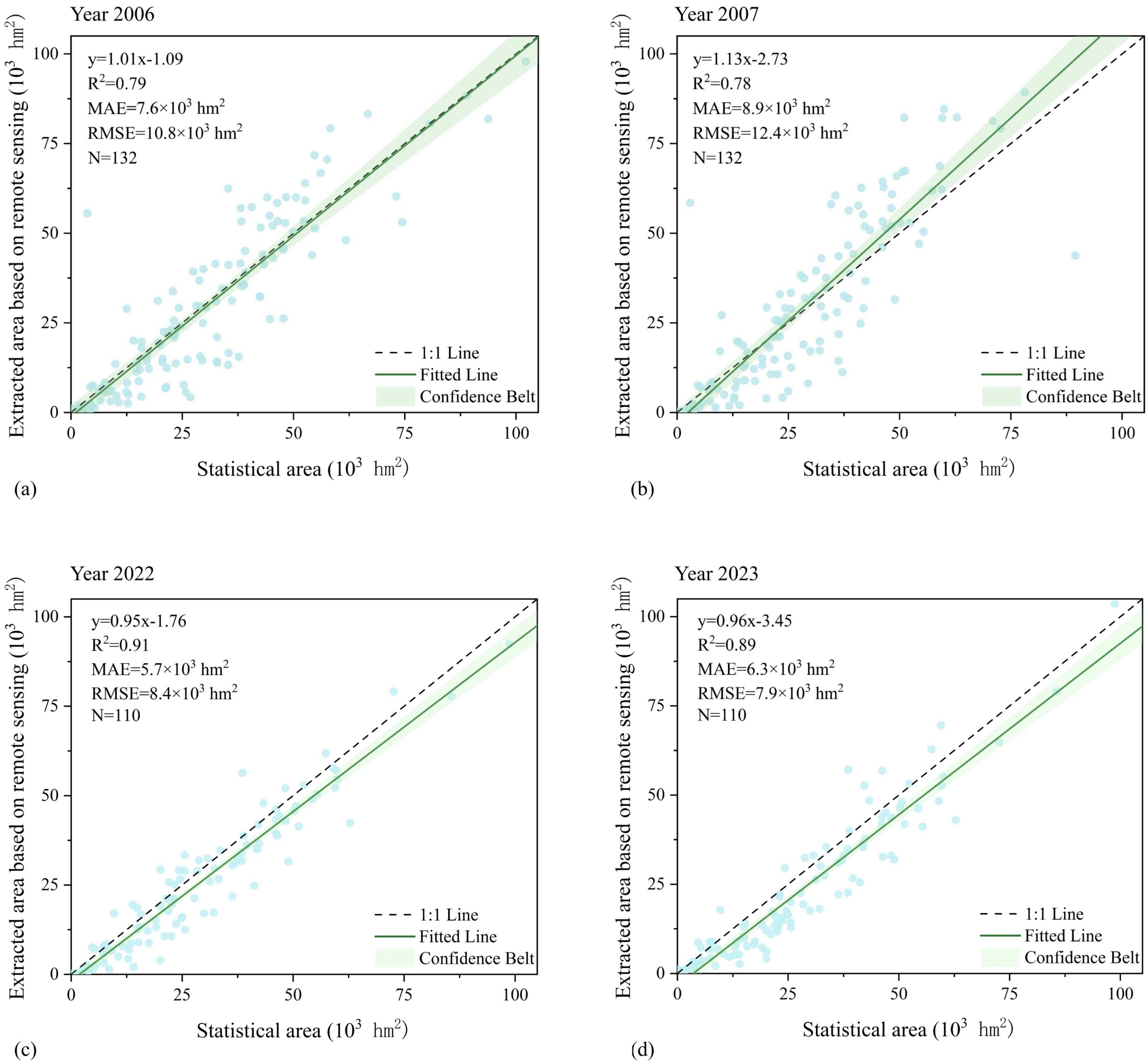

Figure 7. County-level validation of the winter wheat mapping accuracy using CatBoost regarding the multi-temporal datasets derived from Landsat 5: Year 2006 (a) and year 2007 (b), and Landsat 8: Year 2022 (c) and year 2023 (d). (The sample size (*N*) varies across datasets owing to incomplete data, with the total number of counties being 137.).

Table 2. Comparison of recent city-level accuracy: experimental dataset, CNW10, WWDC and IHAC.

Accuracy	Mean	Exp data	CNW10	WWDC	IHAC
R^2	mean_{2015_2023}	0.97	–	0.93	–
	mean_{2015_2019}	0.97	–	0.93	0.91
	mean_{2019_2023}	0.97	0.98	0.94	–
MAE ($\times 10^3$ hm^2)	mean_{2015_2023}	35.37	–	53.28	–
	mean_{2015_2019}	38.67	–	59.73	56.64
	mean_{2019_2023}	31.10	26.59	46.95	–
RMSE ($\times 10^3$ hm^2)	mean_{2015_2023}	46.76	–	67.85	–
	mean_{2015_2019}	49.58	–	74.31	70.72
	mean_{2019_2023}	43.00	37.05	61.34	–

among cities between 2000 and 2024. While several cities, including Dezhou, Binzhou, Dongying, Yantai, Weihai, Qingdao, and Heze, showed an overall increase, others experienced fluctuating declines. Thereinto, Weihai recorded the highest growth rate, increasing at a rate (*Mean* ± *SE*) of 242.2 ± 87.4% per year during the same period whereas Taian showed the fastest decline, with a rate of –2.2 ± 0.9% per

