



Article

Multi-Source Remote Sensing and Stacking Ensemble Learning for Vineyard Mapping and Inventory Updating in Arid Xinjiang

Leiting Yi ^{1,2}, Lei Wang ^{1,3,4,5,*}, Zhi Pu ^{2,*}, Lei Luo ¹, Siyu Zhou ² and Shipeng Wang ²

¹ Xinjiang Uygur Autonomous Region Academy of Forestry Sciences, Urumqi 830063, China; 13035095676@163.com (L.Y.); 13899809063@163.com (L.L.)

² College of Computer and Information Engineering, Xinjiang Agricultural University, Urumqi 830052, China; 19556711845@163.com (S.Z.); w3257516035@163.com (S.W.)

³ College of Geography and Remote Sensing Science, Xinjiang University, Urumqi 830046, China

⁴ Xinjiang Key Laboratory of Desertification Control and Sand Industry, Xinjiang Uygur Autonomous Region Academy of Forestry Sciences, Urumqi 830092, China

⁵ Xinjiang Key Laboratory of Oasis Ecology, Xinjiang University, Urumqi 830046, China

* Correspondence: wanglei4970@163.com (L.W.); puzhi6617914@sina.com (Z.P.)

Abstract

Accurate vineyard mapping is important for agricultural resource monitoring, land-use management, and inventory updating in arid regions. However, vineyard identification in Xinjiang, China, is challenged by fragmented parcels, exposed soil backgrounds, irrigation-driven heterogeneity, and strong accumulated-temperature gradients. This study developed a multi-source remote-sensing framework for vineyard mapping and inventory updating by integrating Sentinel-1/2 data, terrain variables, growing-degree-day-derived agrothermal zones (ATZs), and seasonal-difference features. RF, LightGBM, XGBoost, 1D-CNN, and a Stacking ensemble were evaluated using polygon-level in-distribution testing and Leave-One-ATZ-Out cross-zone validation. The in-distribution test was used to assess vineyard separability under similar sample distributions, whereas cross-ATZ validation was used to evaluate model transferability across heterogeneous thermal domains. Tree-based models and Stacking achieved near-ceiling performance under the in-distribution setting, but cross-ATZ validation revealed substantial performance degradation, indicating that conventional local validation can overestimate operational transferability. RF achieved the highest mean cross-zone F1-score, while Stacking achieved the highest cross-zone AP and Recall and provided a flexible probability surface for thresholding, mosaicking, vector post-processing, and patch-level mapping. In Gaochang District, the Stacking-derived result identified 22,998.84 ha of potential vineyard area, achieved an Area Recall of 0.7908 against the historical inventory, and covered 88.03% of existing parcels at $\geq 30\%$ spatial overlap. The workflow also identified 1356 candidate vineyard patches for inventory updating covering 2502.32 ha for subsequent inventory verification. These results demonstrate that multi-source feature integration and probability-based ensemble mapping can support vineyard mapping and inventory updating in arid regions, while cross-zone validation is essential for assessing operational generalization.

Keywords: vineyard mapping; Sentinel-1/2; stacking ensemble; spatial generalization; oasis agriculture



Academic Editor: Murali Krishna Gumma

Received: 14 June 2026

Revised: 16 July 2026

Accepted: 21 July 2026

Published: 30 July 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

Vineyards are high-value perennial crop systems in arid oasis agriculture, and accurate vineyard inventories are essential for agricultural planning, water allocation, and

climate-adaptation management. Xinjiang is one of the most important grape-growing regions in arid northwestern China, where vineyards are closely linked to oasis agriculture, rural livelihoods, specialty fruit production, and regional land-use management. However, traditional inventory updating based on field surveys and historical records is increasingly insufficient under rapid land-use change, intensifying water-resource constraints, and spatially fragmented agricultural patterns. Precision viticulture increasingly relies on remote sensing, geospatial analysis, and machine learning to characterize vineyard spatial heterogeneity and support inventory updating, irrigation management, and growth monitoring [1–3]. In Xinjiang, vineyard mapping is complicated by diverse canopy structures, exposed soil backgrounds, irrigation conditions, and regional thermal gradients, requiring a mapping approach suitable for complex arid oasis landscapes.

Sentinel-2 has become a core data source for crop classification [4], as its multi-temporal observations effectively capture seasonal spectral variations [5,6]. Phenology metrics derived from Sentinel-2 time series can distinguish vineyards from other tree crops [7], and vineyard mapping studies have increasingly shifted toward multi-source feature integration [8]. In addition, vine growth and spectral variability can be affected by topographic conditions, supporting the inclusion of terrain variables in vineyard monitoring [9]. Although UAV and satellite data have been used for canopy assessment, leaf area index estimation, vine detection, and vineyard zoning [10–12], UAV observations are difficult to scale to regional monitoring, and single-source optical data are limited by cloud contamination, exposed soil backgrounds, and crop spectral similarity.

Multi-sensor fusion offers a practical solution by integrating complementary information from different observation systems [13]. Combining Sentinel-1 SAR and Sentinel-2 optical data is particularly useful in heterogeneous oasis agriculture, because SAR backscatter provides structural and moisture-related information that complements optical reflectance. Radar-optical fusion has been shown to improve crop classification [14,15] and support automated large-scale mapping through cloud platforms such as Google Earth Engine [16,17].

Sentinel-1/2 data fusion and machine-learning approaches have already shown value for crop mapping in Xinjiang, such as cotton-field mapping in northern Xinjiang [18]. Recent multi-year crop datasets based on harmonized Landsat and Sentinel-2 observations have also improved agricultural monitoring in Xinjiang [19]. However, vineyard-specific mapping remains less developed, particularly for fragmented oasis vineyards under strong agrothermal gradients.

Grapevine phenology exhibits distinct leaf-on and leaf-off variations driven by temperature [20], climate shifts [21], and irrigation management [22]. Phenology-based metrics are widely used in agricultural mapping [23–26], and integrating spectrottemporal signatures with accumulated temperature can better link crop development with thermal conditions [27]. This is especially relevant for Xinjiang, where pronounced regional thermal gradients may alter vineyard spectral and structural signals across zones.

Machine learning has advanced crop mapping, but supervised classification remains sensitive to sample design, spatial autocorrelation, and validation strategy [28]. RF, SVM, gradient boosting, and related algorithms have been widely applied to regional crop mapping [29–32], often supported by Google Earth Engine for large-scale processing. Deep learning models, including CNNs, LSTMs, and U-Nets, are increasingly used for multi-temporal crop classification and segmentation [33–38], but their dependence on large samples and structured temporal inputs, as well as their sensitivity to domain transfer, can limit operational use [39,40]. Tree-based ensemble models therefore remain attractive for large-area vineyard mapping because of their stability, interpretability, and deployment efficiency.

Spatial generalizability remains a critical challenge in dynamic crop mapping [41]. Conventional random or spatially close data splits can overestimate accuracy, especially when training and test samples share similar environmental backgrounds. For Xinjiang vineyards, thermal zones differ in climate, terrain, irrigation, and planting structure, making cross-region validation necessary for assessing model transferability.

Several research gaps remain. Existing vineyard studies have often focused on local or region-specific mapping rather than large-area arid-oasis applications [42]. Spectral, textural, and seasonal feature representations for trellised vineyards still require further refinement [43]. Generalization across thermal domains is rarely evaluated. Moreover, converting pixel-level classification outputs into patch- or parcel-level products requires systematic post-processing and inventory comparison for practical updating.

To address these gaps, this study proposes a multi-source Stacking ensemble framework for vineyard mapping in arid Xinjiang. The method integrates Sentinel-1/2 data, terrain variables, thermal zones, and seasonal-difference features. RF, LGBM, and XGB are used as base learners, and their probability outputs are fused by a Stacking model. Polygon-level splitting is adopted to reduce spatial autocorrelation, while Leave-One-ATZ-Out validation is used to assess cross-thermal-zone transferability.

The main contributions of this study are threefold. First, a multi-source and multi-season feature framework is developed for vineyard identification in arid Xinjiang by integrating optical, SAR, terrain, agrothermal, and seasonal-difference information. Second, a Stacking ensemble is evaluated together with RF, LGBM, XGB, and 1D-CNN under both polygon-level in-distribution testing and Leave-One-ATZ-Out cross-zone validation, allowing local separability and cross-zone transferability to be assessed separately. Third, an operational mapping workflow is demonstrated in Gaochang District by converting pixel-level probability outputs into vineyard patch products and comparing them with an existing inventory to evaluate spatial consistency and identify candidate vineyard patches requiring verification.

2. Materials and Methods

2.1. Study Area

This study selected Xinjiang Uygur Autonomous Region, China, as the main study area, with Gaochang District of Turpan City used as a representative case area for vineyard mapping and inventory comparison. Xinjiang is located in the arid and semi-arid region of northwestern China, where oasis agriculture is mainly distributed along basin margins, alluvial plains, and low-elevation areas with irrigation access. Vineyards are often interspersed with cotton fields, maize fields, orchards, shelterbelts, bare land, and built-up areas. Spectral confusion, parcel fragmentation, and pronounced regional thermal gradients therefore pose major challenges for remote-sensing-based vineyard identification.

To characterize thermal heterogeneity within Xinjiang and evaluate cross-region model transferability, agrothermal zones (ATZs) were introduced as both environmental stratification units and generalization-validation domains. Gaochang District, located in the central Turpan Basin, is a major grape-growing area with a concentrated vineyard distribution and an existing historical vineyard inventory. It was therefore selected for final spatial mapping and inventory-consistency analysis. The spatial location of Xinjiang, the ATZ distribution, the reference sample distribution, and the location of Gaochang District are shown in Figure 1.

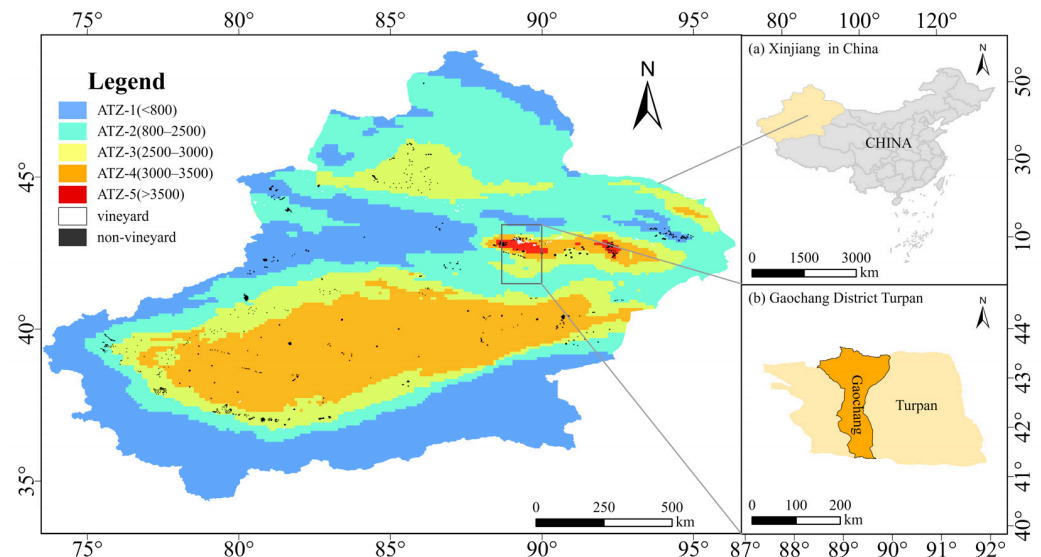


Figure 1. Study area, agrothermal zones, reference sample distribution, and the location of Gaochang District. (a) Xinjiang in China; (b) Gaochang District Turpan.

In this region, vineyard distribution is strongly constrained by the interaction between thermal suitability, terrain background, and irrigation accessibility. Grape cultivation is mainly concentrated in low-elevation oasis areas and basin-margin agricultural zones, where accumulated temperature is sufficient for grapevine development and irrigation infrastructure supports perennial crop growth. In contrast, high-elevation mountainous areas, bare desert surfaces, saline or sparsely vegetated land, and built-up areas are generally unsuitable for vineyard cultivation. These environmental contrasts make Xinjiang an appropriate region for evaluating whether vineyard-mapping models can generalize across agrothermal gradients rather than only within locally similar sample distributions.

In the reference dataset, ATZ1 represented the lowest accumulated-temperature domain and contained no positive vineyard samples; therefore, it was used to examine false-positive tendencies in cold or thermally insufficient regions but was excluded from mean binary F1-score and AP summaries in cross-ATZ validation.

2.2. Data and Preprocessing

Reference samples were constructed from historical vineyard and non-vineyard vector data provided by the Xinjiang Uygur Autonomous Region Academy of Forestry Sciences and were manually checked and redrawn using high-resolution imagery in Google Earth (web platform; <https://earth.google.com/> (accessed on 13 June 2026)) and Google Maps (web platform; <https://maps.google.com/> (accessed on 13 June 2026)). The interpretation imagery dates varied spatially, and exact image dates were not uniformly available for all polygons; therefore, the labels were treated as visually verified reference data rather than field-survey ground truth. During refinement, polygon boundaries were adjusted according to vineyard canopy texture, row-like planting patterns, surrounding land-cover context, and multi-temporal image evidence. To reduce mixed boundary pixels, all polygons were inward-buffered by 5 m before pixel sampling, and extremely large non-vineyard background polygons (>300,000 m²) were removed. The Google Earth Engine (GEE; web platform; <https://earthengine.google.com/> (accessed on 13 June 2026)) workflow exported a unified sample set, whereas deduplication and train/validation/test splitting were performed locally. Because the samples were visually verified rather than field surveyed, possible temporal mismatch between interpretation imagery and the 2020–2024 feature composites remains a source of uncertainty.

