Smart Agricultural Technology 14 (2026) 102030

Contents lists available at ScienceDirect

Smart Agricultural Technology

journal homepage: www.journals.elsevier.com/smart-agricultural-technology

Field-scale potato yield prediction from sentinel-2 time series using lightweight deep learning models

Rahat Tufail, Patrizia Tassinari, Daniele Torreggiani *

Department of Agricultural and Food Sciences, University of Bologna, Viale Fanin 48, 40127 Bologna, Italy

ARTICLE INFO

Keywords:
Crop yield prediction
Potato
Lightweight
Machine learning
Deep learning
Sentinel-2

ABSTRACT

Accurate and timely crop yield prediction and forecasting are important for improving agricultural productivity and supporting informed management decisions. In this study, we developed and evaluated a framework for estimating in-season potato yield at the field scale using Sentinel-2 satellite time series. A Grouped TreeSHAP Stability Selection (GTSS) was first applied to identify a compact, phenology-aware subset of spectral bands and vegetation indices, thereby reducing redundancy and mitigating overfitting in small-data settings. Two deep learning architectures tailored for limited training data were then introduced: LiteTemporalConv, a lightweight temporal convolutional network, and MS-ConvBiGRU-Attn, a hybrid encoder combining multi-scale convolutions, bidirectional GRUs, and an attention mechanism. Both models were benchmarked against widely used machine learning methods, including Random Forest, Support Vector Machine, Extreme Gradient Boost, Partial Least Squares, as well as standard deep learning baselines (CNN and GRU). Results showed that the proposed models outperformed both machine learning and conventional deep learning baselines, with LiteTemporalConv achieving the highest accuracy under 10-fold cross-validation ($R^2 = 0.84$; RMSE = 3.18 t ha^{-1}; rRMSE = 6.36%) and MS-ConvBiGRU-Attn yielding similarly strong performance ($R^2 = 0.82$; RMSE = 3.53 t ha^{-1}; rRMSE = 7.03%). By comparison, the best baseline, XGB, achieved an R^2 of 0.79 with an rRMSE of 9.8%. The two best-performing models were further evaluated on an independent spatial dataset to assess their generalization beyond the training region. In an additional experiment, both deep learning models trained on mid-season observations showed predictive stability for late-season yield estimation. Overall, the results highlight the importance of targeted feature selection and lightweight encoders for yield modeling in data-scarce conditions.

1. Introduction

The global population is expected to grow, reaching around 9.8 billion by 2050 and potentially 11.8 billion by 2100, with an average annual increase of approximately 83 million people [1], However, the amount of cultivated land per person is declining, making it challenging to meet food needs [2]. As the food demand is expected to double [3], intensified agricultural practices and land expansion have increasingly impacted Earth's biogeochemical cycles [4]. Hence, the agricultural sector is under growing pressure from climate change and resource overuse, both of which are intensifying global food security concerns [5]. Rijsberman et al. [6] emphasized the need to increase food production by 40% while preserving ecosystems through environmentally viable agricultural practices [7]. Sustainable Development Goal 2 (SDG 2) promotes sustainable agriculture and empowers producers, where accurate crop yield estimation plays a major role in integrating efficient food production and distribution [8].

Blackmore et al. [9] highlighted the potential of precision agriculture to advance sustainable farming practices. Precision agriculture (PA) technologies, such as Global Positioning System (GPS), Geographic Information System (GIS), and remote sensing (RS), provide timely and cost-effective data on real-time field conditions [10,11]. These data streams not only support better crop yield predictions but also help to reduce operational costs [12]. Estimating and mapping crop yields using Earth observation data provides valuable information to farmers, decision-makers, insurers, and nonprofit agencies, as it supports planning, risk management, and resilience in food systems [13,14]. Forecasting crop yield before harvest allows stakeholders to plan logistics such as collection and storage, and align production with market requirements [15,16]. Early yield estimates at the field scale enable farmers to guide on-farm decisions and help governments in implementing food security initiatives and policy reforms [17,18].

* Corresponding author.
E-mail addresses: rahat.tufail2@unibo.it (R. Tufail), daniele.torreggiani@unibo.it (D. Torreggiani).

https://doi.org/10.1016/j.atech.2026.102030
Received 20 January 2026; Received in revised form 20 March 2026; Accepted 20 March 2026
Available online 20 March 2026

Potato (*Solanum tuberosum L.*) is one of the most important staple crops worldwide. Originating in the Andes Mountains of southern Peru [19], it is the third most consumed food crop after rice and wheat [20] and a critical contributor to global food security [21]. In 2023, potatoes were cultivated on nearly 1.3 million hectares across the EU, yielding over 48 million tons and contributing approximately $334 million in trade surplus [22]. Italy produces approximately 3.2 million tons of potatoes annually across 13.8 thousand hectares in 2020, with most of the production concentrated in Sicily, Campania, and Puglia [23], while Lazio, Veneto, and Calabria also contribute significantly. In 2023, Italy exported potatoes worth $39.8 million, ranking 17th among 155 global exporters [24]. To support both domestic supply strategies and the export potential, accurate potato yield estimation and high-resolution yield maps are crucial for identifying spatial yield variability and improving management decisions [25].

Traditionally, crop growth models have been employed to simulate crop responses to management practices and environmental conditions. Several well-established potato crop models have been developed and discussed in the literature, including APSIM-Potato, LINTUL-Potato, SOLANUM, SUBSTOR-Potato, and SPUDSIM [26-28]. These models have demonstrated their capability to represent key plant responses to factors such as nitrogen fertilization [29], temperature, photoperiod [30], and solar radiation [31], and have been successfully applied for yield estimation during the growing season. However, their operational application at field-to-regional scales often depends on detailed input data, calibration procedures, and modeling assumptions that may not always be readily available across diverse production systems [27,32]. In addition, representing spatial variability and extreme weather impacts over large areas can remain challenging within many modeling frameworks [20,33]. These considerations motivate the exploration of complementary, data-driven, and spatially explicit approaches, particularly those that utilize satellite observations to support practical and scalable potato yield monitoring.

The advent of Earth observation (EO) data has made large-scale crop monitoring and yield prediction both feasible and cost-effective, and remote sensing is now widely applied in precision agriculture [33-35]. However, selecting the most appropriate dataset and method for yield estimation remains a major challenge [36]. Most satellite-based potato yield models rely on vegetation indices (VIs). For example, [37] applied NDVI, while [38] used MODIS-derived indices such as LAI, NDVI, and fPAR to predict potato yield. By contrast, only a few studies [20] have explored the use of raw spectral reflectance for potato yield prediction. Previous research has mainly focused on red and near-infrared-based indices, yet indices and bands from the red-edge region are closely linked to chlorophyll content and nitrogen status, and may offer improved yield prediction accuracy [39,40].

Sentinel-2 imagery provides continuous spectral information that can be used either directly as raw bands or as derived vegetation indices (VIs). VIs capture pigment absorption and canopy water content, making them useful for detecting stress caused by water deficit, temperature extremes, or infections in the potato crop [20], while raw bands preserve complementary reflectance signals linked to crop growth stage. Both inputs offer critical information about potato crop conditions relevant to yield development.

Several techniques were used to explore the relationship between potato yield and satellite-derived data. Statistical models have been widely applied in yield prediction [41,42]. While linear models offer interpretability by quantifying the additive effects of predictors, they often struggle when relationships are highly non-linear [43,44]. To address this limitation, machine learning (ML) models have gained popularity for capturing complex patterns, even with large and high-dimensional datasets [45]. Unlike classical statistical models, ML algorithms do not assume a fixed data structure; instead, they learn relationships directly from the data [46]. Three widely used ML algorithms in crop yield studies are Random Forest (RF), Support Vector Machine (SVM), and Extreme Gradient Boosting (XGB).