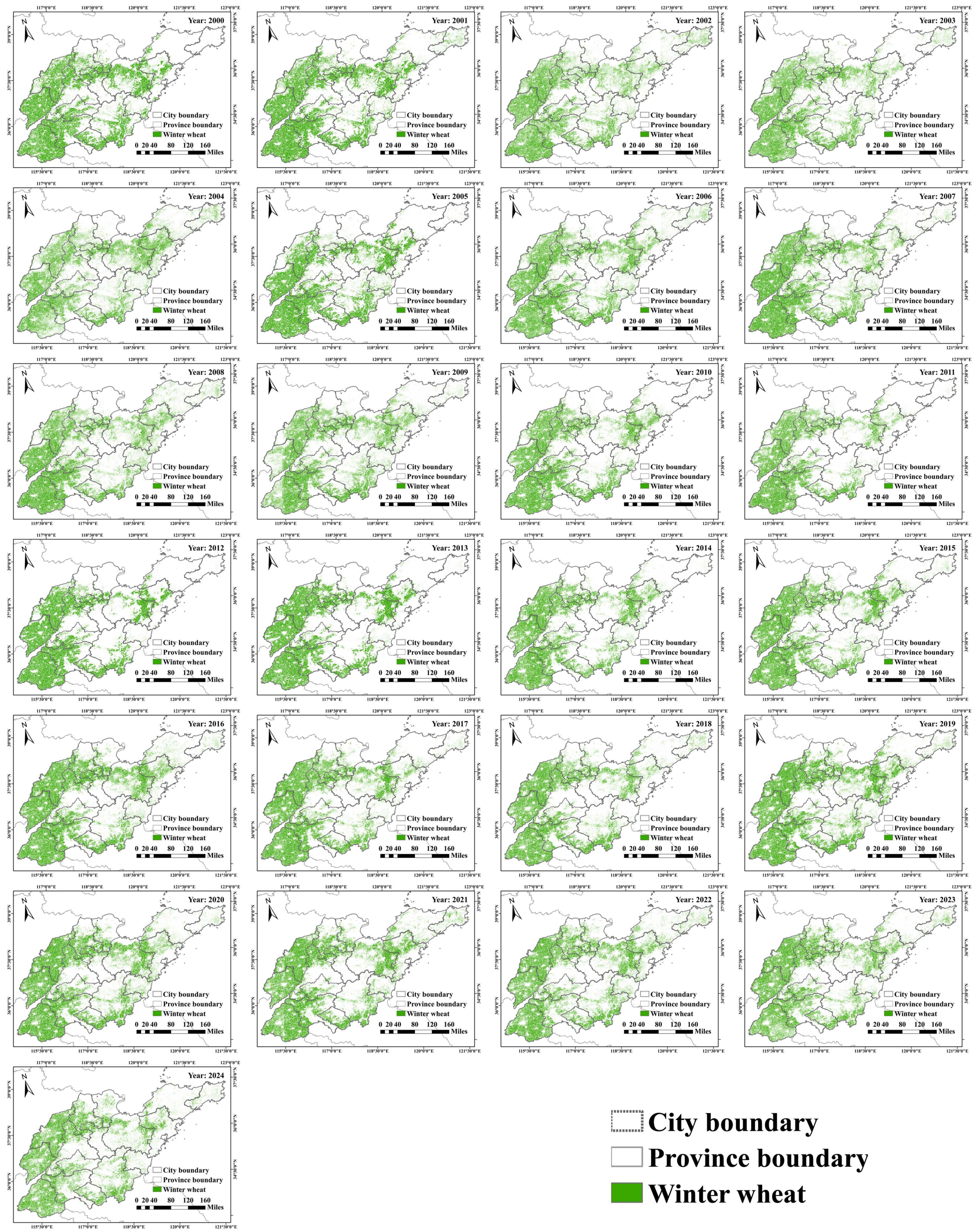

Figure 8. Multi-temporal spatial distribution of winter wheat in Shandong Province (2000–2024) derived from the CatBoost model.

year from 2000 to 2024 (Figure 10b). Notably, the increases in Dezhou, Binzhou, and Qingdao during the period of 2015–2024 were statistically significant according to the Mann-Kendall test ($p < 0.05$). Conversely, significant decreasing trends were observed in Jining and Tainan over the most recent two years.