To avoid accuracy overestimation caused by spatial autocorrelation, model training and validation were strictly conducted at the polygon level, such that all pixels within the same polygon were assigned to only one of the training, validation, or test sets. Attribute fields such as polygon ID, tile ID, source, area_m2, geometry, system:index, and the lower-case sample-level atz field were used only for sample management, spatial grouping, and result tracking, and were not used as model input features. Model inputs were derived only from Sentinel-2, Sentinel-1, DEM, ERA5-Land, and ATZ-related numerical variables.

This study integrated Sentinel-2 optical features, Sentinel-1 SAR features, DEM-derived terrain variables, and ERA5-Land temperature data. Sentinel-2 surface-reflectance data were obtained from COPERNICUS/S2_SR_HARMONIZED, masked using QA60 cloud and cirrus bits, and scaled to 0–1 reflectance. Sentinel-1 features were derived from COPERNICUS/S1_GRD IW-mode VV and VH observations. DEM features were derived from USGS/SRTMGL1_003, and GDD was calculated from ECMWF/ERA5_LAND/DAILY_AGGR temperature_2m. All feature layers were sampled or exported at a 10 m scale in GEE; no additional offline interpolation was applied beyond GEE projection and sampling during feature extraction. The historical vineyard inventory of Gaochang District was not used for model training; it was used only for post-mapping spatial consistency and inventory-update assessment. The main data sources and preprocessing settings are summarized in Table 1.

Table 1. Data sources and key preprocessing settings used in this study.

Data Source	Product or Variable	Resolution/Format	Temporal Range	Main Use
Sentinel-2 MSI	COPERNICUS/S2_SR_HARMONIZED; B2-B8A, B11, B12; NDVI, EVI, NDWI, NDMI, SAVI	10–20 m; sampled/exported at 10 m	2020–2024	Optical and phenological features
Sentinel-1 SAR	COPERNICUS/S1_GRD; IW mode; VV, VH, and VV-VH	10 m	2020–2024	Radar structural features
DEM	USGS/SRTMGL1_003	30 m; sampled/exported at 10 m	Static	DEM, slope, and aspect
ERA5-Land	ECMWF/ERA5_LAND/DAILY_AGGR; temperature_2m	~10 km; exported as ATZ at 10 km and sampled with feature layers	2020–2024	GDD and agrothermal zones
Reference samples	Institutional vineyard/non-vineyard vectors refined using Google Earth/Maps imagery	Vector	Imagery dates varied spatially; exact dates not uniformly available	Model training and validation
Existing inventory	Historical vineyard parcels in Gaochang District	Vector	Historical inventory; exact update dates not uniformly available	Spatial consistency assessment

2.3. Multi-Source Feature Construction and Stacking Classification Framework

To improve separability between vineyards and other oasis agricultural land-cover types, this study constructed leaf-on, leaf-off, and selected seasonal-difference features to characterize spectral, structural, and moisture-related changes between the growing and non-growing seasons. Leaf-on and leaf-off composites were constructed for Sentinel-2 optical bands, vegetation indices, and Sentinel-1 SAR metrics, whereas seasonal-difference features were generated for vegetation indices, Sentinel-1 SAR metrics, and the Sentinel-2 SWIR bands B11 and B12. This design helps represent the relationship between crop development and thermal background. For a feature X with both seasonal composites, the seasonal-difference feature was defined according to Equation (1):

$$X_{\text{diff}} = X_{\text{leaf-on}} - X_{\text{leaf-off}} \quad (1)$$

where $X_{\text{leaf-on}}$ and $X_{\text{leaf-off}}$ denote the corresponding feature values derived from the leaf-on and leaf-off seasonal composites, respectively.

Multi-year compositing was used to represent persistent vineyard conditions during 2020–2024. For both Sentinel-2 and Sentinel-1, monthly median composites were first generated for each month within each year. Monthly composites within the leaf-on and leaf-off windows were then aggregated into annual seasonal composites using the median, and the annual seasonal composites from 2020 to 2024 were further aggregated using the median. This hierarchical median compositing reduced the influence of residual clouds, abnormal observations, year-specific bare-soil exposure, canopy gaps, and management disturbances.

A fixed-calendar strategy was used for operational reproducibility: May–August was defined as the leaf-on window, and April plus September–November were defined as the leaf-off/background window. Seasonal-difference features were calculated as leaf-on minus leaf-off. Although grapevine phenology may vary across ATZs, the Leave-One-ATZ-Out validation was used to evaluate the transferability of this fixed-window representation across thermal zones.

Sentinel-2 features included commonly used optical bands and standard indices, including NDVI, EVI, NDWI, NDMI, and SAVI. Sentinel-1 features were derived from the VV and VH backscatter coefficients provided by the GEE Sentinel-1 GRD product, denoted as VV, VH, and their difference VV-VH. No additional terrain normalization was performed. Previous studies have shown that Sentinel-1/Sentinel-2 fusion and time-series compositing can enhance crop identification in complex agricultural landscapes.

To characterize thermal variation within Xinjiang, annual GDD was accumulated for each year from 2020 to 2024 using ECMWF/ERA5_LAND/DAILY_AGGR daily 2 m air temperature, as defined in Equation (2). For each calendar year, the calculation period was 1 January to 31 December. Temperature was converted from Kelvin to Celsius, and daily GDD was calculated as $\max(T_{\text{mean},d} - 10, 0)$. The 2020–2024 mean annual GDD was classified into five non-overlapping ATZ classes: ATZ1, $\text{GDD} < 800$; ATZ2, $800 \leq \text{GDD} < 2500$; ATZ3, $\text{GDD} \geq 2500$ and < 3000 ; ATZ4, $\text{GDD} \geq 3000$ and < 3500 ; and ATZ5, $\text{GDD} \geq 3500$ degree C days.

$$\text{GDD} = \sum_{d=1}^n \max(T_{\text{mean},d} - T_{\text{base}}, 0) \quad (2)$$

where $T_{\text{mean},d}$ is the daily mean temperature on day d , T_b is the base temperature, which was set to 10 °C in this study, and n is the number of days in the calculation period. In total, 50 numerical input features were constructed: ASPECT, ATZ, B11_DIFF, B11_LEAFOFF, B11_LEAFON, B12_DIFF, B12_LEAFOFF, B12_LEAFON, B2_LEAFOFF, B2_LEAFON, B3_LEAFOFF, B3_LEAFON, B4_LEAFOFF, B4_LEAFON, B5_LEAFOFF, B5_LEAFON, B6_LEAFOFF, B6_LEAFON, B7_LEAFOFF, B7_LEAFON, B8A_LEAFOFF, B8A_LEAFON, B8_LEAFOFF, B8_LEAFON, DEM, EVI_DIFF, EVI_LEAFOFF, EVI_LEAFON, NDMI_DIFF, NDMI_LEAFOFF, NDMI_LEAFON, NDVI_DIFF, NDVI_LEAFOFF, NDVI_LEAFON, NDWI_DIFF, NDWI_LEAFOFF, NDWI_LEAFON, SAVI_DIFF, SAVI_LEAFOFF, SAVI_LEAFON, SLOPE, VH_DIFF, VH_LEAFOFF, VH_LEAFON, VV_DIFF, VV_LEAFOFF, VV_LEAFON, VV_VH_SEASONAL_DIFF, VV_VH_LEAFOFF, VV_VH_LEAFON. The overall multi-source and multi-season feature construction workflow is illustrated in Figure 2.

The classification models included RF, LGBM, XGB, 1D-CNN, and the Stacking ensemble. RF, LGBM, and XGB were used as base learners in the Stacking model, while 1D-CNN was used only as a neural-network baseline under the same tabular input conditions and was not intended to represent a full patch-based or temporal CNN architecture. Previous studies have shown that ensemble machine-learning models and multi-temporal remote-sensing features are suitable for crop classification in complex agricultural regions [44]. Stacking was selected as the main operational model because it can fuse probability outputs from different tree-based models and provide a flexible probability surface for thresholding, mosaicking, raster-to-vector conversion, and patch-level post-processing.

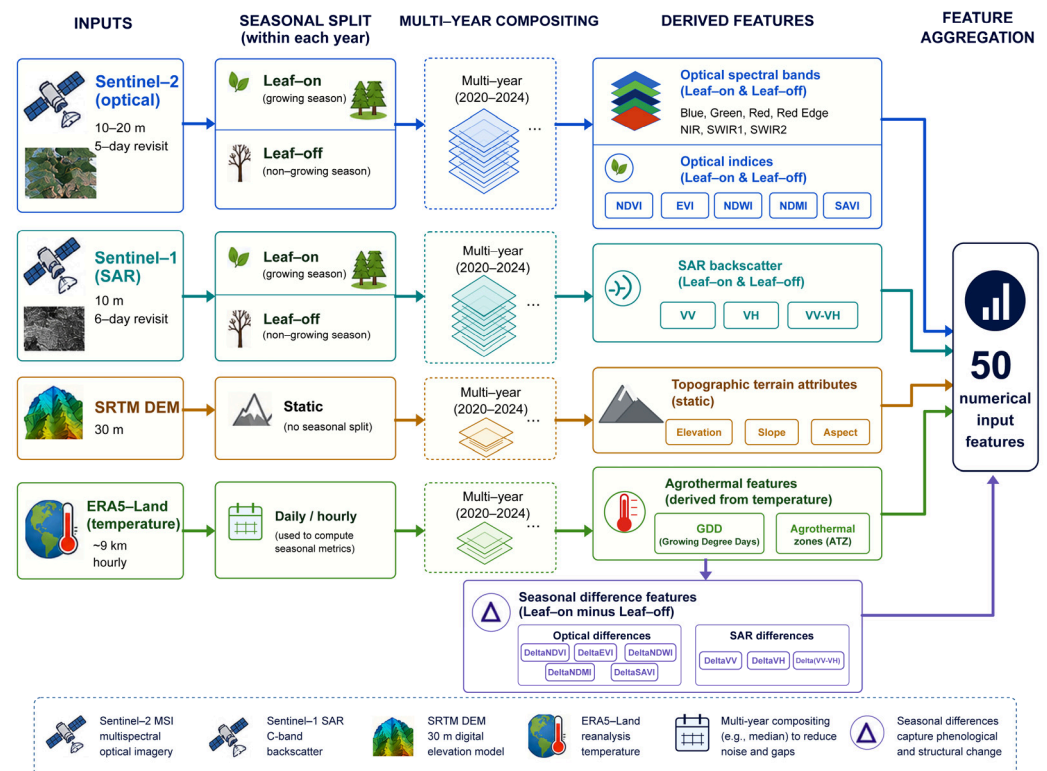


Figure 2. Multi-source and multi-season feature construction framework. Arrows indicate the processing flow, while blue, cyan, orange, green, and purple distinguish Sentinel-2, Sentinel-1, terrain, agrothermal, and seasonal-difference features, respectively.

The first-level models generated vineyard probabilities, denoted as p_{RF} , p_{LGBM} , and p_{XGB} . These probabilities and their statistical combinations were used to construct the meta-feature vector $\mathbf{z}$, as shown in Equation (3):

$$\mathbf{z} = [p_{RF}, p_{LGBM}, p_{XGB}, \mu_p, \max(p), \min(p), \sigma_p, \Delta p] \quad (3)$$

A Logistic Regression meta learner was then used to output the final Stacking probability according to Equation (4):

$$p_{stack} = f_{meta}(\mathbf{z}) \quad (4)$$

where μ_p and σ_p denote the mean and standard deviation of the three base probabilities, respectively, and Δp denotes the pairwise probability–difference combinations among the base learners, including $(p_{RF} - p_{LGBM})$, $(p_{RF} - p_{XGB})$, and $(p_{LGBM} - p_{XGB})$. The final class label was determined using a threshold selected on the validation set.

To avoid using the validation set for both meta-learner training and threshold selection, an out-of-fold stacking protocol was adopted. Specifically, within the training set, five-fold GroupKFold cross-validation was conducted using polygon ID as the grouping unit. For each fold, the base learners were trained on four folds and generated OOF base probabilities for the held-out fold. The OOF probabilities from all training samples were then combined to train the Logistic Regression meta learner. Subsequently, the base learners were retrained using the full training set and used to generate base probabilities for the validation and test sets; the trained meta learner then converted these probabilities into final Stacking probabilities. The validation set was used only to select the F1-optimal threshold, and the test set was used only for final performance evaluation. In the Leave-One-ATZ-Out experiment, the held-out ATZ was used only for final evaluation and did not participate in base-learner training, meta-learner fitting, or threshold selection. The OOF training strategy

and probability-fusion structure of the Stacking ensemble are shown in Figure 3, and the main model configurations and hyperparameters are listed in Table 2.

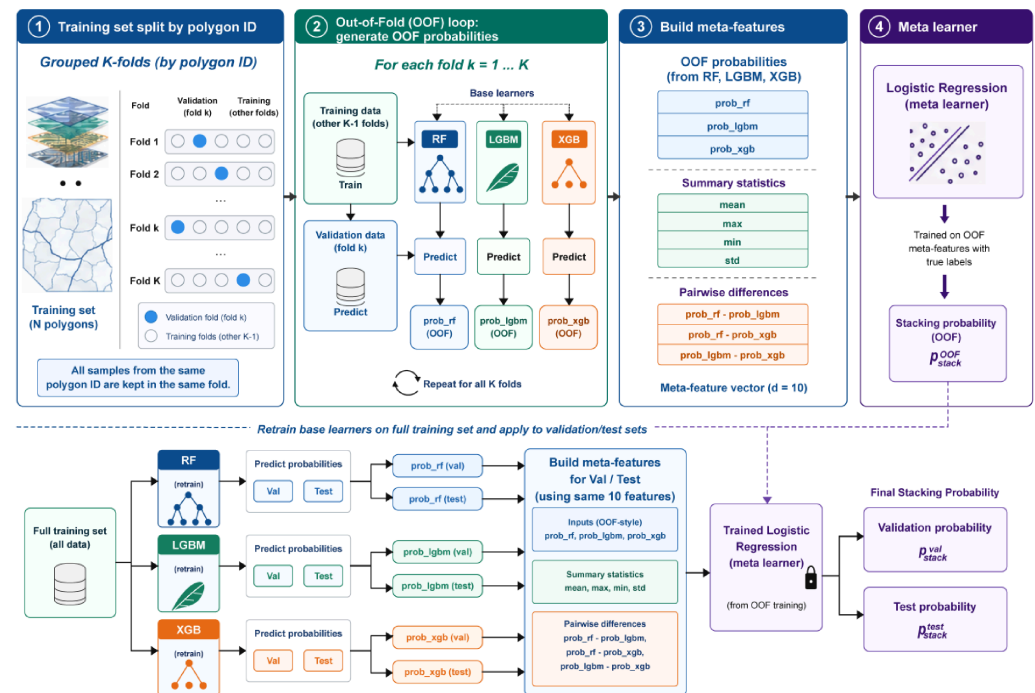


Figure 3. OOF training and probability-fusion framework of the Stacking ensemble. Arrows indicate the probability flow, while blue, green, orange, and purple distinguish RF, LGBM, XGB, and the meta-learner, respectively.