Deep learning (DL) has also shown strong potential by automatically learning complex spatio-temporal patterns from large datasets, which are often difficult to model with traditional approaches [47]. DL algorithms, including Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), have been successfully applied for crop yield modeling [48]. RNNs are particularly effective in yield prediction tasks that involve sequential or time-series data, such as vegetation indices or spectral information over a growing season [49]. DL models often outperform ML when large datasets are available [50]. However, some studies have reported that ML can perform equally well or better at certain scales, depending on the dataset size and features [51].

Despite these advances, a persistent challenge in applying deep learning to yield estimation is the requirement for large labeled datasets, which are rarely available at the field scale. Most studies demonstrating the success of CNNs or RNNs rely on extensive regional or multi-year datasets, whereas single-season, field-level applications often face data scarcity. Recent work suggests that when training data are limited, the choice of model complexity plays an important role in overall performance. Under such conditions, simpler network architectures can achieve performance comparable to that of deeper and more parameterized models, particularly when appropriate regularization and data augmentation are applied [52]. These findings motivate the development of lightweight deep learning approaches. It therefore raises an important question: Can deep learning architectures be designed to perform effectively under small data regimes? Addressing this challenge requires models that are not only expressive enough to capture temporal dependencies in crop growth but also efficient with limited training samples.

In this study, we introduce two lightweight deep learning architectures tailored for field-scale, single-season potato yield estimation. The first model, LiteTemporalConv, is a compact temporal convolutional network designed to emphasize stability, parameter efficiency, and robustness when sample sizes are small. The second, MS-ConvBiGRU-Attn, adopts a hybrid structure that integrates multi-scale convolution, recurrent layers, and an attention mechanism to represent both short-term spectral variations and longer-term phenological patterns. By comparing a simple model with a more complex temporal encoder, this study also explores how model complexity affects performance on small agricultural datasets. In comparison with established machine learning baselines (RF, SVM, XGB, and PLS), we provide a systematic evaluation of these deep learning approaches' strengths and limitations in addressing the challenges of small, high-dimensional agricultural datasets.

2. Materials and methods

2.1. Study area

Ground-truth yield data were collected from potato fields located in the south-western part of the Veneto Region (Italy), centered at 11° 19′ 31″ E, 45° 17′ 36″ N. The Veneto region in north-eastern Italy experiences a varied climate due to its diverse topography. It spans 18,398.9 km^2 and ranges from the Austrian Alps to the Po River Delta. Its geography encompasses mountainous terrains, fertile plains, and hilly regions. The area is crossed by major rivers such as the Po and Adige, and borders the Adriatic Sea with 200 km of coastline. The study area (Fig. 1) is located in the Venetian plain, with a semi-continental climate, marked by hot, humid summers with average temperatures ranging from 25 to 30 °C and cold, foggy winters with temperatures around 0 to 5 °C. Annual precipitation varies between 800 and 1200 mm, with rainfall in spring and autumn. These climatic conditions strongly affect agricultural activities in the region. Potato cultivation in the Veneto region generally follows a seasonal cycle, with sowing typically occurring between late March and early May. However, this timeline can vary depending on the crop variety; some of the early-season varieties are sown as early as February. Similarly, harvesting times differ across varieties. Potatoes

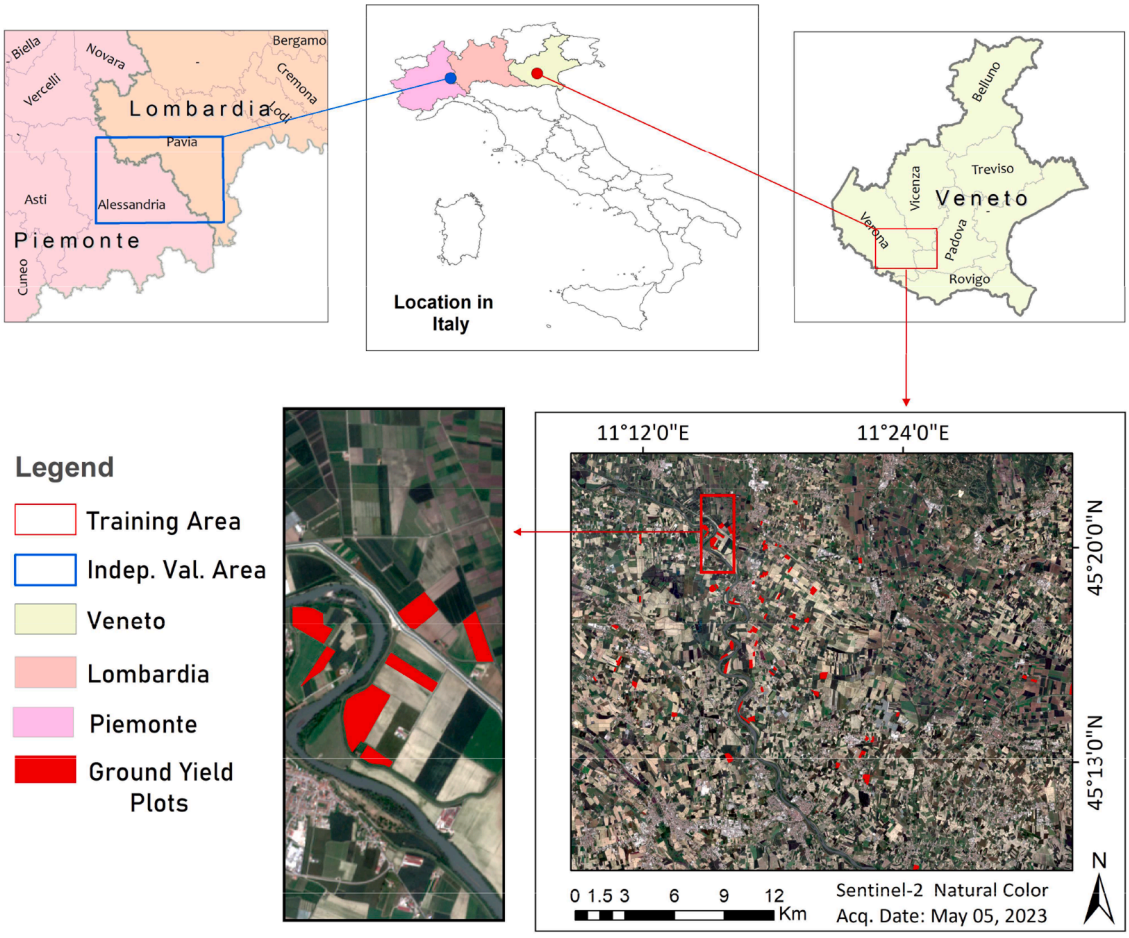

Fig. 1. Spatial distribution of training (Veneto) and independent validation (Lombardia–Piemonte) sites, including Sentinel-2 natural color composite for the training area and representative zoomed ground yield plots.

that are planted early start harvesting in early June, while most of the varieties reach maturity by July. Mid-season varieties are usually harvested between late September and October, whereas late-sown potatoes are typically harvested in November. In addition to the primary study area in Veneto, an independent validation region was considered to assess the spatial transferability of proposed models. This secondary region comprises potato fields located across southern Lombardia and eastern Piemonte in northern Italy (Fig. 1), geographically separated from the Veneto training area.

2.2. Satellite data

The Sentinel-2 satellite constellation was developed under the European Copernicus Earth observation program. It plays a key role in agricultural applications due to its high-temporal frequency and multispectral imaging capabilities [88]. It operates through twin satellites, Sentinel-2A and Sentinel-2B, equipped with the Multispectral Instrument (MSI), which captures data in 13 spectral bands across the visible, near-infrared, and shortwave infrared regions [53]. With a revisit period of five days, Sentinel-2 provides timely observations for crop monitoring. For this study, Sentinel-2 tile T32TPR was selected to cover the potato cultivation regions within the study area. A total of eight Level-2A Bottom-of-Atmosphere (BOA) images were obtained from the Copernicus Open Access Hub, according to the phenological stages of potato growth, including emergence, canopy development, tuber initiation, and maturity. Data acquisition also covered from early planting to post-harvest (Fig. 2). Only scenes with <5% cloud cover over the area of interest were used to maintain data quality. Key spectral bands from each acquisition date were extracted to use for further vegetation indices development and eventually yield prediction.