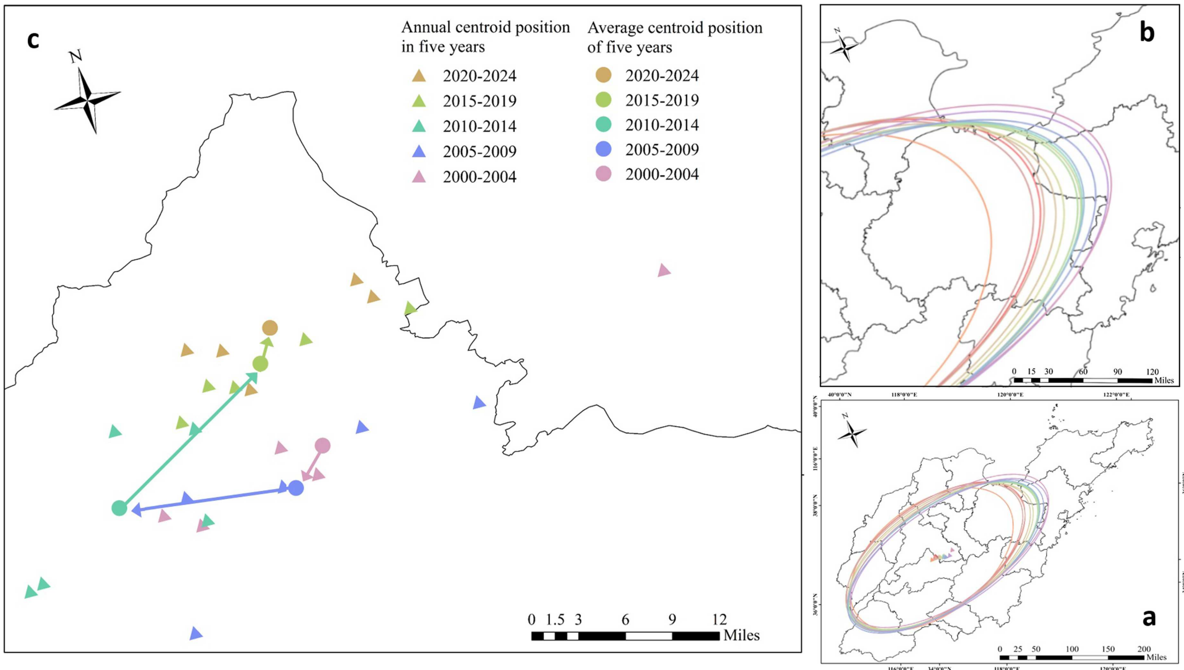

Figure 9. Spatiotemporal shifts in winter wheat cultivation centroids and standard deviation ellipses across Shandong Province (2000–2024): (a) overview map, (b) dynamics of the standard deviation ellipse, and (c) centroid trajectory analysis. Data were generated using the CatBoost model.

3.3.3. The trend in the intensification level of winter wheat

The analysis of winter wheat field fragmentation based on the CatBoost model from 2000 to 2024 (Figure 11) reveals that the PD exhibits a significant increasing trend with notable fluctuations during the study period. Conversely, both the LPI and the AI demonstrate significant decreasing trends, showing significantly negative correlations with PD. These coordinated changes in landscape pattern metrics indicate two typical landscape configuration states in the study area: (1) a 'large-patch-dominated, highly aggregated' type (characterised by low PD, high LPI, and high AI); and (2) a 'small-patch dispersed, non-dominant' type (characterised by high PD, low LPI, and low AI). Comprehensive analysis suggests a continuous decreasing trend in the intensification level of winter wheat planting patterns across the region, with the overall landscape pattern shifting significantly toward a 'small-patch dispersed, non-dominant' development trend.

Supplementary analysis of the spatial autocorrelation of winter wheat planting areas (Local Indicators of Spatial Association, LISA) further verifies the above landscape pattern changes. The observed trends in landscape pattern metrics exhibit strong spatial autocorrelation characteristics across Shandong Province. The shift toward a 'small-patch dispersed, non-dominant' configuration is not randomly distributed but shows significant spatial clustering patterns. Hot spot analysis reveals that high-PD areas (representing more fragmented patterns) are predominantly clustered in the eastern and central hilly regions of Shandong, including Rizhao, Linyi, Qingdao, and Weifang areas, where complex terrain and diverse cropping systems promote fragmentation. Conversely, low-PD areas are significantly clustered in the northwestern and southwestern plains of Shandong, such as Dezhou, Liaocheng, Heze, and Jining, where flat terrain and large-scale farming practices maintain field aggregation.

4. Discussion

4.1. Advantages of the CatBoost model in winter wheat planting area mapping

Accurate mapping of winter wheat planting areas is critical for agricultural policy formulation in China, yet existing methods struggle with balancing long-term temporal analysis and high spatial resolution (Hu