Table 2. Model configurations and key hyperparameters.

Component	Model	Main Hyperparameters	Role
Base learner	RF	n_estimators = 400; max_depth = None; class_weight = balanced_subsample; random_state = 2025	Base probability generation
Base learner	LGBM	n_estimators = 1500; learning_rate = 0.03; num_leaves = 127; subsample = 0.8; colsample_bytree = 0.8; reg_lambda = 1.0; scale_pos_weight = Nneg/Npos	Base probability generation
Base learner	XGB	n_estimators = 800; max_depth = 8; learning_rate = 0.05; subsample = 0.8; colsample_bytree = 0.8; reg_lambda = 1.0; tree_method = hist; scale_pos_weight = Nneg/Npos	Base probability generation
Meta learner	Logistic Regression	C = 1.0; class_weight = balanced; max_iter = 2000; random_state = 2025	Stacking probability fusion

Table 2. *Cont.*

Component	Model	Main Hyperparameters	Role
Neural baseline	1D-CNN	Conv1D blocks = 3 (64 channels; kernel sizes = 5, 3, and 3); BatchNorm + ReLU after each convolution; global pooling + dense output layer; dropout = 0.2; optimizer = Adam; loss = binary cross-entropy with logits; batch size = 2048; learning rate = 0.001; max epochs = 20; early-stopping patience = 4 based on validation AP	Same-input neural-network baseline

The model hyperparameters were not optimized through an exhaustive grid search. Instead, they were selected based on preliminary experiments, commonly used settings in crop-classification studies, class-imbalance considerations, and computational feasibility for large pixel-level datasets. For RF, a relatively large number of trees and balanced subsampling were used to improve stability under class imbalance. For LGBM and XGB, learning rate, tree complexity, subsampling, column sampling, regularization, and positive-class weighting were configured to balance fitting capacity and overfitting risk. The 1D-CNN baseline used the same 50-feature tabular vector after training-set median imputation and standardization. Its architecture consisted of three Conv1D blocks with 64 channels and kernel sizes of 5, 3, and 3, each followed by BatchNorm and ReLU activation, with dropout = 0.2, global pooling, and a dense output layer. It was trained using Adam optimization, binary cross-entropy with logits loss, a batch size of 2048, a learning rate of 0.001, a maximum of 20 epochs, and early stopping based on validation AP with a patience of four epochs.

All local analyses were conducted in Python 3.11.15 within a Conda 24.11.3 environment using JupyterLab 4.2.5 and PyCharm 2025.1. Core packages included NumPy 1.26.4, pandas 2.2.3, SciPy 1.13.1, scikit-learn 1.5.2, XGBoost 2.1.4, LightGBM 4.5.0, PyTorch 2.2.2+cpu, GeoPandas 1.1.3, GDAL 3.9.3, Rasterio 1.4.1, Shapely 2.1.1, pyproj 3.7.1, Matplotlib 3.9.2, and Seaborn 0.13.2.

2.4. Validation Strategy and Evaluation Metrics

To avoid spatial leakage, a polygon-level data split was adopted, ensuring that pixels from the same parcel were assigned to only one subset. In the conventional in-distribution test, the training set was used for model fitting, the validation set for threshold selection and early stopping, and the test set only for final evaluation. This test was used primarily as a baseline assessment of vineyard separability under similar sample distributions, rather than as the sole indicator of operational mapping performance. To further assess transferability under climate-related domain shifts, Leave-One-ATZ-Out validation was conducted. In each iteration, one ATZ was held out as the test domain, while the non-held-out polygons were further divided into training and internal validation subsets at the polygon level. The F1-optimal threshold was selected only on the internal validation subset. The held-out ATZ was never used to select the probability threshold. Because ATZ1 contained no positive vineyard samples, its F1-score and AP were excluded from the mean cross-zone generalization metrics and were used only to inspect false-positive tendencies in a no-positive-sample region. Spatial generalization is a key issue in large-area crop mapping,

and conventional in-distribution validation cannot fully represent model extrapolation ability in unseen regions. The complete validation design is summarized in Figure 4.

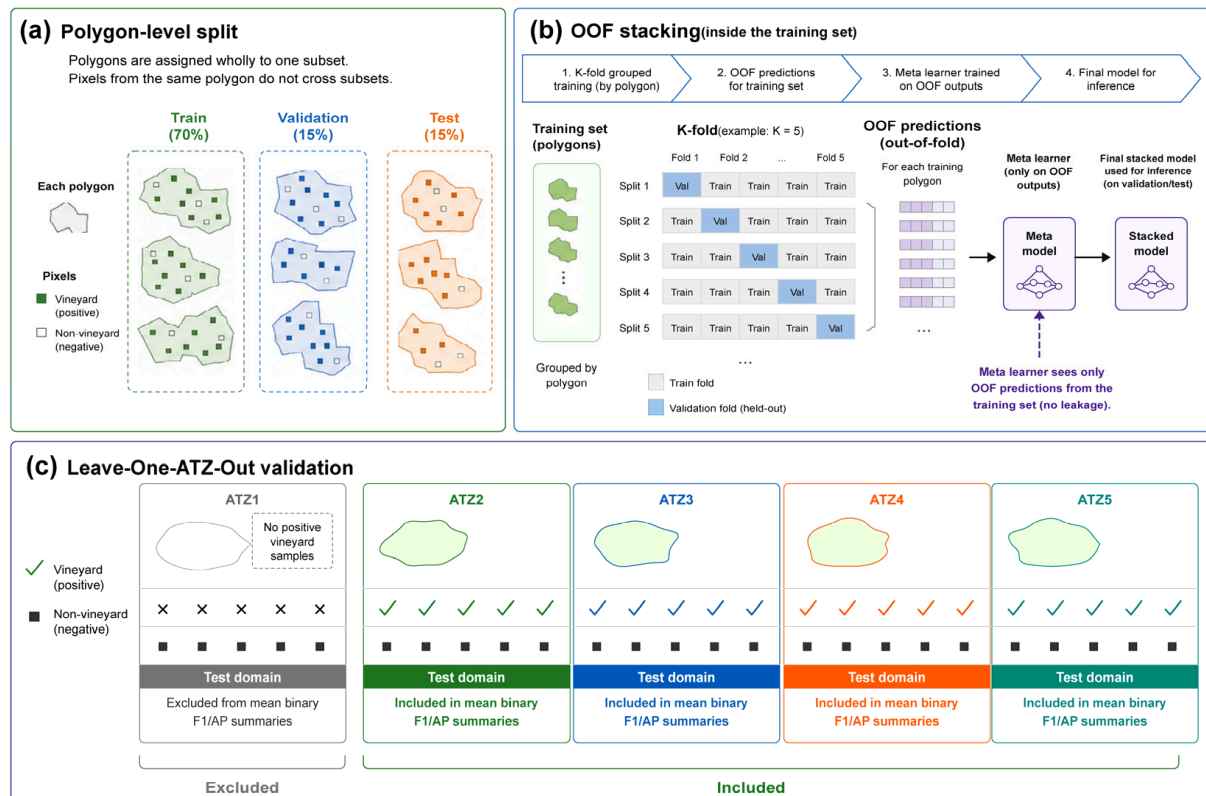


Figure 4. Validation design including (a) polygon-level split; (b) OOF Stacking; and (c) Leave-One-ATZ-Out validation.

Classification performance was evaluated using Precision, Recall, F1-score, Average Precision, Overall Accuracy, and Cohen's Kappa. Precision represents the proportion of predicted vineyard pixels that were true vineyard pixels, while Recall represents the proportion of true vineyard pixels correctly identified by the model. F1-score, the harmonic mean of Precision and Recall, was used for threshold selection and balanced binary-classification evaluation, as defined in Equation (5):

$$F1 = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (5)$$

In addition, a multi-seed stability experiment was conducted to assess the influence of random data splitting on model performance. Five repeated polygon-level splits were generated using random seeds 42, 52, 62, 72, and 82, and the mean and standard deviation of the main metrics were reported.

2.5. Gaochang District Spatial Inference, Post-Processing, and Inventory Comparison

Vineyard mapping in Gaochang District was performed using tile-based inference. For each tile, the feature image was matched and ordered according to the feature columns saved during model training. Missing values were filled using the corresponding feature medians from the training set. During inference, a rule-based candidate mask was used to exclude areas obviously unsuitable for vineyard cultivation. Candidate pixels were retained only when $\text{SLOPE} \leq 15^\circ$, $\text{DEM} \leq 2200 \text{ m}$, $\text{NDVI_LEAFON} \geq 0.15$, $\text{SAVI_LEAFON} \geq 0.05$, and $\text{NDVI_DIFF} \geq 0.00$. Median filling maintained consistency with model training but may introduce local bias in environmentally different tiles. Raster-to-vector post-processing

removed connected components smaller than 16 pixels, filled holes smaller than 800 m², applied a 5 m inward buffer, repaired geometries, and removed polygons smaller than 2500 m² in a local UTM projection.

For each tile, RF, LGBM, and XGB first generated base probabilities, and the meta learner then produced pstack. For Gaochang mapping, the Stacking probability threshold was treated as an operational binarization parameter. To examine its influence, threshold sensitivity was evaluated at 0.85, 0.90, 0.94, and 0.95 using mapped area, patch count, area recall, and area precision (Table S3). All four thresholds were evaluated using the identical candidate mask and the same raster-to-vector post-processing parameters, so that the reported differences reflect the probability-threshold setting rather than changes in vector processing. Probability maps and binary maps from all tiles were mosaicked separately, and a maximum-probability merging strategy was applied in overlapping regions to reduce omission errors along tile boundaries. After binarization, raster-to-polygon conversion and vector post-processing were performed [45]. These operations were used to reduce isolated noise, fill small internal gaps, and suppress boundary over-expansion caused by pixel-level classification. Adjacent predicted pixels were allowed to form continuous vineyard patches, but final dissolve was disabled to avoid excessive merging of spatially distinct objects. Therefore, the final output is more appropriately interpreted as a vineyard patch-level inventory product rather than a strict cadastral parcel-boundary delineation. The Gaochang District inference, post-processing, and inventory comparison workflow is shown in Figure 5.

The newly mapped vineyard patches were compared with the existing vineyard inventory through spatial overlay. For the i -th parcel in the existing inventory, its coverage by the new mapping result was defined according to Equation (6):

$$R_i = \frac{A(P_i^{\text{old}} \cap P^{\text{new}})}{A(P_i^{\text{old}})} \quad (6)$$

where O_i is the i -th parcel in the existing inventory, P^{new} is the spatial union of the newly mapped vineyard patches, and A denotes area. For area-level metrics, O_{all} denotes the spatial union of all existing inventory parcels. Based on R_i , Parcel Recall@30, Parcel Recall@50, and Parcel Recall@70 were calculated as the proportions of historical inventory parcels with $R_i \geq 0.30$, $R_i \geq 0.50$, and $R_i \geq 0.70$, respectively. These metrics measure the degree to which existing inventory parcels were spatially covered by the newly mapped vineyard patches, rather than the geometric boundary accuracy of individual cadastral parcels.

For each predicted patch P_j , its overlap ratio with the historical inventory was calculated as the intersection area between P_j and historical parcels divided by the area of P_j . A predicted patch was labelled as a candidate vineyard patch for inventory updating only when its maximum IoU with historical parcels was <0.10 and its overlap ratio with the historical inventory was <0.20 . The reported candidate-patch area refers to the entire area of these low-overlap predicted patches, rather than only the portion located outside the historical inventory.

Area-level consistency was evaluated using Area Recall, Area Precision, IoU, and the Dice coefficient. Area Recall and Area Precision were defined as the intersection area divided by the existing inventory area and the newly mapped area, respectively. IoU and Dice were defined according to Equations (7) and (8):

$$\text{IoU} = \frac{A(P^{\text{old}} \cap P^{\text{new}})}{A(P^{\text{old}} \cup P^{\text{new}})} \quad (7)$$

$$\text{Dice} = \frac{2A(p^{\text{old}} \cap p^{\text{new}})}{A(p^{\text{old}}) + A(p^{\text{new}})} \quad (8)$$

Because the existing inventory may contain temporal differences, differences in boundary standards, and updating delays, this comparison was interpreted as an assessment of spatial consistency and inventory-updating potential rather than a strict independent ground-truth validation.