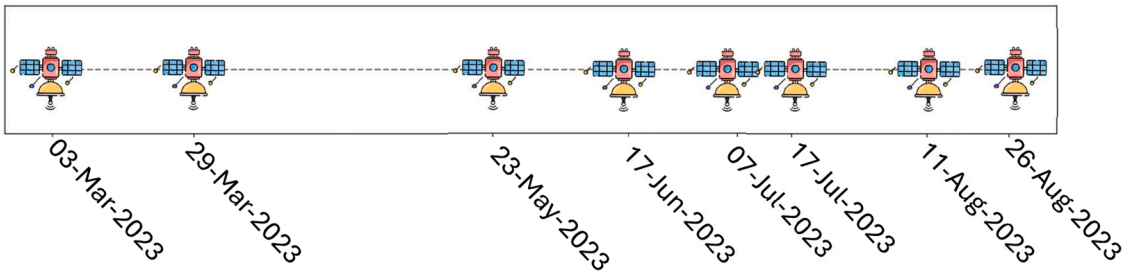

Fig. 2. Satellite data acquisition dates for the potato crop from planting to harvest.

2.3. Ground yield data

Ground yield data were collected from a systematic survey conducted on 114 potato fields located in the study area of Fig. 1, thanks to the collaboration of the Pizzoli Company (Budrio, Italy). Data on the yield harvested for each field were collected during harvest time over a range of different potato varieties, including Innovator, Daisy, and Alverstone R. Each field polygon was accurately delineated and assigned its corresponding fresh matter yield value, measured in tonnes (ton) per hectare (ha). A descriptive statistical summary (Table 1) presents the minimum, maximum, mean, and standard deviation for yield ton per hectare, which provides an overview of the distribution and variability within the dataset.

3. Methodological approach

The methodological approach followed a step-by-step process, starting with the collection of satellite images and preprocessing. Once the images were ready, key features were prepared by calculating vegetation indices from the raw spectral bands. After that, feature selection was carried out to focus on the most relevant inputs for the analysis. The study then advanced into two main experimental streams: (i) the development and evaluation of two specialized deep learning architectures LiteTemporalConv and MS-ConvBiGRU-Attn tailored for potato yield estimation, and (ii) the training of baseline machine learning models (RF, XGB, SVM, and PLS). Performance of all models was assessed and compared using $R^2$and RMSE metrics for single-season, field-scale potato yield prediction (Fig. 3). All analyses were conducted in Python (3.9.20) using pandas, NumPy, and Matplotlib/Seaborn for data handling and visualization, and scikit-learn, statsmodels, XGBoost, and TensorFlow/Keras for model development. Hyperparameters were tuned via grid search for ML models, deep learning models used manually adjusted settings. Efficiency analysis showed that deep learning models were more efficient on a GPU, with faster training and lower memory use than ML approaches. All experiments were performed on a workstation with an Intel i7–12,700, 32 GB RAM, NVIDIA RTX 3060.

3.1. Satellite image preprocessing

Sentinel-2 images were pre-processed automatically using the ESA SNAP Graph Processing Tool (GPT) workflow. This method provides a robust and time-efficient way to handle time-series satellite imagery, which ensures every file is processed consistently and reproducibly. A Linux-based shell script was developed to execute a custom processing graph. This graph included key steps like resampling image bands to 10-meter resolution and clipping the imagery to the boundaries of the study area. The script runs automatically all the necessary operations defined in the SNAP batch processing graph, which significantly reduces the manual effort and processing time. Compared to traditional methods, this automated approach allows for efficient, large-scale processing with minimal user input. The processed images were then used to calculate vegetation indices and to preparation of feature combinations for yield modelling.

Table 1
Ground yield data summary statistics.

Statistics	Area (ha)	Yield (ton/ha)
Minimum	0.59	30.13
Maximum	16.56	63.20
Sum	471.39	4532.47
Mean	5.29	50.11
Std	4.19	8.57

3.2. Feature extraction

In the feature extraction stage, two main datasets were developed to support potato yield prediction. The reflectance values from selected Sentinel-2 spectral bands (Table 2) and a set of 17 vegetation indices were generated for potato yield modelling. The Sentinel-2 bands were selected for their sensitivity to plant traits associated with crop health, particularly those related to photosynthesis, pigment levels, and water content [20]. In Table 3, the indices were selected to ensure consistency with previous applications in potato and related crop yield estimation studies. Collectively, they represent key physiological attributes directly relevant to potato growth and yield prediction. This combination of raw spectral data and crop-specific indices allowed for a more comprehensive representation of crop conditions across developmental stages.

3.3. Data preparation and feature selection

For model development, we first prepared a structured dataset at the parcel level. A total of 114 potato field polygons with observed yields (t/ha) were used as spatial units of analysis. For each polygon, Sentinel-2 surface reflectance data were retrieved on eight cloud-free dates covering the main phenological stages of the 2023 season. From these scenes, we derived two complementary sets of predictors: (i) the ten Sentinel-2 spectral bands spanning visible to shortwave infrared wavelengths, and (ii) a set of 17 vegetation indices capturing canopy greenness, structure, and pigment composition. To obtain representative values per field and date, we computed the mean reflectance or index value of all valid pixels within each polygon. This procedure yielded a spatiotemporal dataset with the structure 114 fields × 8 dates × 27 features, where the feature dimension combines both raw bands and derived indices. These temporally ordered feature sequences provided the input representation for all subsequent sequence-based models and formed the basis for the feature selection analysis.

Deep learning models often perform better when trained on large amounts of data, but adding too many features can have drawbacks [66]. As the number of features grows, models face the curse of dimensionality, which can lead to overfitting and introduce noisy or redundant information that lowers accuracy [67]. Therefore, careful feature selection becomes important in deep learning. Focusing on the right number of meaningful features, models can capture the true patterns in the data more effectively.

A Grouped TreeSHAP [68], Stability Selection (GTSS) procedure was employed to identify a compact and leakage-safe subset of spectral bands, vegetation indices, and acquisition dates relevant for field-level yield prediction. Each field was represented as a tensor

$$X \in R^{N\times T\times C} \tag{1}$$

where N denotes the number of fields, T the number of acquisition dates, and C the number of input channels. For each date $t \in \{1,\ldots\ldots,T\}$, the input consisted of 10 Sentinel-2 spectral bands (SBs) and 14 vegetation indices (VIs), organized as $[SB_{1:10,}\ VI_{1:14}]$. The prediction target was field-level yield (y). For generality, we denote the index grid as $\left\{x_{j,t}^{VI}\right\}$ with $j \in \{1,\ldots\ldots,M_{VI}\}$ and the band grid as $\left\{x_{b,t}^{SB}\right\}$ with $b \in \{1,\ldots\ldots,M_{SB}\}$. To evaluate feature importance, we considered three grouping strategies: i) VI groups – all T columns of a given vegetation index, j: as $\left\{x_{j,t}^{VI}\right\}$ ii) SB groups: all T columns of a given spectral band b iii) Date groups: all VIs or SBs at a specific date t.