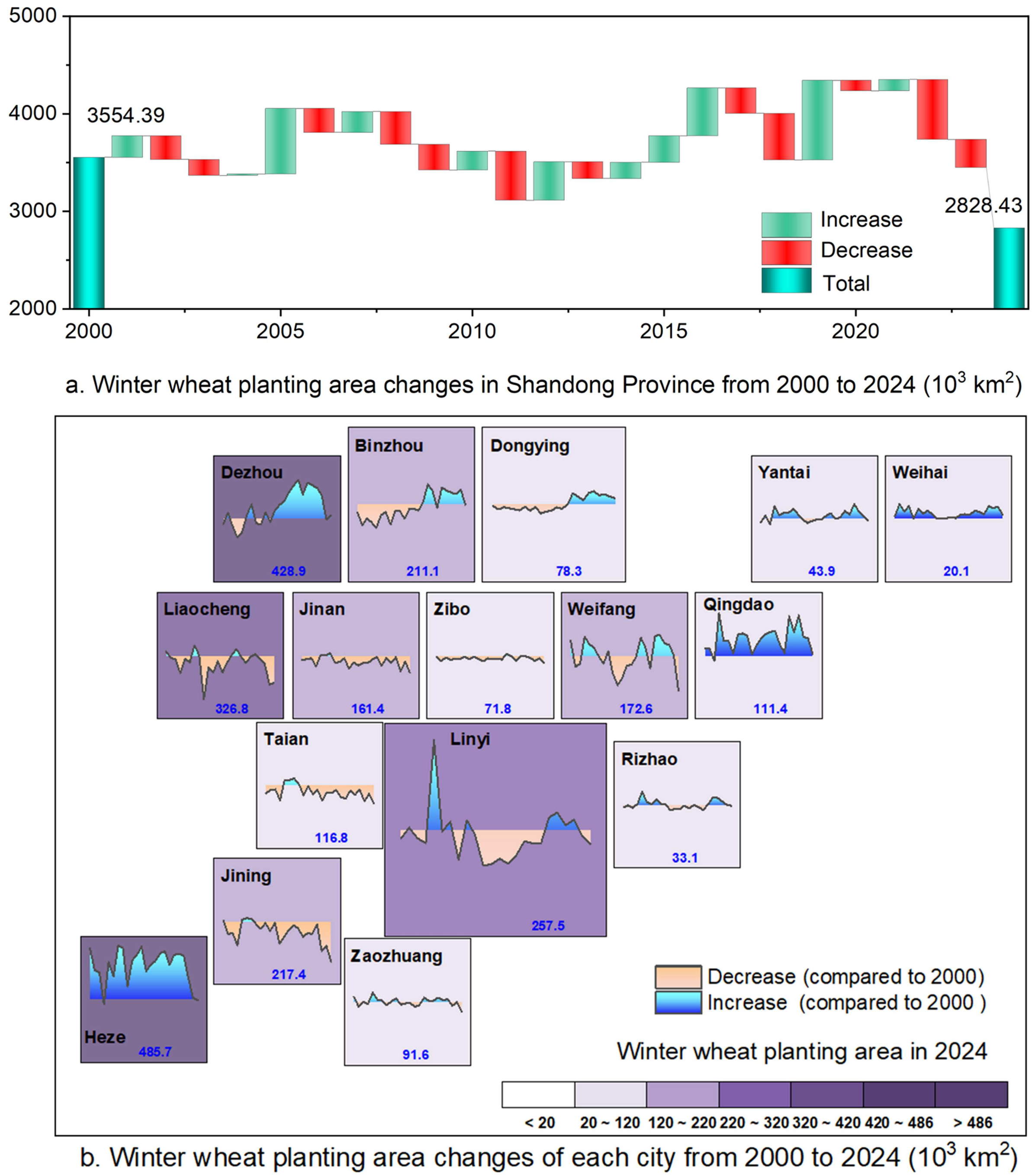

Figure 10. Temporal variation in winter wheat planting area (2000–2024): (a) temporal changes at the provincial scale, and (b) temporal changes at the city level relative to 2000. Data were derived using the CatBoost model.

et al. 2024), labour-intensive sampling (Nasrallah et al. 2018), and limited historical high-precision data (Xiao et al. 2019). Traditional approaches also face computational inefficiency or reliance on auxiliary inputs like meteorological data (Tian et al. 2021). To address these issues, this study employed Landsat-derived monthly maximum NDVI stacks and the CatBoost algorithm, selected for its computational efficiency and accuracy in processing heterogeneous geospatial data. Using this method, winter wheat planting maps with a spatial resolution of 30 m for the period of 2000–2024 were generated (Figure 8). The CatBoost model demonstrated significant advantages in winter wheat planting area mapping, addressing critical limitations of conventional approaches. Its ability to process small training samples (200 annual points: 80 wheat, 120 non-wheat) minimised reliance on labour-intensive ground surveys while

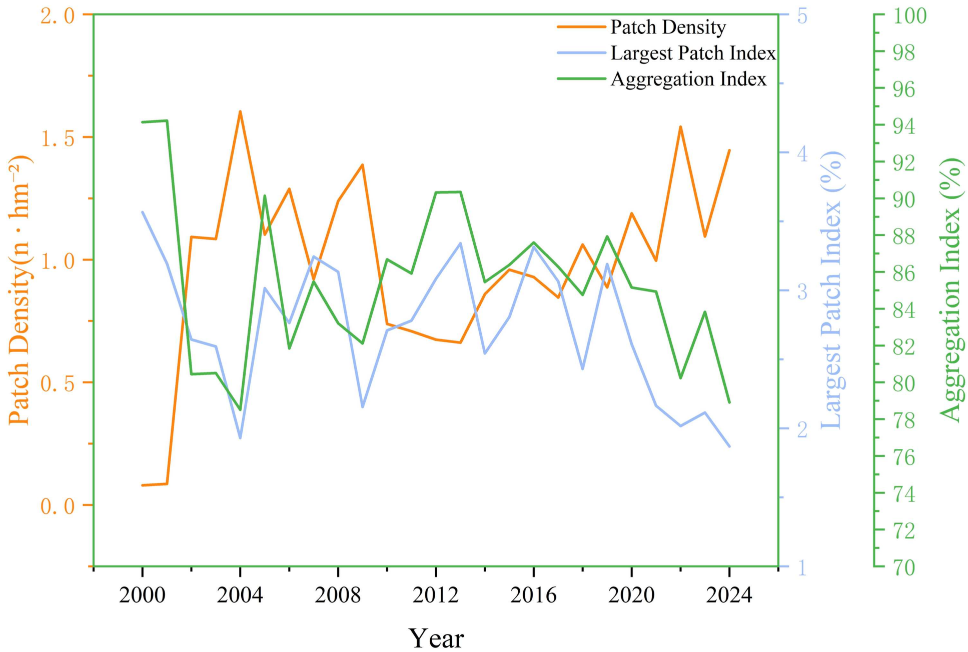

Figure 11. Analysis of winter wheat intensification patterns via PD, LPI, and AI indicators (2000–2024).

maintaining classification accuracy using a 1-km resolution reference dataset (Figure 5). The algorithm's native compatibility with high-dimensional, multi-temporal NDVI stacks eliminated complex preprocessing, enabling direct integration of Landsat time-series data. Robustness to overfitting was ensured through ordered boosting, which enhanced generalisation across a 25-year span (2000–2024) and diverse spatial scales (Figure 8). Coupled with GPU-optimised training efficiency and automatic handling of missing values, CatBoost produced a 30-metre resolution map validated against field surveys and statistical data ($R^2 > 0.9$; Figure 6), achieving precise identification at city and county levels. These capabilities resolve the trade-offs between high spatial resolution and long-term temporal coverage, positioning CatBoost as a scalable, resource-efficient solution for agricultural monitoring in data-constrained environments.