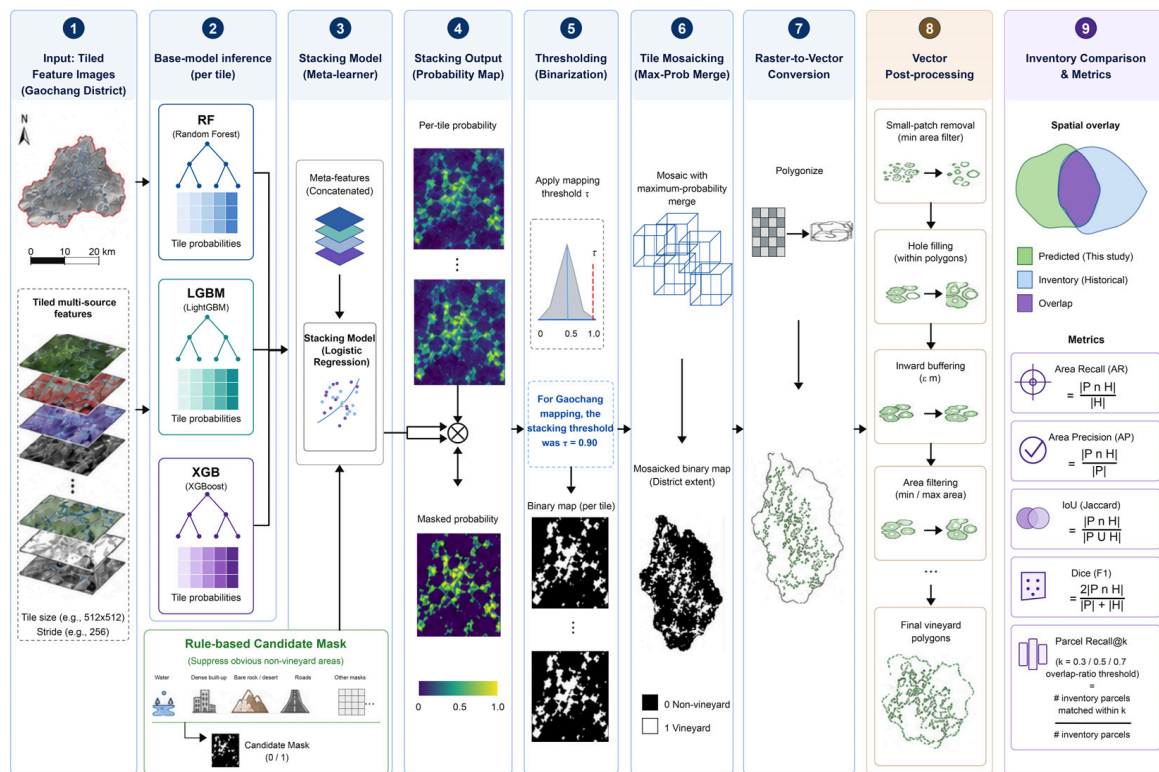


Figure 5. Workflow of Stacking-based inference, vector post-processing, and inventory comparison in Gaochang District.

3. Results

3.1. Dataset Composition and Split Characteristics

After multi-source feature extraction and polygon-level sample organization, the reference-polygon pool contained 4884 polygons. Following inward buffering, image masking, feature extraction, and local filtering, 1,915,776 valid pixel samples were retained for modelling. Among them, vineyard samples accounted for 334,628 pixels, while non-vineyard samples accounted for 1,581,148 pixels, corresponding to a vineyard proportion of 17.47%. All samples contained 50 numerical input features derived from Sentinel-2, Sentinel-1, terrain variables, and ATZ-related information.

The dataset was split at the polygon level into training, validation, and test subsets. The training set contained 1,319,979 pixels from 3419 polygons, the validation set contained 304,326 pixels from 732 polygons, and the test set contained 291,471 pixels from 733 polygons. These three subsets together accounted for all 4884 reference polygons used in this study. No polygon ID overlap was observed among the three subsets, confirming that pixels from the same reference polygon were not shared across training, validation, and test sets. This design reduced the risk of overly optimistic accuracy caused by within-parcel spatial autocorrelation.

The vineyard and non-vineyard samples were unevenly distributed across ATZs. Vineyard samples were mainly concentrated in ATZ2–ATZ5, while ATZ1 contained no positive vineyard samples. Therefore, ATZ1 was not included in the calculation of binary F1-score and AP averages in the Leave-One-ATZ-Out evaluation, and was used only to inspect false-positive tendencies in a no-positive-sample region. The distribution of samples across classes, data splits, and ATZs is shown in Figure 6.

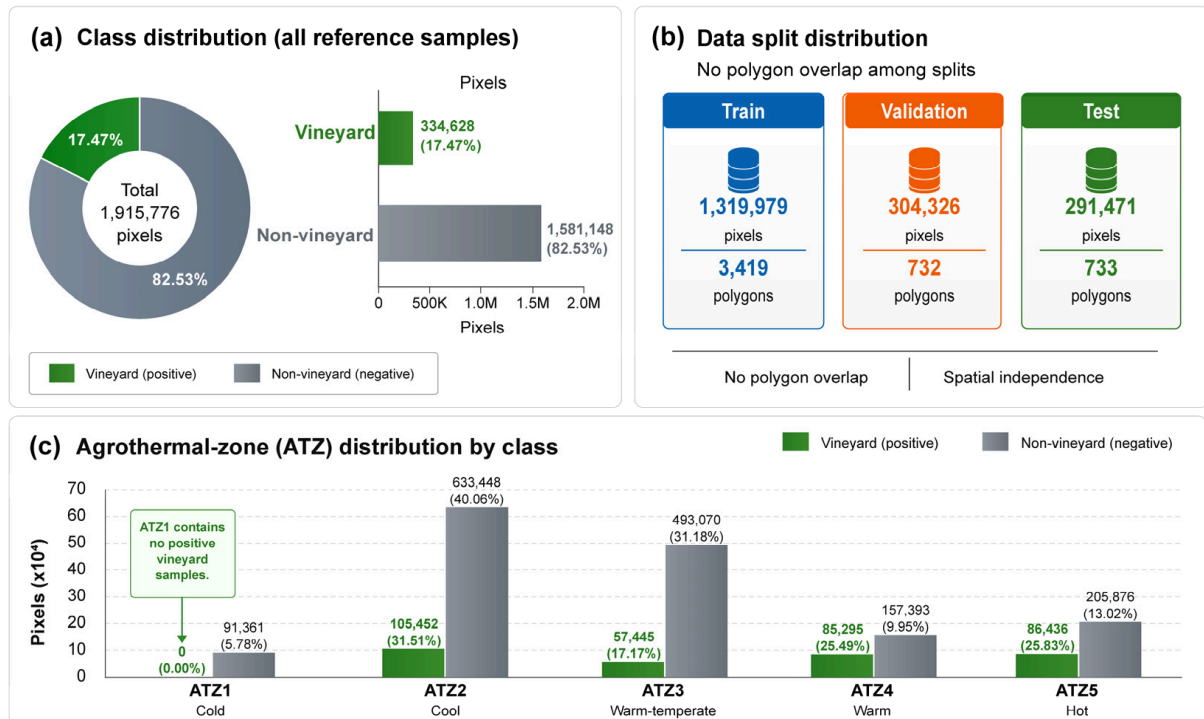


Figure 6. Distribution of reference samples across classes, data splits, and agrothermal zones.

ATZ1 contained 16,166 test pixels and no positive vineyard reference pixels. A common 0.50 threshold was used to compare raw probability-based false-positive tendencies across models. At this threshold, RF, LGBM, XGB, and Stacking produced no false-positive pixels in ATZ1, whereas 1D-CNN produced 15 false-positive pixels, corresponding to a false-positive rate of 0.093% (Table S1).

3.2. In-Distribution Classification Performance

The polygon-level in-distribution test was used as a baseline assessment to examine whether the constructed multi-source feature set could separate vineyards from non-vineyard samples under similar environmental and management conditions. The results are summarized in Table 3, and the corresponding precision–recall curves are shown in Figure 7. Overall, RF, LGBM, XGB, and Stacking all achieved near-ceiling performance, with F1-scores above 0.99 and AP values close to 1.0. LGBM obtained the highest F1-score of 0.9949, while Stacking maintained a comparable F1-score of 0.9928 and an AP of 0.9996. The 1D-CNN baseline showed lower performance than the tree-based models, suggesting that the same tabular feature vector was more effectively exploited by ensemble tree learners in this experiment.

These results confirm that the multi-source feature set provides strong discriminative information for vineyard identification under similar sample distributions. However, in-distribution performance alone cannot represent model transferability across regions with different agrothermal, terrain, irrigation, and planting conditions. There-

fore, the following Leave-One-ATZ-Out experiment was used as the main validation of cross-zone generalization.

Table 3. Baseline in-distribution classification performance on the polygon-level test set using validation-selected thresholds.

Model	Threshold	Precision	Recall	F1-Score	AP	Kappa	OA
RF	0.51	0.9918	0.9909	0.9914	0.9994	0.9895	0.9969
LGBM	0.99	0.9948	0.9949	0.9949	0.9997	0.9938	0.9982
XGB	0.98	0.9953	0.9936	0.9944	0.9997	0.9932	0.9980
1D-CNN	0.92	0.9614	0.9564	0.9589	0.9921	0.9501	0.9854
Stacking	0.94	0.9930	0.9926	0.9928	0.9996	0.9913	0.9974

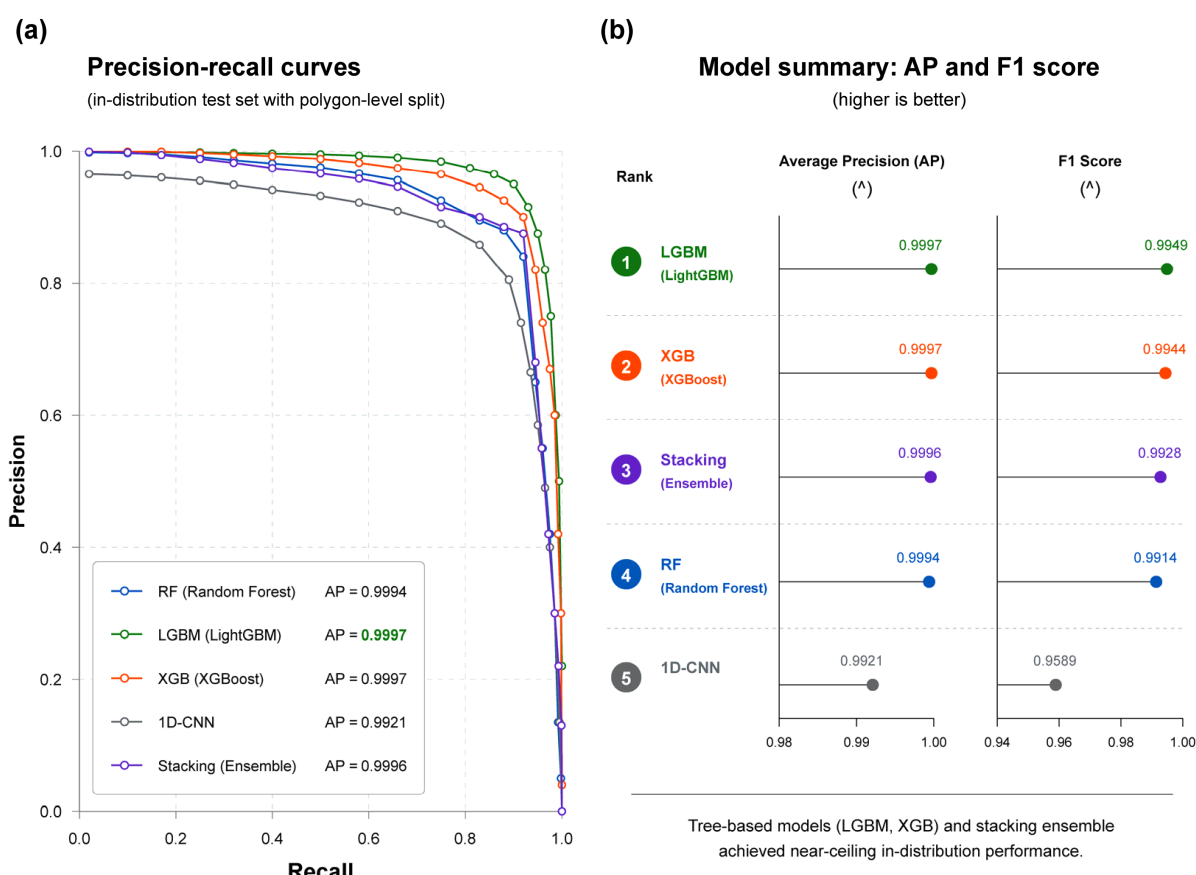


Figure 7. Precision–recall curves under polygon-level in-distribution testing, used as a baseline separability assessment.

3.3. Leave-One-ATZ-Out Cross-Zone Generalization

The Leave-One-ATZ-Out results are summarized in Table 4, with zone-specific model performance visualized in Figure 8. Compared with the in-distribution test results, all models showed lower performance under cross-ATZ validation, indicating that vineyard identification in Xinjiang is affected by clear regional heterogeneity. This decrease confirms that high accuracy under conventional polygon-level validation does not necessarily represent model transferability to unseen thermal zones.

Among the evaluated models, RF achieved the highest mean F1-score of 0.7683 and a mean AP of 0.9049 across valid held-out ATZs. However, RF also showed considerable variation across zones, with an F1-score standard deviation of 0.2246, indicating that its cross-zone performance was affected by zone-specific environmental differences. LGBM, XGB, and Stacking showed similar mean F1-scores, with Stacking achieving a mean F1-score

of 0.7381, the highest mean Recall of 0.9469, and the highest mean AP of 0.9103. Although Stacking did not exceed RF in mean cross-ATZ F1-score, it provided a high-recall and probability-based output suitable for operational thresholding and patch-level mapping. These results suggest that cross-zone evaluation should consider not only mean accuracy but also zone-to-zone variability and downstream mapping requirements.

Table 4. Summary of Leave-One-ATZ-Out cross-zone generalization performance.