A nested GroupKFold strategy was adopted, ensuring that all feature selection occurred exclusively within the outer training fold, while the outer test fold remained unseen until final evaluation. Within each outer training fold, we performed $B = 50$ field-wise bootstraps. For each bootstrap: A compact gradient-boosted tree model (XGB-style, depth 3–4, learning rate = 0.05, up to 500 trees with early stopping, subsample

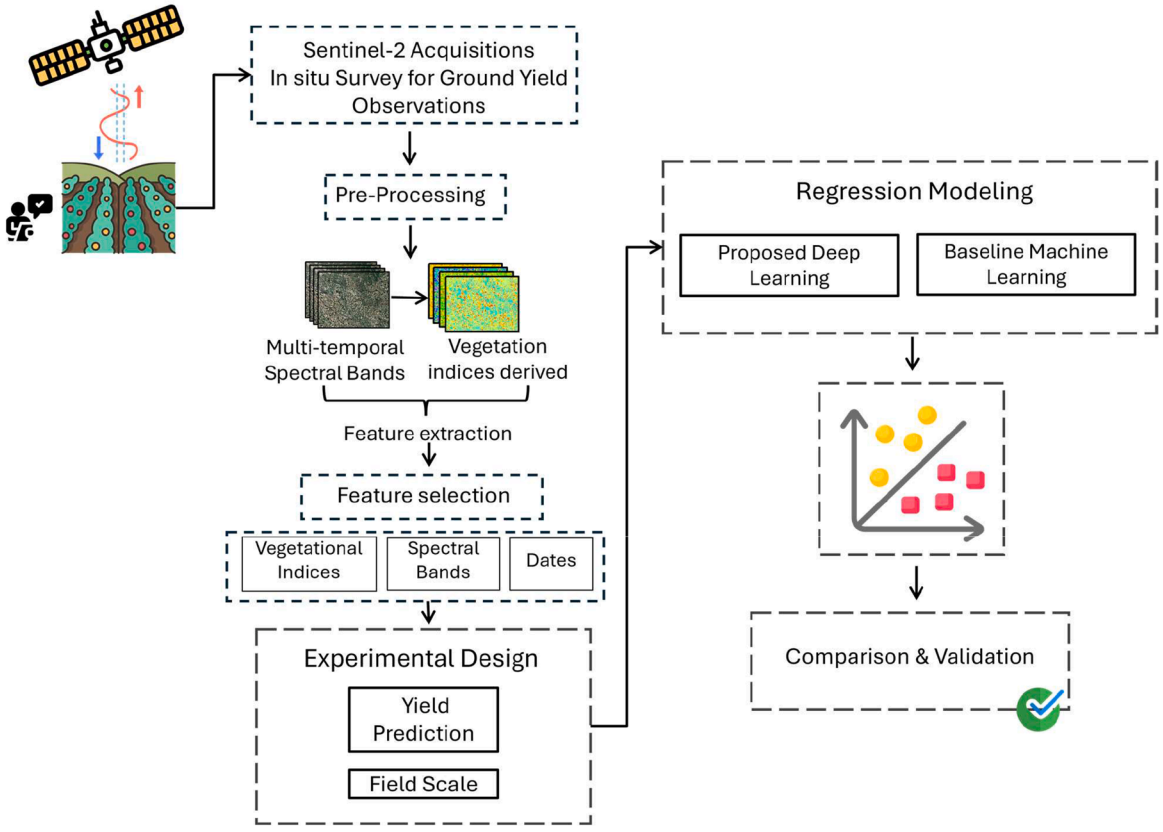

Fig. 3. Methodological workflow carried out for this study.

Table 2
Sentinel-2 bands selected for this study.

Band #	Central λ (nm)	Spectral Width (nm)	Spatial resolution (m)
B2	490	65	10
B3	560	35	10
B4	665	30	10
B5	705	15	20
B6	740	15	20
B7	783	20	20
B8	842	115	10
B8A	865	20	20
B11	1610	90	20
B12	2190	180	20

and column-subsample = 0.9) was fitted. TreeSHAP values were computed on out-of-bootstrap samples (not used in model fitting). The mean absolute SHAP contribution per feature was calculated and aggregated to group scores:

$$\text{Index (VI) score} = S_j^{\text{VI}} = \sum_{t=1}^{T} E\left[\left|\text{SHAP}\left(x_{j,t}^{\text{VI}}\right)\right|\right] \quad (2)$$

$$\text{Band (SB) score} = S_b^{\text{SB}} = \sum_{t=1}^{T} E\left[\left|\text{SHAP}\left(x_{b,t}^{\text{SB}}\right)\right|\right] \quad (3)$$

$$\text{Date score (VIs)} = S_t^{\text{Date, VI}} = \sum_{j=1}^{M_{VI}} E\left[\left|\text{SHAP}\left(x_{j,t}^{\text{VI}}\right)\right|\right] \quad (4)$$

$$\text{Date score (SBs)} = S_t^{\text{Date, SB}} = \sum_{b=1}^{M_{SB}} E\left[\left|\text{SHAP}\left(x_{b,t}^{\text{SB}}\right)\right|\right] \quad (5)$$

To ensure robustness, shadow groups were created by permuting all columns within each group (e.g., all dates of a specific VI). A group was considered a "hit" if its score exceeded the maximum shadow score in that bootstrap. Stability was then defined as the fraction of bootstraps in which a group registered a hit, and only groups with stability ≥ 60% were retained.

3.4. Proposed deep learning framework

We develop a two-stage framework for potato yield estimation from multi-temporal Sentinel-2 data. A light, shared front-end first standardizes the per-timestamp feature vectors; this output is then passed to one of two temporal encoders: LiteTemporalConv, a compact convolutional network, or multi-stage ConvBiGRU-Attn, which combines convolution, recurrent modeling, and attention. The shared front-end ensures both models operate on the same representation, allowing a fair comparison of their predictive yield capabilities.

To assess the temporal contribution of each acquisition date to the encoders, we computed Permutation Importance on R^2 (ΔR^2). For a trained model, we first obtained the baseline R^2 on the (scaled) training set. Then, for each date t, we permuted all features from that date across fields (bands + indices; temporal order preserved elsewhere), re-ran predictions, and calculated $\Delta R^2 = R^2_{base} - R^2_{perm\ (t)}$ .Larger ΔR^2 indicates a greater marginal contribution of that date to the model's fit. Based on this analysis, an additional experiment was performed in which models were retrained using a subset of temporally selected inputs and evaluated on observations corresponding to different temporal stages of the growing season.

3.4.1. Structure of temporal channel projection

The Temporal Channel Projection (TCP) is a minimal front-end that applies the linear transformation to the feature vector at each time step while keeping the sequence length unchanged. Practically, it is implemented as a time-distributed Dense layer with kernel L2 regularization.

Table 3
Description and calculating formulas for vegetational indices from Sentinel-2 spectral bands.

Index	Abbreviation	Formula	Source
PPI	Potato Productivity Index	*(B3 / (B2 + B4)) + ((B9 / B8A) * (B3 / B5))*	[20]
NDVI	Normalized Difference Vegetation Index	$(B8 - B4)/(B8 + B4)$	[54]
LAI	Leaf Area Index	*Neural network inversion of PROSAIL RTM using Sentinel-2 TOC reflectance*	[55]
FVC	Fractional Vegetation Cover	*Neural network inversion of PROSAIL RTM using Sentinel-2 TOC reflectance*	[55]
MTCI	Meris Terrestrial Chlorophyll Index algorithm	*(B6 − B5) / (B5 − B4)*	[56]
IRECI	Inverted Red-Edge Chlorophyll Index	$(B7 - B4)/ (B4 /B4)$	[57]
MCAR	Modified Chlorophyll Absorption Ratio	*[(B5 - B4) - 0.2 * (B5 - B3)] * (B5 / B4)*	[58]
GNDVI	Green Normalized Difference Vegetation Index	*(B8 − B3) / (B8 + B3)*	[59]
SAVI	Soil Adjusted Vegetation Index	*1.5 * (B8 - B4) / (B8 + B4 + 0.5)*	[60]
ND45	Normalized Difference 450 nm Index	*(B5 - B4) / (B5 + B4)*	[61]
CRI2	Carotenoid Reflectance Index	*(1 / B2) - (1 / B5)*	[19]
AR12	Anthocyanin Reflectance Index	*(1 / B3) - (1 / B5)*	[19]
PSRI	Plant Senescence Reflectance Index	*(B4 - B2) / B6*	[62]
NDRE	Normalized difference red edge	*(B8 - B5) / (B8 + B5)*	[63]
CLG	Green Chlorophyll index	*(B7 / B3) - 1*	[64]
FPAR	Fraction of Absorbed Photosynthetically Active Radiation	*Neural network inversion of PROSAIL RTM using Sentinel-2 TOC reflectance*	[55]
REIP	Red-Edge Inflection Point Index	*705+35* ((B4+B7)/2-B5)/ (B6-B5)*	[65]