4.2. Driving forces of winter wheat planting area changes

This study observed a significant decline in winter wheat planting area in Shandong Province from 2000 to 2024 (Figure 8, Figure 10). Additionally, the planting centre exhibited a noticeable northward shift (Figure 9, Figure 10). The observed contraction in winter wheat planting area resulted from the combined effects of agroclimatic stressors, socio-economic transformations, and policy-driven dynamics. Specifically, agroclimatic factors played a pivotal role in this trend. Reduced precipitation and intensified drought stress directly suppressed yields and limited cultivation expansion (Han, Lin, and Wang 2023; Zhao et al. 2022). Although improved irrigation infrastructure in Yellow River diversion districts mitigated water deficits, the saturation of arable land use in key agricultural zones restricted further network expansion. Additionally, while increased thermal accumulation initially enhanced yields, this benefit was offset by a reduction in solar radiation in northern China, exacerbating production uncertainties and prompting crop zoning shifts toward drought-tolerant cultivars or low-vulnerability ecoregions (Dong, Xue, and Pan 2024). Socioeconomic drivers further contributed to this decline. Accelerated urbanisation, rising land opportunity costs, and expanding rural land lease markets collectively drove the conversion of farmland to non-agricultural uses (Zhang et al. 2021). Geospatial heterogeneity and socioeconomic gradient differentials further intensified regional divergence (He, Ding, and Tian 2022). For instance, in coastal and peri-urban plains of the study area, profit-driven land reallocation favoured non-grain production systems like horticulture and photovoltaic land uses, displacing wheat cultivation (Sun et al. 2021; Zhu et al. 2022). Conversely, sustained irrigation investments and policy interventions stabilised wheat planting in the Luxi

Alluvial Plain, though diminishing marginal returns to fertiliser inputs and mechanisation indicate a nearing yield plateau (Figure 8). Policy interventions exhibited a dual influence. While subsidies and farmland protection policies stabilised planting intentions in core production zones, economic restructuring pressures, rural workforce aging, and the expansion of high-value crops like facility agriculture accelerated marginal land abandonment (Wang et al. 2021). For example, the rapid proliferation of non-staple production paradigms, exemplified by facility agriculture expansion on the Jiaodong Peninsula, coupled with accelerated rural workforce aging from urbanisation, amplified marginal land abandonment rates. Overall, these combined forces—particularly in ecotonal fragility zones and economic hotspots— have collectively driven the observed contraction in winter wheat planting area.

4.3. Uncertainty analysis

This study contains several potential uncertainties that may affect identification accuracy and result consistency. First, the limited spatiotemporal coverage of early satellite imagery introduced data gaps, restricting the ability to monitor long-term winter wheat cultivation trends. Although MODIS data supplemented this, its lower spatial resolution reduced the precision of wheat area extraction. Second, inconsistencies in spectral information—specifically, variations in spectral response functions, radiometric calibration, and atmospheric correction algorithms between different satellite sensors—from different data sources during maximum NDVI synthesis may have compromised time-series coherence and result stability. Third, the absence of high-precision county-level validation data constrained model reliability and spatial distribution accuracy, potentially introducing regional biases. To mitigate these uncertainties, future studies should prioritise the integration of higher-resolution input data and improved parameter tuning at the county level to enhance prediction accuracy. Moreover, adopting advanced deep learning algorithms, such as convolutional neural networks (CNN) or spatiotemporal convolutional neural networks (ST-CNN) (Ye et al. 2022), could further improve model robustness by capturing complex spectral and spatiotemporal patterns. Additionally, combining deep learning techniques with innovative approaches like Pixel Matching Models (PMM) may streamline the extraction process, reduce manual labelling efforts, and enhance the accuracy of winter wheat area identification.

5. Conclusion

This study presented a method for extracting winter wheat planting areas in Shandong Province from 2000 to 2024 using decision tree-derived ensemble learning models combined with multisource remote sensing data. By leveraging maximum NDVI change curves, the proposed method achieved dynamic monitoring at a 30 m spatial resolution. Results demonstrated that the CatBoost model outperformed XGBoost and RF in both accuracy and error control, achieving an optimal R^2 of 0.91 and maintaining robust performance across city and county scales. The model's superior alignment with statistical data, along with its improved reliability compared to existing datasets, underscored its effectiveness in capturing the dynamics of wheat planting. Spatiotemporal analysis revealed a significant decline in winter wheat planting area in Shandong Province from 2000 to 2024, and the planting centre exhibited a noticeable northward shift. The spatiotemporal dynamics analysis of winter wheat planting distribution in Shandong Province revealed that the high-resolution distribution maps generated in this study effectively captured the complex influences of climate, terrain, socio-economic factors, and policies on wheat planting patterns. Despite strong model performance, limitations of MODIS data resolution constrained accuracy in certain years. Future research should focus on integrating higher-resolution data, optimising parameters, and refining sampling strategies to improve county-level predictions. This study indicates that winter wheat cultivation is exhibiting a spatial dispersion trend, which complicates irrigation management. Future improvements in water use efficiency could be achieved through measures such as land consolidation, breeding drought-resistant cultivars, and the integration of water and fertiliser management. Overall, this study offered a reliable method for large-scale, long-term wheat area monitoring, providing valuable insights for agricultural planning, resource management, and food security strategies.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (No. 42301327), and the Key Technology Research and Development Program of Shandong Province (No. 2021ZDSYS01). We thank the anonymous reviewers and editors for their helpful comments and suggestions for the manuscript.

Author contributions

CRediT: **Siqi Yang:** Conceptualization, Data curation, Formal analysis, Methodology, Visualization, Writing – original draft; **Caiyun Deng:** Conceptualization, Data curation, Formal analysis, Funding acquisition, Methodology, Writing – review & editing; **Tianhe Xu:** Funding acquisition, Project administration, Supervision, Writing – review & editing; **Ran Kang:** Data curation, Validation, Writing – review & editing; **Huiying Yin:** Data curation, Methodology; **Jian Guo:** Data curation, Visualization; **Li Zhang:** Data curation, Visualization; **Lulu Si:** Data curation; **Hermann Josef Kaufmann:** Supervision, Writing – review & editing.