Model	Valid Folds	Precision Mean	Recall Mean	F1-Score Mean	F1-Score Std	AP Mean
RF	4	0.7817	0.8426	0.7683	0.2246	0.9049
LGBM	4	0.6373	0.9351	0.7414	0.1598	0.8999
XGB	4	0.6377	0.9267	0.7364	0.1777	0.8993
Stacking	4	0.6318	0.9469	0.7381	0.1714	0.9103
1D-CNN	4	0.7571	0.7823	0.6997	0.2352	0.8754

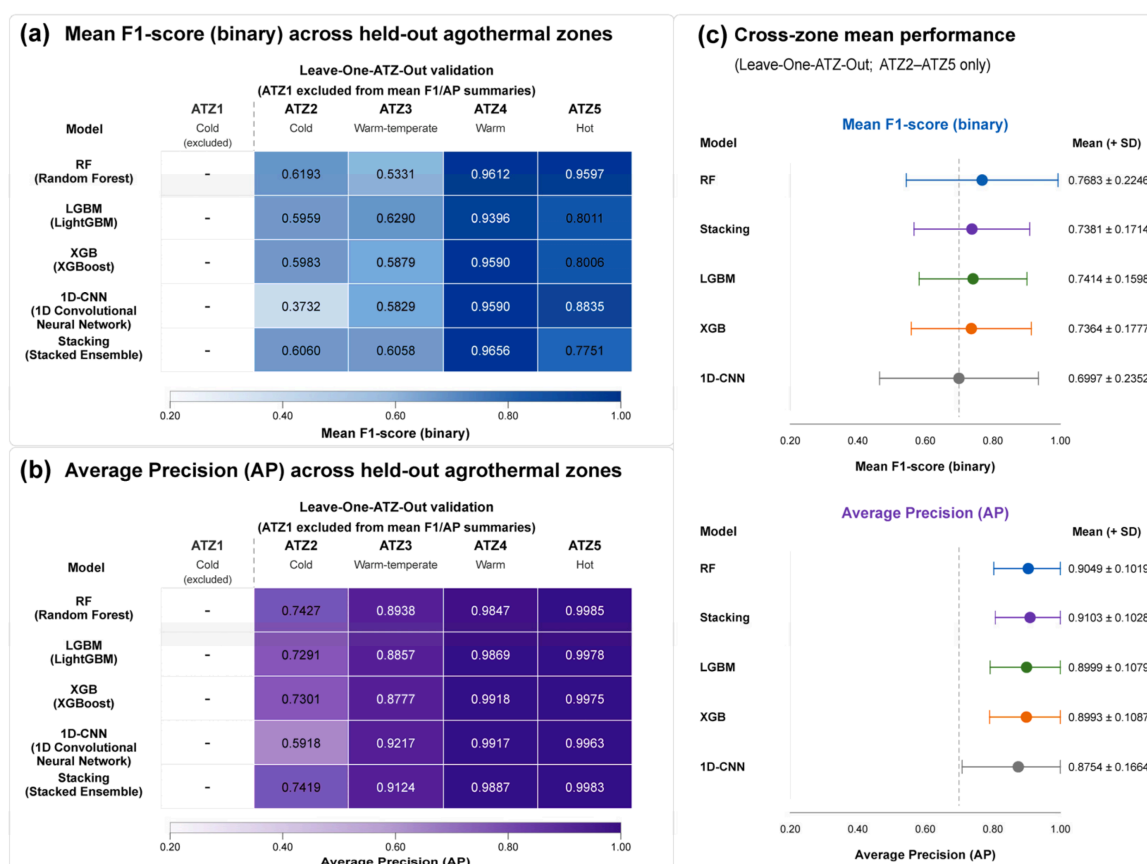


Figure 8. Heatmap of model performance across held-out agrothermal zones under Leave-One-ATZ-Out validation.

The different behaviors of the models can be partly explained by their learning mechanisms. RF, as a bagging-based ensemble, averages many decorrelated trees and is generally less sensitive to local sample peculiarities, which may help explain its relatively strong mean F1-score under cross-ATZ transfer. In contrast, boosting models such as LGBM and XGB sequentially focus on difficult samples and can fit fine-grained decision boundaries under the training distribution, resulting in excellent in-distribution performance but potentially greater sensitivity to domain shifts. The Stacking model integrates the probability outputs of multiple tree-based learners and therefore provides a flexible probability surface, but the fusion process does not guarantee superior performance in every held-out thermal zone. These results suggest that cross-zone robustness depends not only on model

accuracy under local validation but also on how each algorithm responds to environmental heterogeneity and class-distribution shifts.

Performance varied substantially among held-out ATZs. ATZ2 and ATZ3 showed lower F1-scores, suggesting stronger spectral, structural, or phenological shifts relative to the training domains. In contrast, ATZ4 and ATZ5 achieved higher classification performance, indicating better feature consistency between the held-out zone and the remaining training zones. These results highlight the importance of evaluating vineyard mapping models under cross-environment conditions rather than relying only on in-distribution accuracy.

3.4. Stability, Ablation, and Feature Importance

The multi-seed stability results are shown in Table 5. Across repeated polygon-level splits, tree-based models and the Stacking ensemble maintained relatively stable performance, whereas the 1D-CNN baseline showed larger variation. This indicates that the tree-based models were less sensitive to random split variation under the constructed tabular feature space. The Stacking ensemble achieved a mean F1-score of 0.9914 ± 0.0019 , confirming that its performance remained stable across different random seeds. LGBM showed the highest mean F1-score across repeated splits, while Stacking remained close to the top-performing tree-based models.

Table 5. Multi-seed stability of different models under polygon-level splitting.

Model	Number of Seeds	F1-Score Mean $\pm$ Std	Precision Mean $\pm$ Std	Recall Mean $\pm$ Std	AP Mean $\pm$ Std	Kappa Mean $\pm$ Std	OA Mean $\pm$ Std
RF	5	0.9889 ± 0.0019	0.9874 ± 0.0071	0.9905 ± 0.0047	0.9994 ± 0.0003	0.9866 ± 0.0022	0.9962 ± 0.0007
LGBM	5	0.9951 ± 0.0018	0.9972 ± 0.0019	0.9931 ± 0.0036	0.9997 ± 0.0004	0.9941 ± 0.0022	0.9983 ± 0.0006
XGB	5	0.9941 ± 0.0024	0.9971 ± 0.0020	0.9911 ± 0.0050	0.9997 ± 0.0003	0.9929 ± 0.0028	0.9980 ± 0.0007
1D-CNN	5	0.9639 ± 0.0093	0.9656 ± 0.0147	0.9624 ± 0.0113	0.9928 ± 0.0044	0.9565 ± 0.0109	0.9876 ± 0.0028
Stacking	5	0.9914 ± 0.0019	0.9916 ± 0.0041	0.9912 ± 0.0054	0.9972 ± 0.0035	0.9896 ± 0.0022	0.9970 ± 0.0005

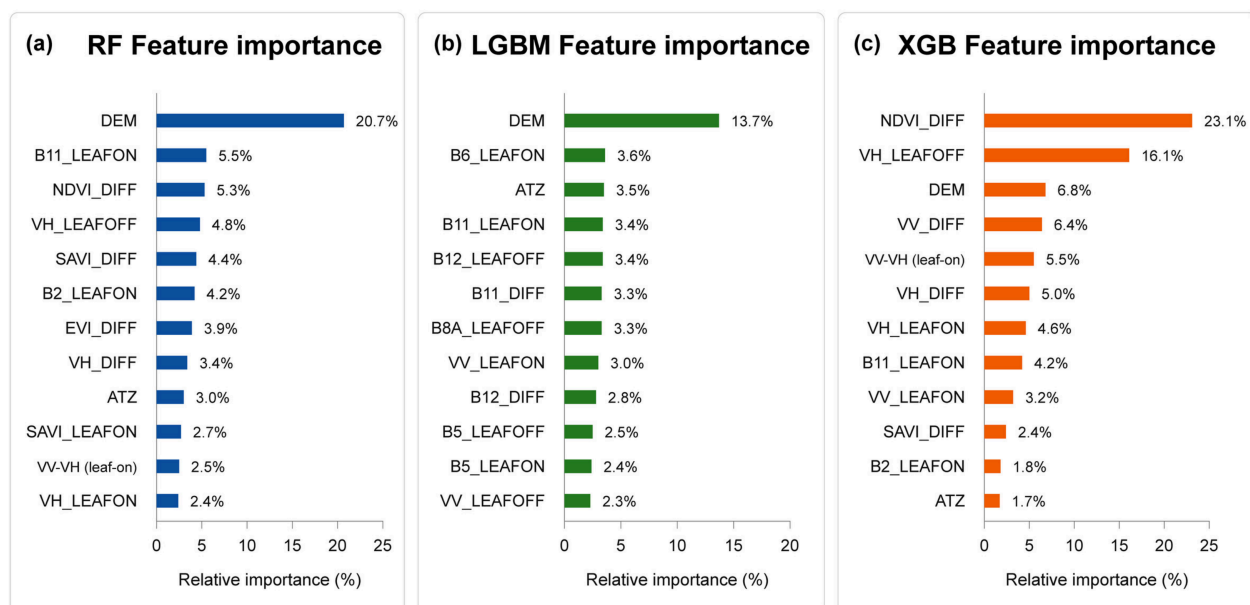
The selected cross-ATZ ablation results are listed in Table 6. The ALL rows in Table 6 use the same main Leave-One-ATZ-Out baseline values reported in Table 4 for the corresponding models, so that feature-set ablation effects are compared against the same reference baseline. Removing Sentinel-1 features reduced the cross-zone F1-score for RF, LGBM, and 1D-CNN, indicating that SAR backscatter provided complementary structural information beyond optical features. In contrast, removing topographic variables improved the Leave-One-ATZ-Out mean F1-score for RF and LGBM, suggesting that terrain variables may have a dual role: they can provide useful environmental constraints under in-distribution conditions but may also encode region-specific background patterns and reduce transferability across thermal zones. The effect of removing ATZ also varied among models, with little change for RF and LGBM but an improvement for the 1D-CNN baseline.

Because the feature set included adjacent spectral bands, vegetation indices, seasonal composites, and seasonal-difference variables, predictor collinearity was expected. Pearson correlation and VIF diagnostics confirmed strong correlations among related features: 37 feature pairs had $|r| \geq 0.90$, with a maximum $|r|$ of 0.9948, and 47 variables exceeded the VIF threshold of 10, including three variables with infinite VIF values caused by exact or near-exact linear combinations among seasonal variables (Table S2). Highly correlated variables were retained because the tree-based models and Stacking ensemble can handle redundant predictors, and because removing them would reduce the comparability of the predefined feature groups used in the ablation analysis. Therefore, feature-importance results were interpreted as relative indicators of predictor use rather than independent causal effects.

Table 6. Selected ablation results under Leave-One-ATZ-Out validation.

Model	Setting	F1-Score Mean	F1-Score Std	AP Mean
RF	ALL	0.7683	0.2246	0.9049
RF	NO_S1	0.7509	0.1931	0.8673
RF	NO_TOPO	0.8075	0.1930	0.9037
RF	NO_ATZ_FEATURE	0.7780	0.1686	0.9167
LGBM	ALL	0.7414	0.1598	0.8999
LGBM	NO_S1	0.6891	0.1714	0.8547
LGBM	NO_TOPO	0.8143	0.1555	0.9318
LGBM	NO_ATZ_FEATURE	0.7321	0.1348	0.9092
1D-CNN	ALL	0.6997	0.2352	0.8754
1D-CNN	NO_S1	0.6723	0.2457	0.8414
1D-CNN	NO_TOPO	0.6414	0.2557	0.8634
1D-CNN	NO_ATZ_FEATURE	0.7513	0.1659	0.8923

Model-derived feature importance was used to summarize the relative contribution of predictors in the tree-based base learners. RF importance was based on the impurity-based `feature_importances_` attribute, LGBM importance was based on the built-in `split-count` `feature_importances_` scores, and XGB importance was based on the built-in `gain-based` `feature_importances_` scores. The values were normalized within each model to support relative comparison. Because these importance values are model-specific ranking indicators rather than causal effects, they were interpreted as relative evidence of predictor use by each classifier. The global feature-importance results are shown in Figure 9.

**Figure 9.** Global feature importance derived from tree-based base learners.

The feature-importance results indicated that vineyard identification relied on a combination of terrain, agrothermal, optical, SAR, and seasonal-difference features. In RF, the leading predictors included DEM, B11_LEAFON, NDVI_DIFF, VH_LEAFOFF, and SAVI_DIFF. In LGBM, DEM, B6_LEAFON, ATZ, B11_LEAFON, and B12_LEAFOFF ranked highly. XGB emphasized NDVI_DIFF, VH_LEAFOFF, DEM, VV_DIFF, and VV-VH leaf-on features. The complete feature rankings are shown in Figure 9.

Overall, these patterns suggest that vineyard identification in arid oasis regions was not driven by a single spectral index or data source. Instead, classification depended on the joint contribution of elevation-related context, agrothermal suitability, optical reflectance,

seasonal vegetation changes, and SAR backscatter characteristics. The high importance of DEM and ATZ indicates that environmental background information was useful for local discrimination, but these variables should be interpreted cautiously because they may encode region-specific sample distributions and may not always transfer well across agrothermal zones.

3.5. Gaochang District Vineyard Mapping and Inventory Comparison

The final Stacking model was applied to Gaochang District for operational vineyard mapping. The spatial distribution of the mapped vineyard patches is shown in Figure 10. The Stacking-derived map contained 2295 vineyard patches with a total area of 22,998.84 ha. In comparison, the existing vineyard inventory contained 4904 parcels with a total area of 19,155.56 ha. The mapped result showed a higher total area than the existing inventory, while the number of mapped patches was lower than the number of inventory parcels. Threshold sensitivity analysis showed a consistent trade-off between inventory coverage and spatial precision (Table S3). As the probability threshold increased from 0.85 to 0.95, the mapped area decreased from 24,390.14 to 20,785.70 ha, Area Recall decreased from 0.8149 to 0.7434, and Area Precision increased from 0.6401 to 0.6852. A threshold of 0.90 was selected as an operational compromise, retaining an Area Recall of 0.7908 while improving Area Precision relative to the 0.85 threshold. Therefore, the mapped product should be interpreted as a patch-level screening result for inventory updating rather than as an independent cadastral inventory.