As shown in Fig. 4(a), TCP introduced a lightweight front-end that performs a time-shared linear projection of per-step features from $R^{F_{in}}$ to $R^{F_{out}}$ while preserving sequence length.

$$\text{For } X \in R^{T \times F_{in}} \text{ with } T = 6,\ F_{in} = 18 \tag{6}$$

The module computes

$$\widetilde{X} = XW + 1_T b^T \in R^{T \times F_{out}},\ \widetilde{X} \in R^{6 \times F_{out}} \tag{7}$$

The result $\widetilde{X} \in R^{T \times F_{out}}$ is then fed to the encoder. This layer serves three roles: it mixes correlated spectral inputs into a compact set of features, suppresses noisy or redundant directions [69], and conditions the data for downstream temporal layers without imposing assumptions on the time axis. We tune it by cross-validation over a small grid (e.g., 8, 12, 14, 16, 24) and fix the best setting for all subsequent training and evaluation. In our experiments, adding TCP provides consistent gains in R^2 and RMSE with negligible computational overhead. After TCP, the resulting sequence is forwarded to one of two alternative encoders that differ in how they capture spatiotemporal signals for yield prediction.

3.4.2. *LiteTemporalConv*

The LiteTemporalConv begins with a lightweight TCP output layer that harmonizes inputs to a fixed temporal–feature grid. In our implementation, inputs are mapped to 14 features per step and 6-time steps, ensuring a consistent interface for the encoder. As shown in Fig. 4(b), the core of LiteTemporalConv is a shallow stack of one-dimensional convolutions operating along the temporal sequence with the same padding Conv1D 64 filters, kernel size 3, ReLU Conv1D 64 filters, kernel size 3, ReLU Layer Normalization → Dropout ($p = 0.1$) Conv1D 128 filters, kernel size 3, ReLU. This sequence favors small kernels to capture short-to-mid-range temporal signals while keeping the parameter count low [70]. With unit stride, the arrangement provides an effective receptive field suited to sparsely sampled seasonal data. Layer normalization stabilizes activations under small batch sizes, and dropout offers additional regularization [71]. The convolutional feature map is summarized by global average pooling across time to produce a fixed-length representation. A minimalist prediction head follows: Dropout ($p = 0.20$) → Dense (64 units, ReLU, L2 regularization) → Dense (1 unit) for yield

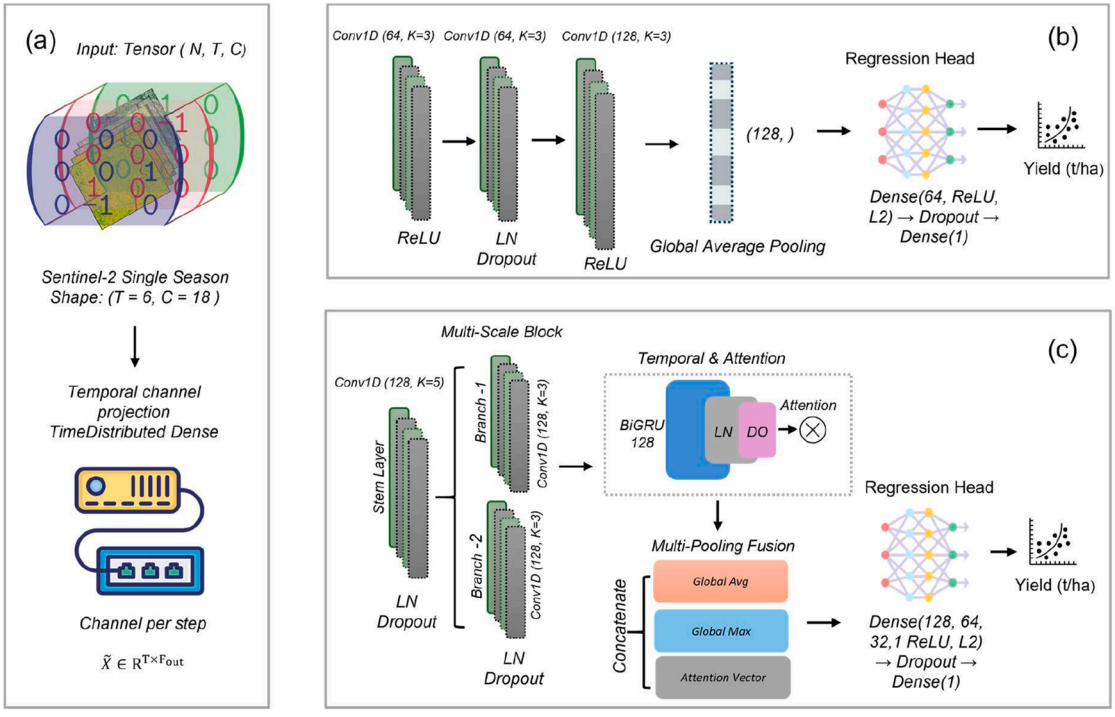

Fig. 4. Structure of the proposed deep learning framework for crop yield prediction. (a) Temporal Channel Projection (TCP) (b) LiteTemporalConv (c) MS - ConvBiGRU-Attn.

regression. The encoder output h (6 × *C*), pooled by GlobalAveragePooling1D or a (C) vector, feeds a compact MLP head that maps global average pooling sequence aggregation z to the yield prediction $\widehat{y}$.

$$z = \mathrm{ReLU}(W_{1h} + b_1),\ \widehat{y} = W_2 z + b_2 \tag{8}$$

Where $W_1 \in \mathrm{R}^{64 \times C_3}$, $W_2 \in \mathrm{R}^{1 \times 64}$, C_3 =the number of channels coming out of the last Conv1D layer, b_1 = the bias vector for the first dense layer, b_2 = the bias vector for the final output layer. Global pooling limits overfitting, removes sensitivity to minor variations in sequence length, and keeps computation low. LiteTemporalConv is intended to be efficient, stable, and easy to train. The adapter enforces a uniform input geometry; the shallow Conv1D stack emphasizes local temporal patterns without excessive capacity [72]; and the pooling-plus-dense head offers a clear, regularized path to the target. Together, these choices yield a practical encoder for multi-temporal agricultural signals, well-suited to datasets with limited sample sizes.

3.4.3. MS - ConvBiGRU-Attn

In this study, the ConvBiGRU-Attn model is organized as a staged sequence encoder that couples multi-scale convolution, bidirectional recurrent modeling, and a lightweight attention mechanism Fig. 4(c). The network comprises an input TCP, a convolutional stem with parallel branches, a bidirectional GRU block, an attention layer for temporal weighting, a multi-pooling fusion module, and a compact projection head followed by a shallow regression head. Model begins with a multi-scale convolutional stem with a moderate-width temporal convolution (128 filters, kernel size 5) to capture short-range dynamics [73], followed by Layer Normalization and dropout to stabilize activations. To enhance feature diversity without inflating parameters, two parallel branches operate in tandem: (i) a standard Conv1D with 128 filters and kernel size 3, and (ii) a dilated Conv1D with 128 filters, kernel size 3, and dilation rate 2 to extend the temporal receptive field. The branch outputs are concatenated and compressed with a pointwise (1 × 1) convolution to 192 channels, then normalized. This multi-scale design extracts complementary temporal signals while maintaining computational balance. Downstream of the stem, a bidirectional GRU (128 units per direction) models longer-range dependencies by processing the sequence in both forward and backward directions [71]. The layer returns a time-aligned feature map, preserving step-wise information for subsequent weighting. Layer Normalization and dropout are applied to improve stability on small datasets. An attention layer assigns a data-driven importance score to each time step [71] and forms a context vector as the weighted combination of the GRU outputs. Placed immediately after the recurrent block and before pooling, this component highlights phenological periods most informative for yield estimation and provides an interpretable summary of the sequence. To increase robustness, the attention context is fused with global average pooling and global max pooling computed over the same GRU outputs. Concatenating these three summaries, attention context, average pool, and max pool balances typical temporal patterns with salient peaks while retaining the focused signal emphasized by attention [73]. The fused vector is passed through a compact projection layer (Dense 256, ReLU), followed by Layer Normalization and comparatively strong dropout. The resulting feature embedding constitutes the encoder output and is designed to be expressive yet well-regularized. For completeness, the encoder feeds a conservative regression head consisting of sequential dense blocks (128 → 64 → 32 units with ReLU), each interleaved with normalization or dropout, and a final linear unit that produces the yield estimate.

$$\widehat{y} = W_2 ReLU(W_1 f + b_1) + b_2 \tag{9}$$

Where $\widehat{y}$ = predicted yield, f = multi pooling fusion, W_1 = Weight matrix of the first fully connected (Dense) layer, W_2 = weight matrix of the final output layer, b_1 = the bias vector for the first dense layer, b_2 = the bias vector for the final output layer.