Disclosure statement

The authors declare that there are no conflicts of interest regarding the publication of this article.

Data availability statement

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

References

Asadollah, S. B. H. S., A. Jodar-Abellan, and M. Á Pardo. 2024. "Optimizing Machine Learning for Agricultural Productivity: A Novel Approach with RScv and Remote Sensing Data over Europe." *Agricultural Systems* 218: 103955. https://doi.org/10.1016/j.agsy.2024.103955.

Bentéjac, C., A. Csörgő, and G. Martínez-Muñoz. 2021. "A comparative analysis of gradient boosting algorithms." *Artificial Intelligence Review* 54: 1937–1967. https://doi.org/10.1007/s10462-020-09896-5.

Blickensdörfer, L., M. Schwieder, D. Pflugmacher, C. Nendel, S. Erasmi, and P. Hostert. 2022. "Mapping of Crop Types and Crop Sequences with Combined Time Series of Sentinel-1, Sentinel-2 and Landsat 8 Data for Germany." *Remote Sensing of Environment* 269: 112831. https://doi.org/10.1016/j.rse.2021.112831.

Dong, Z., X. Xue, and Z. Pan. 2024. "An Assessment of the Vulnerability to Climate Change of Winter Wheat and Summer Maize Production in Shandong Province, China." *Theoretical and Applied Climatology* 155: 1885–1896. https://doi.org/10.1007/s00704-023-04729-1.

Dong, J., Z. Pang, Y. Fu, Q. Peng, X. Li, and W. Yuan. 2024. "Annual Winter Wheat Mapping Dataset in China from 2001 to 2020." *Scientific Data* 11: 1218. https://doi.org/10.1038/s41597-024-04065-7.

Dong, J., Y. Fu, J. Wang, H. Tian, Z. Niu, W. Han, Y. Zheng, J. Huang, and W. Yuan. 2020. "Early-Season Mapping of Winter Wheat in China Based on Landsat and Sentinel Images." *Earth System Science Data* 12: 3081–3095. https://doi.org/10.5194/essd-12-3081-2020.

Fan, L., S. Chen, S. Liang, X. Sun, L. You, W. Wu, and P. Yang. 2020. "Assessing Long-Term Spatial Movement of Wheat Area Across China." *Agricultural Systems* 185: 102933. https://doi.org/10.1016/j.agsy.2020.102933.

Fan, L., L. Xia, J. Yang, X. Sun, S. Wu, B. Qiu, and J. Chen. 2024. "A Temporal-Spatial Deep Learning Network for Winter Wheat Mapping Using Time-Series Sentinel-2 Imagery." *ISPRS Journal of Photogrammetry and Remote Sensing* 214: 48–64. https://doi.org/10.1016/j.isprsjprs.2024.06.005.

Feng, L., Y. Wang, Z. Zhang, and Q. Du. 2021. "Geographically and Temporally Weighted Neural Network for Winter Wheat Yield Prediction." *Remote Sensing of Environment* 262: 112514. https://doi.org/10.1016/j.rse.2021.112514.

Feng, Y., B. Chen, W. Liu, X. Xue, L. Zhu, and H. Xing. 2024. "Winter Wheat Mapping in Shandong Province of China with Multi-Temporal Sentinel-2 Images." *Applied Science* 14: 3940. https://doi.org/10.3390/app14093940.

Han, W., X. Lin, and D. Wang. 2023. "Uncovering the Primary Drivers of Regional Variability in the Impact of Climate Change on Wheat Yields in China." *Journal of Cleaner Production* 421: 138479. https://doi.org/10.1016/j.jclepro.2023.138479.

He, H., R. Ding, and X. Tian. 2022. "Spatiotemporal Characteristics and Influencing Factors of Grain Yield at the County Level in Shandong Province, China." *Scientific Reports* 12: 12001. https://doi.org/10.1038/s41598-022-14801-x.

Hou, P., Y. Liu, H. Yang, R. Xie, K. Wang, B. Ming, J. Xue, et al. 2021. "Quantifying Maize Grain Yield Losses Caused By Climate Change Based on Extensive Field Data Across China." *Resources, Conservation and Recycling* 174: 105811. https://doi.org/10.1016/j.resconrec.2021.105811.

Hu, J., B. Zhang, D. Peng, J. Huang, B. Zhao, Y. Li, E. Cheng, et al. 2024. "Mapping 10-m Harvested Area in the Major Winter Wheat-Producing Regions of China from 2018 to 2022." *Scientific Data* 11: 1038. https://doi.org/10.1038/s41597-024-03867-z.

Li, J., and H. Lei. 2021. “Tracking the Spatio-Temporal Change of Planting Area of Winter Wheat-Summer Maize Cropping System in the North China Plain during 2001–2018.” *Computers and Electronics in Agriculture* 187: 106222. https://doi.org/10.1016/j.compag.2021.106222.

Li, J., L. He, J. Chen, and A. Plaza. 2020. “Spatio-Temporal Fusion for Remote Sensing Data: An Overview and New Benchmark.” *Science China Information Sciences* 63: 140301. https://doi.org/10.1007/s11432-019-2785-y.

Liu, X., X. Li, L. Gao, J. Zhang, D. Qin, and K. Wang. 2023. “Early-Season and Refined Mapping of Winter Wheat Based on Phenology algorithms-a Case of Shandong, China.” *Frontiers in Plant Science* 14: 1016890. https://doi.org/10.3389/fpls.2023.1016890.

Liu, X., H. Zhai, Y. Shen, B. Lou, C. Jiang, T. Li, and S. B. Hussain. 2020. “Large-Scale Crop Mapping from Multisource Remote Sensing Images in Google Earth Engine.” *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 13: 414–427. https://doi.org/10.1109/JSTARS.2019.2963539.