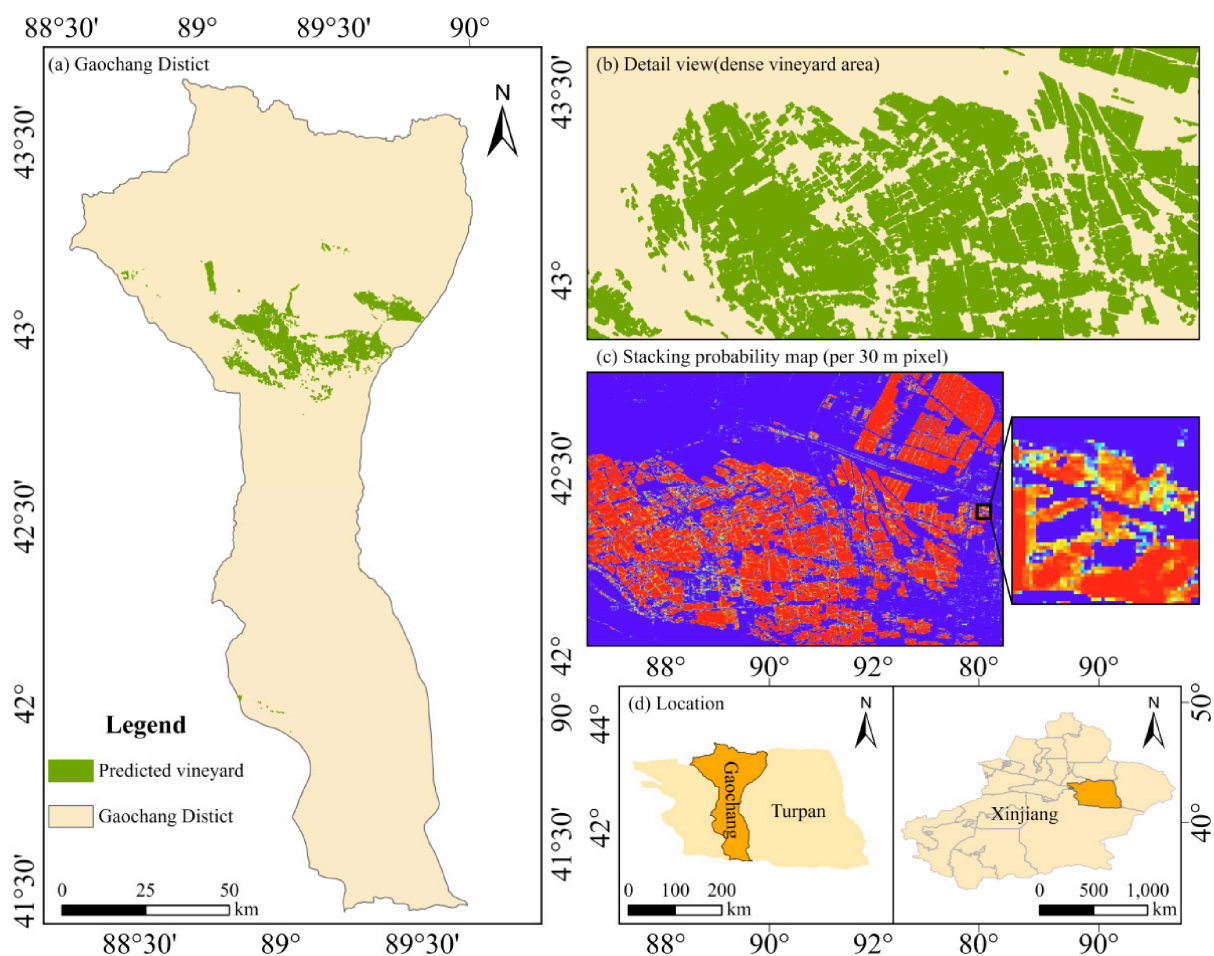


Figure 10. Stacking-derived vineyard map of Gaochang District. In (b), green marks the vineyard plots we identified. In (c), red (high) and blue (low) show the likelihood of vineyards.

The patch-level statistics are summarized in Table 7. The mean mapped patch area was 10.02 ha, with a median area of 0.82 ha. The maximum mapped patch area reached 2169.48 ha, indicating that some adjacent vineyard areas were represented as larger continuous patches after raster-to-vector conversion and post-processing. Therefore, the Gaochang output should be interpreted as a vineyard patch-level inventory product rather than a strict cadastral parcel-boundary reconstruction.

Table 7. Parcel and area statistics of the existing inventory and the Stacking-derived vineyard map in Gaochang District.

Dataset	Number of Polygons	Total Area (ha)	Mean Polygon Area (ha)	Median Polygon Area (ha)	Maximum Polygon Area (ha)
Existing inventory	4904	19,155.56	3.91	2.62	71.72
Stacking-derived map	2295	22,998.84	10.02	0.82	2169.48

The spatial agreement between the newly mapped vineyard patches and the existing inventory is summarized in Table 8, and the overlay patterns are shown in Figure 11. The intersection area between the two datasets was 15,148.83 ha, corresponding to an Area Recall of 0.7908 and an Area Precision of 0.6587. The IoU and Dice coefficient were 0.5610 and 0.7187, respectively. These results indicate that the mapped vineyard patches covered a large proportion of the existing inventory, while also identifying areas outside the inventory.

At the parcel level, 88.03% of existing inventory parcels were covered by at least 30% of the new mapping result, while Parcel Recall@50 and Parcel Recall@70 were 82.34% and 69.62%, respectively. The mapped result also identified 1356 candidate vineyard patches for inventory updating, corresponding to 2502.32 ha and 10.88% of the mapped vineyard area. These low-overlap patches should not be directly interpreted as confirmed new vineyards, because the existing inventory may contain temporal inconsistencies and boundary differences. Instead, they represent candidate areas for inventory updating and subsequent manual verification.

Table 8. Spatial agreement between the Stacking-derived vineyard map and the existing vineyard inventory in Gaochang District.

Metric	Value
Intersection area (ha)	15,148.83
Union area (ha)	27,005.58
Area Recall	0.7908
Area Precision	0.6587
IoU	0.5610
Dice coefficient	0.7187
Parcel Recall@30	0.8803
Parcel Recall@50	0.8234
Parcel Recall@70	0.6962
Candidate inventory-update patches	1356
Candidate patch area (ha)	2502.32
Candidate patch area ratio	10.88%

The total mapped area outside the historical inventory was 7850.01 ha, whereas the low-overlap candidate patches accounted for 2502.32 ha. The remaining outside-inventory area was associated mainly with predicted patches that still partially overlapped histori-

cal parcels and may reflect boundary expansion, patch merging, or inventory boundary differences rather than fully independent candidate patches.

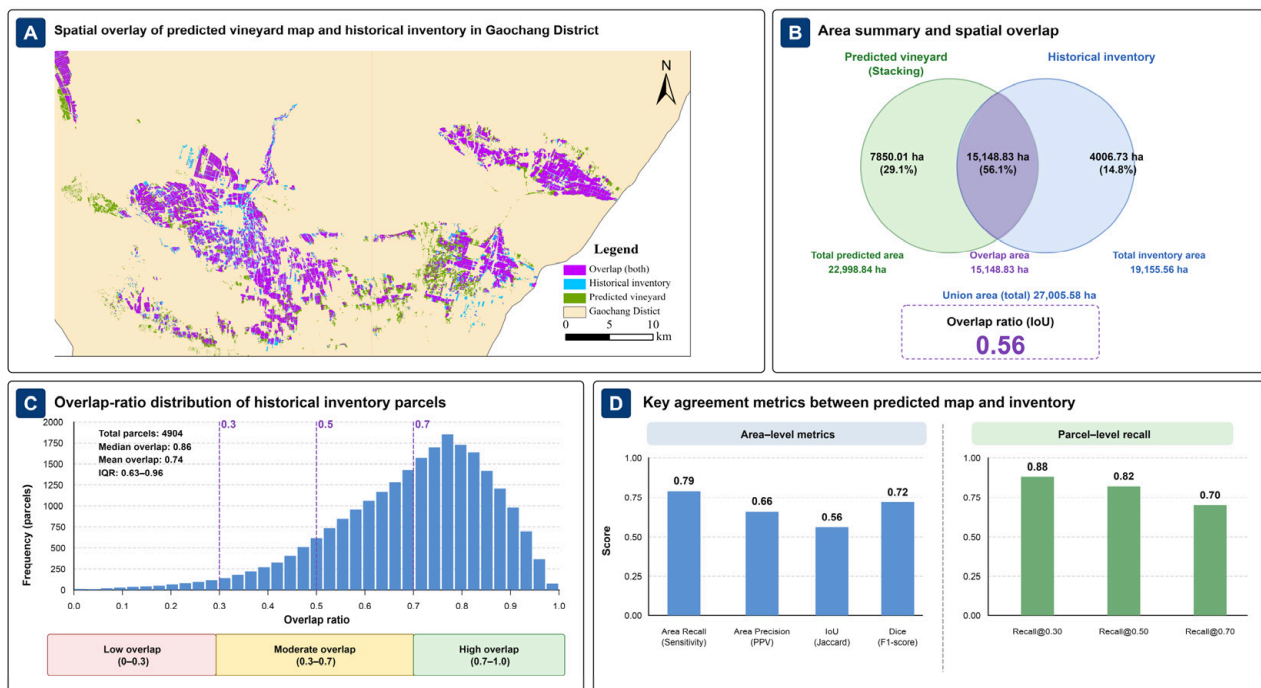


Figure 11. Spatial overlay and overlap-ratio distribution between the existing vineyard inventory and the Stacking-derived vineyard map in Gaochang District.

4. Discussion

4.1. Model Selection and Operational Transferability

Model selection for vineyard mapping should consider not only local classification accuracy but also cross-zone transferability, probability interpretability, and compatibility with downstream spatial post-processing. Under the polygon-level in-distribution test, several tree-based models achieved very high accuracy, indicating that the multi-source feature set effectively separates vineyards from non-vineyard samples under similar environmental conditions. However, the Leave-One-ATZ-Out validation revealed substantial performance degradation across all models, confirming that accumulated-temperature gradients, terrain background, irrigation context, and vineyard structure can strongly affect operational transferability.

Although RF achieved the highest mean cross-zone F1-score, its relatively large standard deviation across held-out ATZs indicates that its performance was not uniformly stable across thermal domains. Stacking did not produce the highest mean F1-score, but it achieved the highest cross-zone AP and Recall and provided a continuous probability surface for thresholding, mosaicking, raster-to-vector conversion, and patch-level post-processing. In the warmer ATZ4-ATZ5 domains most relevant to the Gaochang application, Stacking maintained high Recall and AP. In ATZ4, Stacking achieved a Recall of 0.9440 and an AP of 0.9853, compared with 0.9304 and 0.9864 for RF. In ATZ5, Stacking achieved a Recall of 0.9995 and an AP of 0.9990, compared with 0.9906 and 0.9990 for RF. Therefore, Stacking was selected primarily because it maintained high recall, produced a continuous probability surface, and was compatible with downstream thresholding and patch-level post-processing, rather than because it was superior to RF in every cross-zone metric.

The lower performance of the 1D-CNN baseline should also be interpreted cautiously. It does not indicate that deep learning is unsuitable for vineyard mapping. Rather, it reflects the limitation of using a fixed tabular feature vector to represent complex spatial

and temporal vineyard dynamics. Deep learning approaches often require structured multi-temporal sequences, spatial image patches, or explicit temporal representations to fully exploit their advantages. In this study, tree-based models were more suitable for the constructed tabular feature space.

4.2. Multi-Source Feature Importance and the Dual Role of Terrain Variables

The feature-importance and ablation results jointly indicate that vineyard mapping in arid oasis landscapes depends on complementary information from optical, SAR, terrain, agrothermal, and seasonal-difference features. Optical and seasonal-difference features captured canopy vigor, vegetation moisture, soil background, and phenological contrast, while Sentinel-1 SAR features provided structural and moisture-related backscatter information that was not fully represented by optical reflectance. The reduction in cross-ATZ performance after removing Sentinel-1 features further suggests that radar information contributed to model robustness under heterogeneous oasis conditions.

Terrain variables and ATZ information should be interpreted as environmental context rather than universally transferable vineyard indicators. DEM and ATZ ranked highly in the feature-importance results, indicating that they helped local discrimination by representing elevation–thermal suitability and oasis background conditions. However, the ablation results showed that removing topographic variables improved cross-ATZ performance for some models, suggesting that static environmental variables may also encode region-specific sample distributions. Therefore, terrain and agrothermal features have a context-dependent role: they can improve local fitting when training and test samples share similar backgrounds, but may weaken extrapolation when models are transferred to unseen agrothermal zones. This also supports the need for future GDD-driven dynamic phenological windows and region-specific phenological metrics.

The feature-importance results should therefore be interpreted with caution. Several optical bands, vegetation indices, and seasonal-difference variables were strongly correlated, which may distribute importance among related predictors. The rankings were used to identify broad evidence groups, such as terrain, SWIR, SAR, and seasonal-difference information, rather than to infer independent effects of individual variables.

4.3. Operational Mapping and Inventory-Update Potential in Gaochang District

The Gaochang District case confirms the viability of the proposed workflow for operational vineyard inventory updating. The Stacking-derived map covered most historical vineyard areas, achieving an Area Recall of 0.7908 and Parcel Recall values of 0.8803, 0.8234, and 0.6962 at the 30%, 50%, and 70% overlap thresholds, respectively. These results indicate that the workflow can reproduce a large proportion of known vineyard distributions and generate useful spatial evidence for inventory review.