The head is intentionally shallow to limit overfitting while leveraging the rich temporal representation learned by the encoder. Overall, ConvBiGRU-Attn integrates three complementary capabilities: (i) multi-scale convolution to harvest local temporal cues, (ii) bidirectional recurrence to capture longer-range phenology, and (iii) attention to prioritize informative growth windows. The multi-pooling fusion and staged regularization (Layer Normalization and dropout throughout) further improve stability and generalization on limited training data.

3.4.4. Parameter initialization and optimization

Both models employed Keras defaults for parameter initialization: convolutional and dense kernels were drawn from Glorot uniform distributions, biases were initialized to zero, and layer normalization parameters were set with scale $\gamma = 1$ and shift $\beta = 0$. Additionally, the recurrent unit within the MS-ConvBiGRU-Attn encoder utilized orthogonal initialization for the recurrent kernel, thereby enhancing stability during sequence modeling. Optimization was performed using the Adam algorithm with the Huber loss function. Regularization was applied through L2 wt penalties and dropout layers to improve generalization under limited data conditions. Training was conducted for up to 120 epochs with a batch size of 8, and early stopping based on validation loss was used to prevent overfitting. An initial learning rate of 3×10^{-3} was used. Data augmentation was applied during training, including temporal shifting, per-feature scaling, additive Gaussian noise, feature dropout, and mixup. All augmentation procedures were implemented with fixed random seeds to ensure reproducibility. A summary of the model configurations and training parameters is provided in Table 4.

3.5. Baseline machine learning models

Four widely acknowledged machine learning regression algorithms were employed to capture complex, nonlinear relationships between spectral predictors and potato yield: Random Forest (RF), developed by [74], is an ensemble learning method that builds multiple decision trees and aggregates their predictions. It effectively handles variable interactions and reduces overfitting [75]. Partial Least Squares (PLS) Regression was also employed as a dimensionality reduction technique that projects the original variables into a smaller set of components while preserving the relationship with the target variable [76]. It is particularly useful when we need to predict a dependent variable from a large set of independent variables [77]. Support vector regression (SVR), based on the theory of Support Vector Machines, was introduced by [78]. It constructs a regression function by minimizing error within a specified margin, offering robustness in high-dimensional feature spaces. Extreme gradient boost (XGB) regression, proposed by [79], is a scalable and efficient implementation of gradient boosting, which

Table 4
Parameter optimization of the proposed deep learning framework.

Aspect	LiteTemporalConv	MS - ConvBiGRU-Attn
Regularization	L2 = 1×10^{-4}; Dropout p = 0.1–0.2	L2 = 1×10^{-3}; Dropout p = 0.2–0.4
Optimizer	Adam ($\eta = 3 \times 10^{-3}$, β_1=0.9, β_2=0.999, ε=1e^{-7})	Adam ($\eta = 3 \times 10^{-3}$, β_1=0.9, β_2=0.999, ε=1e^{-7})
Loss	Huber ($\delta = 1.0$)	Huber ($\delta = 1.0$)
LR schedule	ReduceLROnPlateau (factor 0.5, patience 20)	ReduceLROnPlateau (factor 0.5, patience 20, min_lr=1e^{-6}, min_delta=0.001)
Early stopping	Patience = 5, restore best weights	Patience = 5, restore best weights
Batch size	8	8
Epochs	120	120
Augmentation	Enabled: time shift, scaling, noise, feature dropout, mixup	Enabled: time shift, scaling, noise, feature dropout, mixup
Validation	k-fold CV ($k = 10$)	k-fold CV ($k = 10$)

optimizes prediction performance through additive tree models and regularization techniques. To further evaluate the effectiveness of the proposed lightweight architectures, two standard deep learning baselines were also implemented. A conventional one-dimensional convolutional neural network (CNN) was used to model temporal patterns through stacked convolutional layers, while a gated recurrent unit (GRU)-based model was employed to capture sequential dependencies in the input data [71]. These models represent commonly used deep learning approaches for temporal modeling and provide a reference for assessing performance.

3.6. Model evaluation using cross-validation and independent spatial testing

Primary model evaluation was conducted using grouped k-fold cross-validation. Cross-validation is widely recommended for model evaluation in small-sample settings. It has efficient use of limited data to improve predictive performance [80] demonstrated that cross-validation can yield more stable and accurate predictions when independent validation data is scarce. Similarly, [81] adopted a cross--validation–based strategy for cotton yield estimation under limited sample conditions.

In this study, all reported performance metrics were computed on held-out test folds obtained through grouped k-fold cross-validation at the field level. For each fold, models were trained on approximately (k–1)/k of the fields and evaluated on the remaining held-out fields.

Model performance was quantified using three statistical metrics: the coefficient of determination (R^2), the root mean square error (RMSE), and the relative root mean square error (rRMSE), calculated as follows:

$$R^2 = 1 - \left[\sum_{i=1}^{n}(y_i - \widehat{y}_i)^2\right] \Big/ \left[\sum_{i=1}^{n}(y_i - \overline{y})^2\right] \tag{10}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2} \tag{11}$$

$$rRMSE = \frac{RMSE}{\overline{y}} \cdot 100 \tag{12}$$

where y_i and $\widehat{y}_i$ denote the observed and predicted potato yields for field i, respectively, $\overline{y}$ is the mean observed yield across all fields, and n is the total number of fields in the dataset. Each evaluation metric presents how well the model performed. The R^2 value indicates the proportion of variation in the observed data that is explained by the predictions, while RMSE and rRMSE reflect the extent to which absolute and relative predicted values deviate from the actual observations [18]. Final performance metrics were reported as the mean and standard deviation across all folds.

To further examine the spatial transferability of the proposed framework, the two highest-performing models identified through cross-validation were additionally evaluated using an independent hold-out dataset. This validation experiment was conducted using an independent set of 20 potato fields in the validation test area (Fig. 1). For this independent evaluation, Sentinel-2 imagery from the same growing season was processed using the identical preprocessing workflow applied to the training data. Vegetation indices and spectral bands were computed following the same procedures. The feature subset identified through the GTSS feature selection was retained to ensure consistency between training and validation inputs. Models trained on the Veneto dataset were directly applied to the validation fields without any retraining or fine-tuning, allowing an unbiased evaluation of their ability to generalize spatially. Predicted yields for the independent fields were then compared against observed field-level yield measurements to quantify performance under cross-regional conditions. This validation was not intended to represent a comprehensive multi-year or multi-environment assessment, but rather to provide an initial indication of spatial generalization capacity in the context of limited labeled data availability.