Luo, Y., Z. Zhang, Z. Li, Y. Chen, J. Cao, and F. Tao. 2020. “Identifying the Spatiotemporal Changes of Annual Harvesting Areas for Three Staple Crops in China By Integrating Multi-Data Sources.” *Environmental Research Letters*. 15: 074003. https://doi.org/10.1088/1748-9326/ab80f0.

Luo, L., S. Sun, J. Xue, Z. Gao, J. Zhao, Y. Yin, and X. Luan. 2023. “Crop Yield Estimation Based on Assimilation of Crop Models and Remote Sensing Data: a Systematic Evaluation.” *Agricultural Systems* 210: 103711. https://doi.org/10.1016/j.agsy.2023.103711.

N. A. S. A. *MODIS Data*. Available online: (accessed on 08 December 2024). https://modis.gsfc.nasa.gov/.

Nasrallah, A., N. Baghdadi, M. Mhawej, G. Faour, T. Darwish, H. Belhouchette, and S. Darwich. 2018. “A Novel Approach for Mapping Wheat Areas Using High Resolution Sentinel-2 Images.” *Sensors (Basel, Switzerland)* 18: 2089. https://doi.org/10.3390/s18072089.

Potapov, P., S. Turubanova, M. C. Hansen, A. Tyukavina, V. Zalles, A. Khan, X. Song, A. Pickens, Q. Shen, and J. Cortez. 2022. “Global Maps of Cropland Extent and Change Show Accelerated Cropland Expansion in the Twenty-First Century.” *Nature Food* 3: 19–28. https://doi.org/10.1038/s43016-021-00429-z.

Qu, C., P. Li, and C. Zhang. 2021. “A Spectral Index for Winter Wheat Mapping Using Multi-Temporal Landsat NDVI Data of Key Growth Stages.” *ISPRS Journal of Photogrammetry and Remote Sensing* 175: 431–447. https://doi.org/10.1016/j.isprsjprs.2021.03.015.

Ren, S., B. Guo, X. Wu, L. Zhang, M. Ji, and J. Wang. 2021. “Winter Wheat Planted Area Monitoring and Yield Modeling Using MODIS Data in the Huang-Huai-Hai Plain, China.” *Computers and Electronics in Agriculture* 182: 106049. https://doi.org/10.1016/j.compag.2021.106049.

Roznik, M., M. Boyd, and L. Porth. 2022. “Improving Crop Yield Estimation By Applying Higher Resolution Satellite NDVI Imagery and High-Resolution Cropland Masks.” *Remote Sensing Applications: Society and Environment* 25: 100693. https://doi.org/10.1016/j.rsase.2022.100693.

Ryu, J. -H., D. Oh, and J. Cho. 2021. “Simple Method for Extracting the Seasonal Signals of Photochemical Reflectance Index and Normalized Difference Vegetation Index Measured Using a Spectral Reflectance Sensor.” *Journal of Integrative Agriculture* 20: 1969–1986. https://doi.org/10.1016/S2095-3119(20)63410-4.

Sadeh, Y., X. Zhu, D. Dunkerley, J. P. Walker, Y. Zhang, O. Rozenstein, V. Manivasagam, and K. Chenu. 2021. “Fusion of Sentinel-2 and PlanetScope Time-Series Data into Daily 3 M Surface Reflectance and Wheat LAI Monitoring.” *International Journal of Applied Earth Observation and Geoinformation* 96: 102260. https://doi.org/10.1016/j.jag.2020.102260.

Sagan, V., M. Maimaitijiang, S. Bhadra, M. Maimaitiyiming, D. R. Brown, P. Sidike, and F. B. Fritschi. 2021. “Field-Scale Crop Yield Prediction Using Multi-Temporal WorldView-3 and PlanetScope Satellite Data and Deep Learning.” *ISPRS Journal of Photogrammetry and Remote Sensing* 174: 265–281. https://doi.org/10.1016/j.isprsjprs.2021.02.008.

Statistics, S. P. B. o. 2023. “The Shandong Statistical Yearbook.” *The Journal*. 2001.

Sun, Y., Y. Chang, J. Liu, X. Ge, and F. Chen. 2021. “Spatial Differentiation of Non-Grain Production on Cultivated Land and Its Driving Factors in Coastal China.” *Sustainability* 13: 13064. https://doi.org/10.3390/su132313064.

Tian, H., P. Wang, K. Tansey, J. Zhang, and H. Li. 2021. “An LSTM Neural Network for Improving Wheat Yield Estimates By Integrating Remote Sensing Data and Meteorological Data in the Guanzhong Plain, PR China.” *Agricultural and Forest Meteorology* 310: 108629. https://doi.org/10.1016/j.agrformet.2021.108629.

Uribeetxebarria, A., A. Castellón, and A. Aizpurua. 2023. “Optimizing Wheat Yield Prediction Integrating Data from Sentinel-1 and Sentinel-2 with CatBoost Algorithm.” *Remote Sensing* 15: 1640. https://doi.org/10.3390/rs15061640.

USGS. Landsat Missions. Available online: (accessed on 08 December 2024). https://www.USGS.gov/landsat-missions.

Vannoppen, A., and A. Gobin. 2021. “Estimating Farm Wheat Yields from NDVI and Meteorological Data.” *Agronomy* 11: 946. https://doi.org/10.3390/agronomy11050946.

Wang, Y., W. Shi, and T. Wen. 2023. “Prediction of Winter Wheat Yield and Dry Matter in North China Plain Using Machine Learning Algorithms for Optimal Water and Nitrogen Application.” *Agricultural Water Management* 277: 108140. https://doi.org/10.1016/j.agwat.2023.108140.