Due to raster-to-vector conversion and spatial post-processing, the new map produced fewer but larger aggregated patches than the existing inventory. Specifically, the Stacking-derived map contained 2295 patches with a total area of 22,998.84 ha, whereas the existing inventory contained 4904 parcels with a total area of 19,155.56 ha. This difference reflects the patch-level nature of the output. Adjacent vineyard pixels and spatially connected agricultural vegetation may be merged during polygonization, hole filling, small-object removal, and other post-processing operations; therefore, candidate vineyard patches for inventory updating should be prioritized for manual verification rather than treated as confirmed new parcels.

The lower Area Precision of 0.6587 and IoU of 0.5610 indicate that part of the mapped area occurred outside the historical inventory. However, these areas should not be interpreted directly as false positives. They may indicate inventory omissions, boundary

discrepancies, temporal changes, newly planted vineyards, or classification errors. In this study, 1356 low-overlap mapped patches were identified as candidate vineyard patches for inventory updating, accounting for 10.88% of the mapped area. These patches provide useful targets for subsequent manual verification, field checking, or high-resolution image interpretation.

Figure 12 illustrates representative cases in Gaochang District, including correctly detected vineyard patches, omission cases, over-expanded patches, and low-overlap candidate patches requiring verification. These examples further demonstrate that the mapping result is useful for inventory screening, but still requires manual verification when used for parcel-level updating.

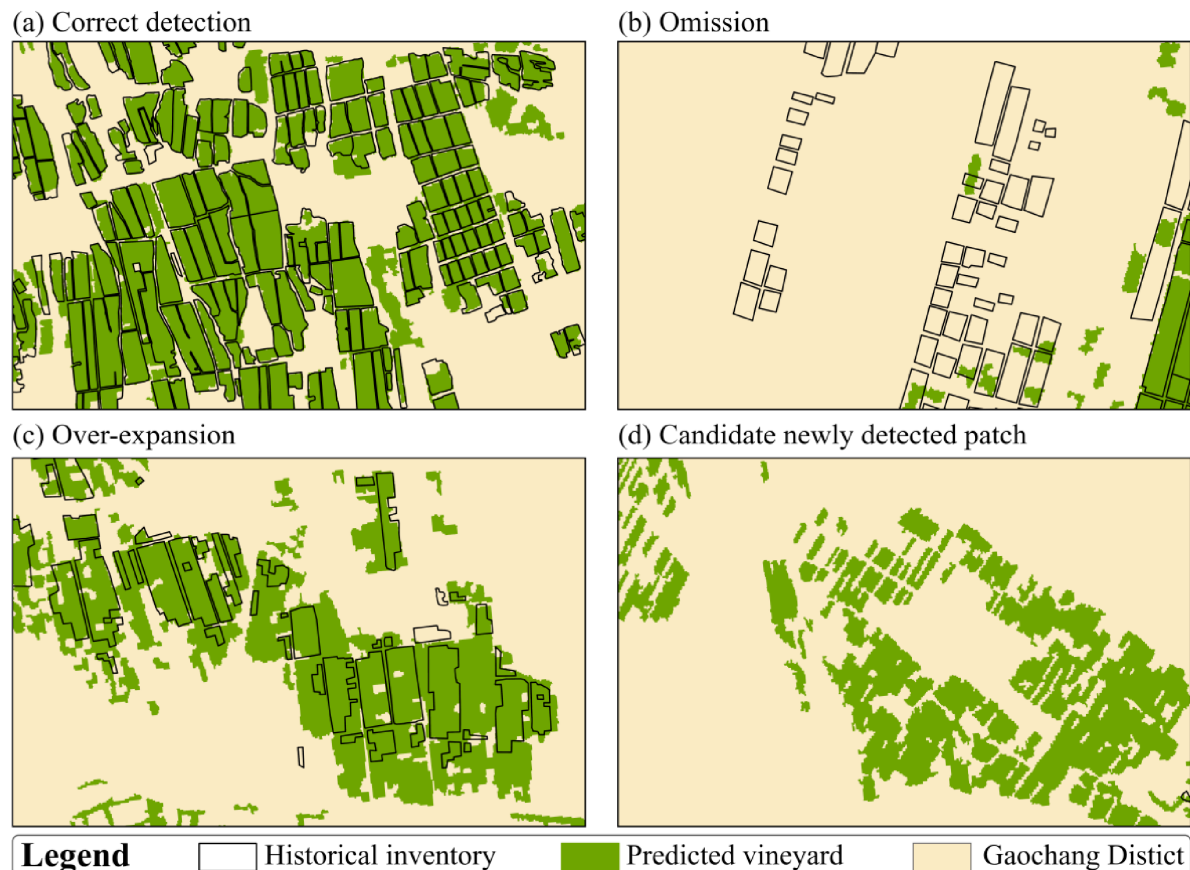


Figure 12. Typical examples of correct detection, omission, over-expansion, and low-overlap candidate patches requiring verification in Gaochang District.

Therefore, the final Gaochang product should be regarded as an operational patch-level screening product rather than a precise cadastral boundary reconstruction. It is suitable for regional vineyard distribution mapping, historical inventory comparison, and candidate update detection. If the goal is strict parcel-boundary delineation, additional high-resolution imagery, object-based segmentation, row-structure features, and boundary refinement methods would be required.

From an engineering perspective, the Gaochang workflow was implemented using tile-based inference and separate raster-to-vector post-processing. This design limits peak memory demand because each tile can be processed independently. The tile-based workflow is scalable to larger regions, although actual runtime depends on hardware configuration, tile size, and valid-pixel density.

4.4. Limitations and Future Work

First, the leaf-on and leaf-off composites were based on fixed calendar-month windows. This design improved reproducibility and simplified large-area GEE implementation, but it cannot fully account for interannual or cross-ATZ phenological differences. Future work should test adaptive phenological windows or time-series models aligned with local grapevine development stages.

Second, polygon-level splitting reduces within-parcel leakage, but it does not fully eliminate spatial dependence among nearby parcels. The Leave-One-ATZ-Out experiment provides a stronger assessment of cross-thermal-zone transferability, but additional spatial block validation or county-level independent validation would be needed to more rigorously test geographic extrapolation.

Third, the reference samples and historical inventory may contain temporal, boundary, and interpretation uncertainties. Although manual refinement using high-resolution imagery improved sample quality, the labels were visually verified rather than field surveyed. Therefore, the Gaochang inventory comparison should be interpreted as spatial consistency and inventory-update potential rather than independent ground-truth validation.

Fourth, candidate-area filtering and vector post-processing strongly affect the final patch morphology and mapped area. These steps improve map usability by reducing noise and suppressing obvious non-vineyard areas, but they may also remove weak-signal vineyards, merge adjacent patches, or shift boundaries. The threshold sensitivity analysis added in this revision partly addresses this uncertainty, but future work should further test candidate-mask thresholds, minimum patch size, hole filling, inward buffering, and dissolve settings.

Finally, the 1D-CNN baseline used the same tabular feature vector as the machine-learning models and therefore did not exploit spatial image patches or explicit temporal sequences. Its lower performance should not be interpreted as evidence that deep learning is unsuitable for vineyard mapping. Future deep learning experiments should use structured multi-temporal or spatial inputs [46].

5. Conclusions

This study developed a multi-source remote-sensing and Stacking ensemble framework for vineyard mapping and inventory updating in arid Xinjiang. The framework integrated optical, SAR, terrain, agrothermal, and seasonal-difference information and evaluated RF, LGBM, XGB, 1D-CNN, and Stacking models using both in-distribution testing and Leave-One-ATZ-Out cross-zone validation.

The in-distribution test confirmed that the constructed multi-source feature set provided strong separability between vineyards and non-vineyards under similar sample distributions. However, cross-ATZ validation showed that model performance declined substantially when transferred to unseen agrothermal zones, indicating that conventional local validation may overestimate operational transferability. RF achieved the highest mean cross-zone F1-score, whereas Stacking achieved the highest cross-zone AP and Recall and provided a continuous probability surface suitable for thresholding, mosaicking, vector conversion, and patch-level post-processing.

Feature-importance and ablation analyses indicated that vineyard identification relied on the combined effects of optical phenology, SAR backscatter, terrain background, agrothermal context, and seasonal-difference features. Sentinel-1 features provided complementary structural information, while seasonal-difference features enhanced vineyard phenological contrast. Terrain and ATZ variables showed a dual role, supporting local discrimination but potentially introducing region-specific dependence under cross-zone transfer.

In Gaochang District, the Stacking-derived vineyard map contained 2295 mapped patches with a total potential vineyard area of 22,998.84 ha. Compared with the existing vineyard inventory, the mapped result achieved an Area Recall of 0.7908, an Area Precision of 0.6587, an IoU of 0.5610, and a Dice coefficient of 0.7187. At the parcel level, 88.03% of historical inventory parcels were covered by at least 30% of the new mapping result. The workflow also identified 1356 candidate vineyard patches for inventory updating covering 2502.32 ha, which can support subsequent inventory verification and updating. Overall, the proposed framework provides a practical approach for vineyard mapping in arid regions, and the comparison between in-distribution and cross-zone validation highlights the need to evaluate both recognition accuracy and spatial transferability before operational deployment.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/agriengineering8080318/s1>, Table S1: ATZ1 false-positive statistics under a common 0.50 probability threshold; Table S2: (a) Summary of collinearity diagnostics, (b) Feature pairs with Pearson $|r| \geq 0.90$, (c) VIF values for the 50 input predictors; Table S3: sensitivity of Gaochang vineyard mapping results to the Stacking probability threshold after identical raster-to-vector post-processing.

Author Contributions: Conceptualization, L.Y.; methodology, L.Y., L.W. and Z.P.; software, L.Y.; validation, L.Y., L.W. and Z.P.; formal analysis, L.Y.; investigation, L.Y.; resources, L.W.; data curation, L.Y., L.W., Z.P., S.Z. and S.W.; writing—original draft preparation, L.Y.; writing—review and editing, L.Y., L.W., Z.P. and L.L.; visualization, L.Y.; supervision, L.W. and Z.P.; project administration, L.W.; funding acquisition, L.W. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the project “Construction of Forest and Fruit Resource Data System Based on Space-Air-Ground Multi-Source Remote Sensing Monitoring Technology” (Project No. 2023B02004-1), supported by the Xinjiang Uygur Autonomous Region Academy of Forestry Sciences and the Xinjiang Key Laboratory of Desertification Control and Sand Industry. The Article Processing Charge (APC) was also funded by this project.

Data Availability Statement: The Sentinel-1, Sentinel-2, ERA5-Land, and DEM datasets used in this study are publicly available through Google Earth Engine. The derived feature tables, trained models, source code, and derived evaluation results are available from the corresponding authors upon reasonable request. The reference vineyard polygons and historical inventory data are not publicly available because they were derived from institutional inventory data and manually refined interpretation results, and are subject to data-use restrictions.

Acknowledgments: The authors gratefully acknowledge the Xinjiang Uygur Autonomous Region Academy of Forestry Sciences and the Xinjiang Key Laboratory of Desertification Control and Sand Industry for their support in data provision and technical assistance. The authors also acknowledge the Google Earth Engine platform for providing access to cloud-based geospatial processing resources. During the preparation of this manuscript, the authors used ChatGPT (web-based service; no fixed local version; <https://chatgpt.com/>; accessed on 20 July 2026), developed by OpenAI, for language polishing, structural editing, and manuscript formatting assistance. The authors have reviewed and edited the output and take full responsibility for the content of this publication.

Conflicts of Interest: The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

Abbreviations

The following abbreviations are used in this manuscript:

AP	Average Precision
ATZ	Agrothermal Zone
CNN	Convolutional Neural Network
DEM	Digital Elevation Model
GDD	Growing Degree Days
GEE	Google Earth Engine
IoU	Intersection over Union
LGBM	Light Gradient Boosting Machine
OA	Overall Accuracy
OOF	Out-of-Fold
RF	Random Forest
SAR	Synthetic Aperture Radar
SVM	Support Vector Machine
XGB	Extreme Gradient Boosting
1D-CNN	One-Dimensional Convolutional Neural Network