4. Results

4.1. Feature selection evaluation

The Grouped TreeSHAP Stability Selection (GTSS) analysis identified a phenology-aware subset of predictors for potato yield estimation based on sequence means across eight acquisition dates. Eight vegetation indices (VIs) achieved stability values $\geq$ 60% and were retained for further analysis. Ranked by GTSS score (sum of mean absolute TreeSHAP values across dates), the highest-scoring indices were PPI (100%), MCAR (90%), PSRI (90%), and LAI (80%), followed by ND45 (75%), CLG (70%), FPAR (65%), and NDVI (60%) (Fig. 5a).

Spectral band importance further corroborated this trend. Red-edge and near-infrared bands (B05, B07, B8A, B08) together with shortwave infrared (B11, B12) consistently dominated, while visible blue, green, and red bands (B02–B04) contributed moderately. Most bands registered stability values between 60 and 80%, reflecting repeated selection across bootstraps. For comparability across models, we retained all 10 Sentinel-2 bands in the modeling stage, although GTSS revealed a clear hierarchy in their contribution (Fig. 5a).

In terms of temporal selection, the six most stable dates for VIs were 17 June 2023 (90%), 23 May 2023 (90%), 07 July 2023 (≈90%), 17 July 2023 (80%), 26 August 2023 (70%), and 11 August 2023 (60%). A nearly identical pattern was observed for SBs, with the same six dates retained and top scores on 07 July 2023 (85%), 17 June 2023 (80%), and 23 May 2023 (80%) (Fig. 5b). These findings underscore that the most informative periods align with mid- to late-season phenological windows, when canopy development and biomass accumulation strongly influence final yield.

Overall, selecting eight vegetation indices and six dates while retaining all spectral bands provided concise yet information-rich input details. This approach improves the sample-to-parameter ratio for deep learning models, reduces the risk of overfitting under small datasets, and preserves the phenological signals most critical for yield prediction.

4.2. Training dynamics of deep learning models

The training behavior of the proposed deep learning architectures, LiteTemporalConv and MS-ConvBiGRU-Attn, was monitored to evaluate convergence stability and potential risks of overfitting. As shown in Fig. 6(a), the training loss dropped rapidly within the first 10 epochs, stabilizing below 0.03 thereafter. Validation loss followed a similar trajectory, converging near 0.04–0.05 with mild oscillations during mid-training. Importantly, the validation curve did not diverge from the training curve, suggesting that the combination of global average pooling, dropout, and L2 regularization effectively mitigated overfitting despite the small sample size.

In contrast, the MS-ConvBiGRU-Attn model, trained with a batch size of eight for 120 epochs, displayed a more gradual but consistent convergence. The training loss decreased from an initial value above 2.0 to ~0.25, while validation loss closely tracked this decline, reaching ~0.28 at the end of training (Fig. 6(b)). The close alignment of training and validation trajectories across all epochs demonstrates that the model generalized well and did not suffer from memorization of training samples.

These training histories confirm that both architectures achieved stable optimization under the chosen parameterization. LiteTemporalConv exhibited rapid early convergence with low variance, while MS-ConvBiGRU-Attn converged more gradually but demonstrated generalization capacity. The results highlight how architectural differences, in particular the presence of recurrent units and attention, impact the dynamics of model fitting while maintaining robustness under cross-

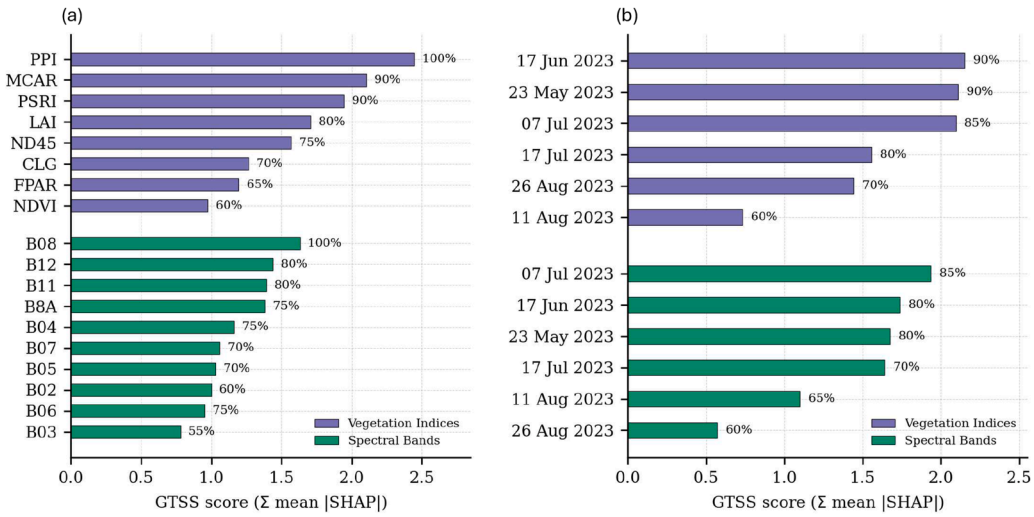

Fig. 5. GTSS feature selection results. (a) Stability-selected vegetation indices and Sentinel-2 spectral bands, ranked by GTSS score. (b) Stability selected acquisition dates for VIs and SBs.

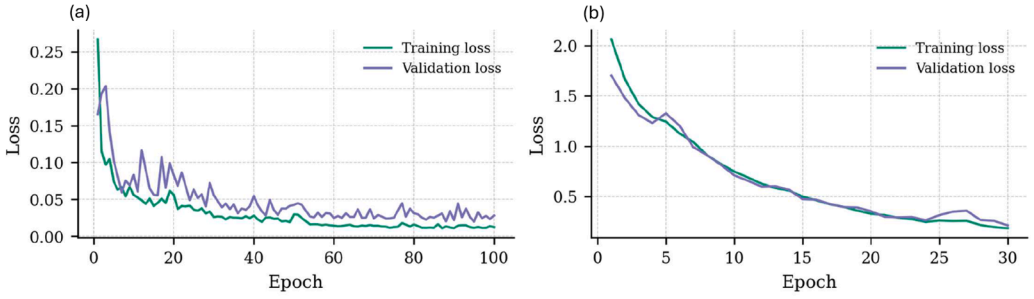

Fig. 6. Training and validation loss curves of the (a) LiteTemporalConv model (b) MS-ConvBiGRU-Attn model.

validation.

4.3. Performance of deep learning models

The proposed deep learning architectures, LiteTemporalConv and MS-ConvBiGRU-Attn, were evaluated to determine their capacity for yield prediction under the constraints of limited field data. Both models produced encouraging results, achieving high predictive accuracy while maintaining stability across cross-validation folds (Table 5). The LiteTemporalConv model attained an average coefficient of determination (R^2) of 0.84 ± 0.10 with a root mean squared error (RMSE) of 3.18 ± 0.54 t/ha, corresponding to a relative RMSE of 6.36%. These results demonstrate that the lightweight design comprising a compact stack of convolutional layers, normalization, dropout, and global average pooling enables the model to extract yield-relevant temporal patterns with better efficiency. The architecture's simplicity reduces the risk of overparameterization while retaining sufficient capacity to capture the short- to mid-range dynamics encoded in the seasonal satellite observations. This balance between model simplicity and representational strength is reflected in the consistent performance observed across cross-validation folds, as indicated by the low variance of the error metrics.

The MS-ConvBiGRU-Attn model achieved an R^2 of 0.82 ± 0.18 and an RMSE of 3.53 ± 1.94 t/ha, corresponding to a relative RMSE of 7.03%. Although its variance across folds was somewhat higher, the model integrates a multi-scale convolutional stem, bidirectional GRU units, and an attention mechanism, enabling it to capture both local spectral fluctuations and long-range phenological dependencies. In addition to cross-validation, both deep learning models were further evaluated using an independent validation test of a spatially different geographic region to assess their generalization beyond the training area (Table 5). LiteTemporalConv maintained performance on unseen data, achieving an R^2 of 0.86 with an RMSE of 4.21 t/ha, while MS-ConvBiGRU-Attn attained an R^2 of 0.84 and an RMSE of 5.10 t/ha. The consistency between cross-validation and hold-out results indicates that both architectures generalize well despite the limited sample size and different spatial locations. The scatter plots of observed versus predicted yields for cross-validation methods (Fig. 7) highlight the effectiveness of this design: predictions closely follow the 1:1 line, with strong calibration across the yield spectrum. While LiteTemporalConv exhibited tighter clustering in the upper yield range, MS-ConvBiGRU-Attn produced more nuanced fits at mid-yield levels, reflecting the complementary strengths of the two architectures.