Wang, Q., F. Li, J. Yu, L. Fleskens, and C. J. Ritsema. 2021. “Price Decline, Land Rental Markets and Grain Production in the North China Plain.” *China Agricultural Economic Review* 13: 124–149. https://doi.org/10.1108/CAER-04-2019-0066.

Wang, S., J. Chen, M. Shen, T. Shi, L. Liu, L. Zhang, and Q. Dong. 2022. “Characterizing Spatiotemporal Patterns of Winter Wheat Phenology from 1981 to 2016 in North China By Improving Phenology Estimation.” *Remote Sensing* 14: 4930. https://doi.org/10.3390/rs14194930.

Wu, X., Y. Qi, Y. Shen, W. Yang, Y. Zhang, and A. Kondoh. 2019. "Change of Winter Wheat Planting Area and Its Impacts on Groundwater Depletion in the North China Plain." *Journal of Geographical Sciences* 29: 891–908. https://doi.org/10.1007/s11442-019-1635-9.

Xiao, G., X. Zhu, C. Hou, Y. Liu, and K. Xu. 2019. "A Spatialization Method for Grain Yield Statistical Data: A Study on Winter Wheat of Shandong Province, China." *Agronomy Journal* 111: 1892–1903. https://doi.org/10.2134/agronj2018.09.0555.

Xu, X., J. Liu, S. Zhang, R. Li, C. Yan, and S. Wu. 2018. "China's multi-period land use land cover remote sensing monitoring data set (CNLUCC)." *Resource and Environmental Data Cloud Platform (Beijing, China)*. https://doi.org/10.12078/2018070201.

Yang, J., and X. Huang. 2021. "The 30 M Annual Land Cover Dataset and Its Dynamics in China from 1990 to 2019." *Earth System Science Data*. 13: 1–29. https://doi.org/10.5194/essd-13-3907-2021.

Yang, G., X. Li, P. Liu, X. Yao, Y. Zhu, W. Cao, and T. Cheng. 2023. "Automated In-Season Mapping of Winter Wheat in China with Training Data Generation and Model Transfer." *Remote Sensing Applications: Society and Environment* 202: 422–438. https://doi.org/10.1016/j.isprsjprs.2023.07.004.

Ye, F., Z. Zhou, Y. Wu, and B. Enkhtur. 2022. "Application of Convolutional Neural Network in Fusion and Classification of Multi-Source Remote Sensing Data." *Frontiers in Neurorobotics*. 16: 1095717. https://doi.org/10.3389/fnbot.2022.1095717.

Yuan, Y., L. Lin, Q. Liu, R. Hang, and Z. Zhou. 2022. "SITS-Former: A Pre-Trained Spatio-Spectral-Temporal Representation Model for Sentinel-2 Time Series Classification." *International Journal of Applied Earth Observation and Geoinformation* 106: 102651. https://doi.org/10.1016/j.jag.2021.102651.

Zhai, W., C. Li, S. Fei, Y. Liu, F. Ding, Q. Cheng, and Z. Chen. 2023. "CatBoost Algorithm for Estimating Maize Above-Ground Biomass Using Unmanned Aerial Vehicle-Based Multi-Source Sensor Data and SPAD Values." *Computers and Electronics in Agriculture*. 214: 108306. https://doi.org/10.1016/j.compag.2023.108306.

Zhang, X., F. Qiu, and F. Qin. 2019. "Identification and Mapping of Winter Wheat By Integrating Temporal Change Information and Kullback–Leibler Divergence." *International Journal of Applied Earth Observation and Geoinformation*. 76: 26–39. https://doi.org/10.1016/j.jag.2018.11.002.

Zhang, W., M. Brandt, A. V. Prishchepov, Z. Li, C. Lyu, and R. Fensholt. 2021. "Mapping the Dynamics of Winter Wheat in the North China Plain from Dense Landsat Time Series (1999 to 2019)." *Remote Sensing* 13: 1170. https://doi.org/10.3390/rs13061170.

Zhang, Y., C. Xia, X. Cheng, G. Feng, Y. Wang, and Q. Gao. 2021. "Estimating the Maize Biomass By Crop Height and Narrowband Vegetation Indices Derived from UAV-based Hyperspectral Images." *Ecological Indicators* 129: 107985. https://doi.org/10.1016/j.ecolind.2021.107985.

Zhang, Y., X. Lu, B. Liu, D. Wu, G. Fu, Y. Zhao, and P. Sun. 2021. "Spatial Relationships between Ecosystem Services and Socioecological Drivers Across a Large-Scale Region: A Case Study in the Yellow River Basin." *Science of the Total Environment* 766: 142480. https://doi.org/10.1016/j.scitotenv.2020.142480.

Zhang, S., J. Zhang, Y. Bai, and F. Yao. 2018. "Extracting Winter Wheat Area in Huanghuaihai Plain Using MODIS-EVI Data and Phenology Difference Avoiding Threshold." *Transactions of the Chinese Society of Agricultural Engineering* 34: 150–158. https://doi.org/10.11975/j.issn.1002-6819.2018.11.019.

Zhao, X., N. Xia, Y. Xu, X. Huang, and M. Li. 2021. "Mapping Population Distribution Based on XGBoost Using Multisource Data." *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*. 14: 11567–11580. https://doi.org/10.1109/JSTARS.2021.3125197.

Zhao, Y., X. Wang, Y. Guo, X. Hou, and L. Dong. 2022. "Winter Wheat Phenology Variation and Its Response to Climate Change in Shandong Province, China." *Remote Sensing* 14: 4482. https://doi.org/10.3390/rs14184482.

Zhu, Z., Z. Dai, S. Li, and Y. Feng. 2022. "Spatiotemporal Evolution of Non-Grain Production of Cultivated Land and Its Underlying Factors in China." *INternational Journal Of Environmental Research And Public Health* 19: 8210. https://doi.org/10.3390/ijerph19138210.