References

1. Mucalo, A.; Matić, D.; Morić-Španić, A.; Čagalj, M. Satellite Solutions for Precision Viticulture: Enhancing Sustainability and Efficiency in Vineyard Management. *Agronomy* **2024**, *14*, 1862. [\[CrossRef\]](#)
2. Brandmeier, M.; Heßdörfer, D.; Siebenlist, P.; Meyer-Spelbrink, A.; Kraus, A. Time Series Analysis of Multisensor Data for Precision Viticulture—Assessing Microscale Variations in Plant Development with Respect to Irrigation and Topography. *Remote Sens.* **2024**, *16*, 1419. [\[CrossRef\]](#)
3. Ohana-Levi, N.; Gao, F.; Knipper, K.; Kustas, W.P.; Anderson, M.C.; Alsina, M.M.; Sanchez, L.A.; Karnieli, A. Time-Series Clustering of Remote Sensing Retrievals for Defining Management Zones in a Vineyard. *Irrig. Sci.* **2022**, *40*, 801–815. [\[CrossRef\]](#)
4. Drusch, M.; Del Bello, U.; Carlier, S.; Colin, O.; Fernandez, V.; Gascon, F.; Hoersch, B.; Isola, C.; Laberinti, P.; Martimort, P.; et al. Sentinel-2: ESA's Optical High-Resolution Mission for GMES Operational Services. *Remote Sens. Environ.* **2012**, *120*, 25–36. [\[CrossRef\]](#)
5. Gómez, C.; White, J.C.; Wulder, M.A. Optical Remotely Sensed Time Series Data for Land Cover Classification: A Review. *ISPRS J. Photogramm. Remote Sens.* **2016**, *116*, 55–72. [\[CrossRef\]](#)
6. Vuolo, F.; Neuwirth, M.; Immitzer, M.; Atzberger, C.; Ng, W.T. How Much Does Multi-Temporal Sentinel-2 Data Improve Crop Type Classification? *Int. J. Appl. Earth Obs. Geoinf.* **2018**, *72*, 122–130. [\[CrossRef\]](#)
7. Abubakar, M.A.; Chanzy, A.; Flamain, F.; Pouget, G.; Courault, D. Delineation of Orchard, Vineyard, and Olive Trees Based on Phenology Metrics Derived from Time Series of Sentinel-2. *Remote Sens.* **2023**, *15*, 2420. [\[CrossRef\]](#)
8. Pace, L.; Coluzzi, R.; Imbrenda, V.; D'Emilio, M.; Falcone, A.; Nuzzo, V.; Lanfredi, M. Combining Multi-Source Data to Map Vineyards in a Specialized District of Basilicata (Southern Italy). *J. Maps* **2024**, *20*, 2327859. [\[CrossRef\]](#)
9. Tassopoulos, D.; Kalivas, D.; Giovos, R.; Lougkos, N.; Priovolou, A. Sentinel-2 Imagery Monitoring Vine Growth Related to Topography in a Protected Designation of Origin Region. *Agriculture* **2021**, *11*, 785. [\[CrossRef\]](#)
10. Campos, J.; García-Ruiz, F.; Gil, E. Assessment of Vineyard Canopy Characteristics from Vigour Maps Obtained Using UAV and Satellite Imagery. *Sensors* **2021**, *21*, 2363. [\[CrossRef\]](#) [\[PubMed\]](#)
11. Ilniyaz, O.; Du, Q.; Shen, H.; He, W.; Feng, L.; Azadi, H.; Kurban, A.; Chen, X. Leaf Area Index Estimation of Pergola-Trained Vineyards in Arid Regions Using Classical and Deep Learning Methods Based on UAV-Based RGB Images. *Comput. Electron. Agric.* **2023**, *207*, 107723. [\[CrossRef\]](#)
12. Gavrilović, M.; Jovanović, D.; Božović, P.; Benka, P.; Govedarica, M. Vineyard Zoning and Vine Detection Using Machine Learning in Unmanned Aerial Vehicle Imagery. *Remote Sens.* **2024**, *16*, 584. [\[CrossRef\]](#)
13. Samadzadegan, F.; Toosi, A.; Dadrass Javan, F. A Critical Review on Multi-Sensor and Multi-Platform Remote Sensing Data Fusion Approaches: Current Status and Prospects. *Int. J. Remote Sens.* **2025**, *46*, 1327–1402. [\[CrossRef\]](#)
14. Adrian, J.; Sagan, V.; Maimaitijiang, M. Sentinel SAR-Optical Fusion for Crop Type Mapping Using Deep Learning and Google Earth Engine. *ISPRS J. Photogramm. Remote Sens.* **2021**, *175*, 215–235. [\[CrossRef\]](#)
15. Dobrinić, D.; Gašparović, M.; Medak, D. Sentinel-1 and 2 Time-Series for Vegetation Mapping Using Random Forest Classification: A Case Study of Northern Croatia. *Remote Sens.* **2021**, *13*, 2321. [\[CrossRef\]](#)

16. Saad El Imanni, H.; El Harti, A.; Hssaisoune, M.; Velastegui-Montoya, A.; Elbouzidi, A.; Addi, M.; El Iysaouy, L.; El Hachimi, J. Rapid and Automated Approach for Early Crop Mapping Using Sentinel-1 and Sentinel-2 on Google Earth Engine; A Case of a Highly Heterogeneous and Fragmented Agricultural Region. *J. Imaging* **2022**, *8*, 316. [\[CrossRef\]](#) [\[PubMed\]](#)
17. Luo, C.; Liu, H.J.; Lu, L.P.; Liu, Z.R.; Kong, F.C.; Zhang, X.L. Monthly Composites from Sentinel-1 and Sentinel-2 Images for Regional Major Crop Mapping with Google Earth Engine. *J. Integr. Agric.* **2021**, *20*, 1944–1957. [\[CrossRef\]](#)
18. Hu, T.; Hu, Y.; Dong, J.; Zhang, X.; Huang, J.; Liu, J.; Du, G.; Liu, X.; Zheng, Y.; Xie, Q. Integrating Sentinel-1/2 Data and Machine Learning to Map Cotton Fields in Northern Xinjiang, China. *Remote Sens.* **2021**, *13*, 4819. [\[CrossRef\]](#)
19. Liang, Q.; Di, Y.; Hao, X.; Zhang, J.; Ci, M.; Sun, F.; Wang, C.; Fan, X.; Guo, X. A 30 m Multi-Year Dataset of Major Crop Distributions in Xinjiang, China (2013–2024) Based on Harmonized Landsat-Sentinel-2 Data. *Sci. Data* **2026**, *13*, 730. [\[CrossRef\]](#) [\[PubMed\]](#)
20. Cameron, W.; Petrie, P.R.; Barlow, E.W.R.; Howell, K.; Jarvis, C.; Fuentes, S. A Comparison of the Effect of Temperature on Grapevine Phenology between Vineyards. *OENO One* **2021**, *55*, 283–301. [\[CrossRef\]](#)
21. Faralli, M.; Mallucci, S.; Bignardi, A.; Varner, M.; Bertamini, M. Four Decades in the Vineyard: The Impact of Climate Change on Grapevine Phenology and Wine Quality in Northern Italy. *OENO One* **2024**, *58*, 8083. [\[CrossRef\]](#)
22. Bambach, N.; Kustas, W.; Alfieri, J.; Gao, F.; Prueger, J.; Hipps, L.; McKee, L.; Castro, S.J.; Alsina, M.M.; McElrone, A.J. Inter-Annual Variability of Land Surface Fluxes across Vineyards: The Role of Climate, Phenology, and Irrigation Management. *Irrig. Sci.* **2022**, *40*, 463–480. [\[CrossRef\]](#) [\[PubMed\]](#)
23. Belgiu, M.; Bijker, W.; Csillik, O.; Stein, A. Phenology-Based Sample Generation for Supervised Crop Type Classification. *Int. J. Appl. Earth Obs. Geoinf.* **2021**, *95*, 102264. [\[CrossRef\]](#)
24. Li, R.; Xu, M.; Chen, Z.; Gao, B.; Cai, J.; Shen, F.; He, X.; Zhuang, Y.; Chen, D. Phenology-Based Classification of Crop Species and Rotation Types Using Fused MODIS and Landsat Data: The Comparison of a Random-Forest-Based Model and a Decision-Rule-Based Model. *Soil Tillage Res.* **2021**, *206*, 104838. [\[CrossRef\]](#)
25. Guo, Y.; Xia, H.; Pan, L.; Zhao, X.; Li, R.; Bian, X.; Wang, R.; Yu, C. Development of a New Phenology Algorithm for Fine Mapping of Cropping Intensity in Complex Planting Areas Using Sentinel-2 and Google Earth Engine. *ISPRS Int. J. Geo-Inf.* **2021**, *10*, 587. [\[CrossRef\]](#)
26. Pan, L.; Xia, H.; Yang, J.; Niu, W.; Wang, R.; Song, H.; Guo, Y.; Qin, Y. Mapping Cropping Intensity in Huaihe Basin Using Phenology Algorithm, All Sentinel-2 and Landsat Images in Google Earth Engine. *Int. J. Appl. Earth Obs. Geoinf.* **2021**, *102*, 102376. [\[CrossRef\]](#)
27. Zhang, L.; Gao, L.; Huang, C.; Wang, N.; Wang, S.; Peng, M.; Zhang, X.; Tong, Q. Crop Classification Based on the Spectrotemporal Signature Derived from Vegetation Indices and Accumulated Temperature. *Int. J. Digit. Earth* **2022**, *15*, 626–652. [\[CrossRef\]](#)
28. Khatami, A.; Mountrakis, G.; Stehman, S.V. A Meta-Analysis of Remote Sensing Research on Supervised Pixel-Based Land-Cover Image Classification Processes: General Guidelines for Practitioners and Future Research. *Remote Sens. Environ.* **2016**, *177*, 89–100. [\[CrossRef\]](#)
29. Awad, M. Google Earth Engine (GEE) Cloud Computing Based Crop Classification Using Radar, Optical Images and Support Vector Machine Algorithm (SVM). In Proceedings of the 2021 IEEE 3rd International Multidisciplinary Conference on Engineering Technology (IMCET), Beirut, Lebanon, 8–10 December 2021; pp. 1–6. [\[CrossRef\]](#)
30. Gorelick, N.; Hancher, M.; Dixon, M.; Ilyushchenko, S.; Thau, D.; Moore, R. Google Earth Engine: Planetary-Scale Geospatial Analysis for Everyone. *Remote Sens. Environ.* **2017**, *202*, 18–27. [\[CrossRef\]](#)
31. Liu, M.; Wang, J.; Fetisov, D.; Li, K.; Xu, C.; Jiang, J. Machine Learning-Based Fine Classification of Agricultural Crops in the Cross-Border Basin of the Heilongjiang River between China and Russia. *Remote Sens.* **2024**, *16*, 1670. [\[CrossRef\]](#)
32. Yao, J.; Wu, J.; Xiao, C.; Zhang, Z.; Li, J. The Classification Method Study of Crops Remote Sensing with Deep Learning, Machine Learning, and Google Earth Engine. *Remote Sens.* **2022**, *14*, 2758. [\[CrossRef\]](#)
33. Ji, S.; Zhang, C.; Xu, A.; Shi, Y.; Duan, Y. 3D Convolutional Neural Networks for Crop Classification with Multi-Temporal Remote Sensing Images. *Remote Sens.* **2018**, *10*, 75. [\[CrossRef\]](#)
34. Pelletier, C.; Webb, G.I.; Petitjean, F. Temporal Convolutional Neural Network for the Classification of Satellite Image Time Series. *Remote Sens.* **2019**, *11*, 523. [\[CrossRef\]](#)
35. Zhong, L.; Hu, L.; Zhou, H. Deep Learning Based Multi-Temporal Crop Classification. *Remote Sens. Environ.* **2019**, *221*, 430–443. [\[CrossRef\]](#)
36. Zhou, Y.; Luo, J.; Feng, L.; Yang, Y.; Chen, Y.; Wu, W. Long-Short-Term-Memory-Based Crop Classification Using High-Resolution Optical Images and Multi-Temporal SAR Data. *GIScience Remote Sens.* **2019**, *56*, 1170–1191. [\[CrossRef\]](#)
37. Wei, S.; Zhang, H.; Wang, C.; Wang, Y.; Xu, L. Multi-Temporal SAR Data Large-Scale Crop Mapping Based on U-Net Model. *Remote Sens.* **2019**, *11*, 68. [\[CrossRef\]](#)
38. Chamorro Martinez, J.A.; Cué La Rosa, L.E.; Feitosa, R.Q.; Sanches, I.D.; Happ, P.N. Fully Convolutional Recurrent Networks for Multidate Crop Recognition from Multitemporal Image Sequences. *ISPRS J. Photogramm. Remote Sens.* **2021**, *171*, 188–201. [\[CrossRef\]](#)

39. Zhao, H.; Duan, S.; Liu, J.; Sun, L.; Reymondin, L. Evaluation of Five Deep Learning Models for Crop Type Mapping Using Sentinel-2 Time Series Images with Missing Information. *Remote Sens.* **2021**, *13*, 2790. [\[CrossRef\]](#)
40. Xu, J.; Yang, J.; Xiong, X.K.; Li, H.; Huang, J.; Ting, K.C.; Ying, Y.; Lin, T. Towards Interpreting Multi-Temporal Deep Learning Models in Crop Mapping. *Remote Sens. Environ.* **2021**, *264*, 112599. [\[CrossRef\]](#)
41. Xu, J.; Zhu, Y.; Zhong, R.; Lin, Z.; Xu, J.; Jiang, H.; Huang, J.; Li, H.; Lin, T. DeepCropMapping: A Multi-Temporal Deep Learning Approach with Improved Spatial Generalizability for Dynamic Corn and Soybean Mapping. *Remote Sens. Environ.* **2020**, *247*, 111946. [\[CrossRef\]](#)
42. López-Pérez, E.; Sanchis-Ibor, C.; Jiménez-Bello, M.Á.; Pulido-Velazquez, M. Mapping of Irrigated Vineyard Areas through the Use of Machine Learning Techniques and Remote Sensing. *Agric. Water Manag.* **2024**, *302*, 108988. [\[CrossRef\]](#)
43. Han, X.; Ye, H.; Zhang, Y.; Nie, C.; Wen, F. Study on the Method of Vineyard Information Extraction Based on Spectral and Texture Features of GF-6 Satellite Imagery. *Agronomy* **2024**, *14*, 2542. [\[CrossRef\]](#)
44. Abdali, E.; Valadan Zoej, M.J.; Taheri Dehkordi, A.; Ghaderpour, E. A Parallel-Cascaded Ensemble of Machine Learning Models for Crop Type Classification in Google Earth Engine Using Multi-Temporal Sentinel-1/2 and Landsat-8/9 Remote Sensing Data. *Remote Sens.* **2024**, *16*, 127. [\[CrossRef\]](#)
45. Xue, H.; Xu, X.; Zhu, Q.; Yang, G.; Long, H.; Li, H.; Yang, X.; Zhang, J.; Yang, Y.; Xu, S.; et al. Object-Oriented Crop Classification Using Time Series Sentinel Images from Google Earth Engine. *Remote Sens.* **2023**, *15*, 1353. [\[CrossRef\]](#)
46. Zhang, X.; Cai, Z.; Hu, Q.; Yang, J.; Wei, H.; You, L.; Xu, B. Improving Crop Type Mapping by Integrating LSTM with Temporal Random Masking and Pixel-Set Spatial Information. *ISPRS J. Photogramm. Remote Sens.* **2024**, *218*, 87–101. [\[CrossRef\]](#)

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.