Table 5
Performance of the proposed deep learning models R^2, RMSE (t/ha), and relative RMSE (%), expressed as mean ± standard deviation across 10 folds and hold-out independent test data.

Variant	R^2 mean ± std	RMSE (t/ha) mean ± std	rRMSE (%)	R^2 Independent Test	RMSE (t/ha) Independent Test
LiteTemporalConv	0.84 ± 0.10	3.18 ± 0.54	6.36	0.86	4.21
MS-ConvBiGRU-Attn	0.82 ± 0.18	3.53 ± 1.94	7.03	0.83	5.10

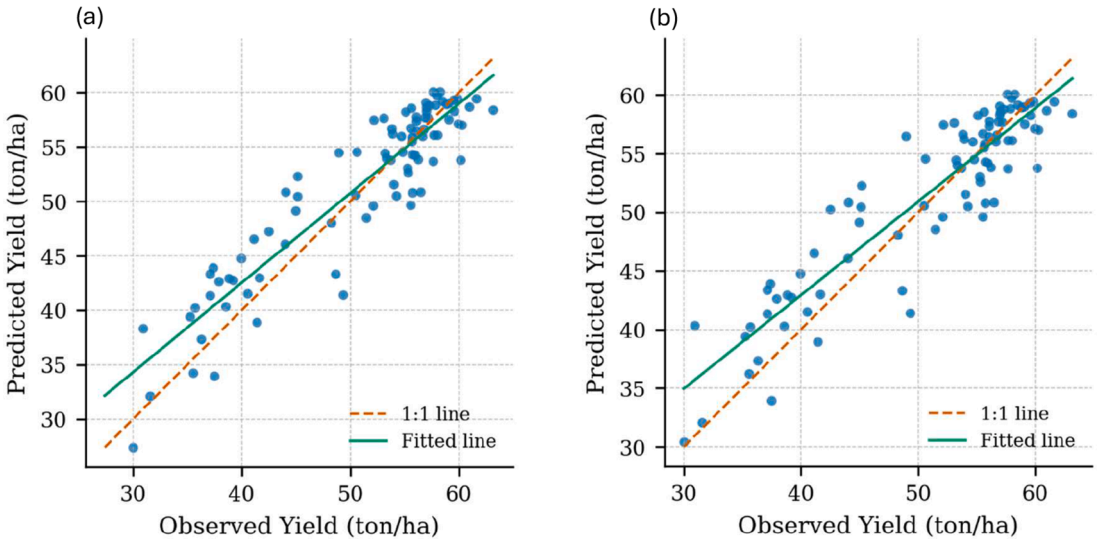

Fig. 7. Scatter plot for observed versus predicted potato yields (a) LiteTemporalConv (b) MS-ConvBiGRU-Attn. The fitted regression line aligns closely with the 1:1 reference line, demonstrating calibration across the yield range.

An important outcome of this study is that both models achieved reliable accuracy under a small data regime. This is particularly relevant in agricultural applications, where large labeled datasets are rarely available. The LiteTemporalConv model shows that a simple encoder can achieve stable and reproducible predictions with modest training samples. Meanwhile, the MS-ConvBiGRU-Attn architecture, though slightly more variable in small-sample settings, is designed to leverage richer temporal cues and has the potential to scale its performance further as more annotated data becomes available. Its inclusion of attention not only aids interpretability but also positions it to benefit disproportionately from larger training cohorts by learning refined temporal weighting schemes. In summary, both architectures demonstrated appropriate predictive power, surpassing typical expectations under limited data availability. LiteTemporalConv provides a fast, stable, and computationally efficient solution, while MS-ConvBiGRU-Attn offers a more flexible and expressive framework with the capacity to improve further under expanded datasets.

4.4. Comparison with baseline models

To evaluate the performance of the proposed deep learning architectures, we compared them with widely used baseline models, including Random Forest (RF), Partial Least Squares Regression (PLS), Support Vector Machine (SVM), Extreme Gradient Boosting (XGB), Convolutional Neural Networks (CNN), and Gated Recurrent Unit (GRU). The comparative results are summarized in Tables 5 and 6 and visualized in Fig. 8.

Across all metrics, the deep learning models consistently outperformed the baseline methods. LiteTemporalConv achieved the highest predictive accuracy with an R^2 of 0.84 ± 0.10 and an RMSE of 3.18 ± 0.54 t/ha, outperforming the best baseline, XGB (R^2 = 0.79 ± 0.11, RMSE = 4.9 ± 0.45 t/ha). The relative RMSE of LiteTemporalConv was 6.36%, compared with 9.8% for XGB and over 10% for Random Forest, confirming that the lightweight convolutional design can deliver under small-sample conditions.

Table 6
Performance of baseline machine learning models for single-season potato yield prediction. Metrics are reported as mean ± standard deviation across 10 folds, including coefficient of determination (R^2), root mean squared error (RMSE, t/ha), and relative RMSE (%).

Variant	R^2 mean ± std	RMSE (t/ha) mean ± std	rRMSE (%)
RF	0.77 ± 0.12	5.3 ± 0.87	10.1%
PLS	0.72 ± 0.10	6.5 ± 0.94	13.0%
SVM	0.69 ± 1.21	8.8 ± 0.62	17.0%
XGB	0.79 ± 0.11	4.9 ± 0.45	9.8%
CNN	0.68 ± 1.41	10.1 ± 0.94	20.1%
GRU	0.68 ± 1.32	9.8 ± 0.91	19.5%

The MS-ConvBiGRU-Attn model also performed well, with an R^2 of 0.82 ± 0.18 and RMSE of 3.53 ± 1.94 t/ha (relative RMSE 7.03%). While its variance across folds was higher than LiteTemporalConv, it still achieved better statistics as compared to baseline models by a substantial margin. This reflects the strength of its hybrid design, which integrates multi-scale convolutional filters, recurrent gating, and attention-based weighting to capture both short- and long-range temporal dependencies. Importantly, its performance suggests high potential for further improvements when trained on larger multi-year datasets, where its greater representational capacity can be more fully exploited.

Among the baseline models, XGB and RF were the highest, achieving R^2 values of 0.79 and 0.77, respectively. PLS and SVM performed notably worse, with R^2 values of 0.72 and 0.69, and relative errors exceeding 13% and 17%, respectively. These models also performed well due to their ability to fully capture the nonlinearities and complex temporal interactions embedded in the multispectral time series. To further contextualize the performance of the proposed lightweight architectures, additional baseline experiments were conducted using standard CNN and GRU models with conventional structures. Both models showed lower predictive performance compared to the proposed encoders and traditional machine learning baselines. The CNN and GRU achieved mean R^2 values of 0.68 ± 1.41 and 0.68 ± 1.32, respectively, with corresponding RMSE values of 10.1 ± 0.94 t ha^{-1} (rRMSE = 20.1%) and 9.8 ± 0.91 t ha^{-1} (rRMSE = 19.5%).

The results clearly demonstrate the advantage of the proposed design over conventional deep learn, regression and ensemble methods. While tree-based ensembles such as XGB provide relatively strong baselines, they may be constrained by the lack of explicit sequence modeling. In contrast, both LiteTemporalConv and MS-ConvBiGRU-Attn use temporal features learning directly from raw sequences, enabling better predictive accuracy and more stable calibration. This indicates that even under modest training sample sizes, deep sequence architectures can provide a scalable solution for crop yield prediction.

4.5. Temporal response of the deep learning encoders

Permutation-based ΔR^2 analysis revealed distinct temporal

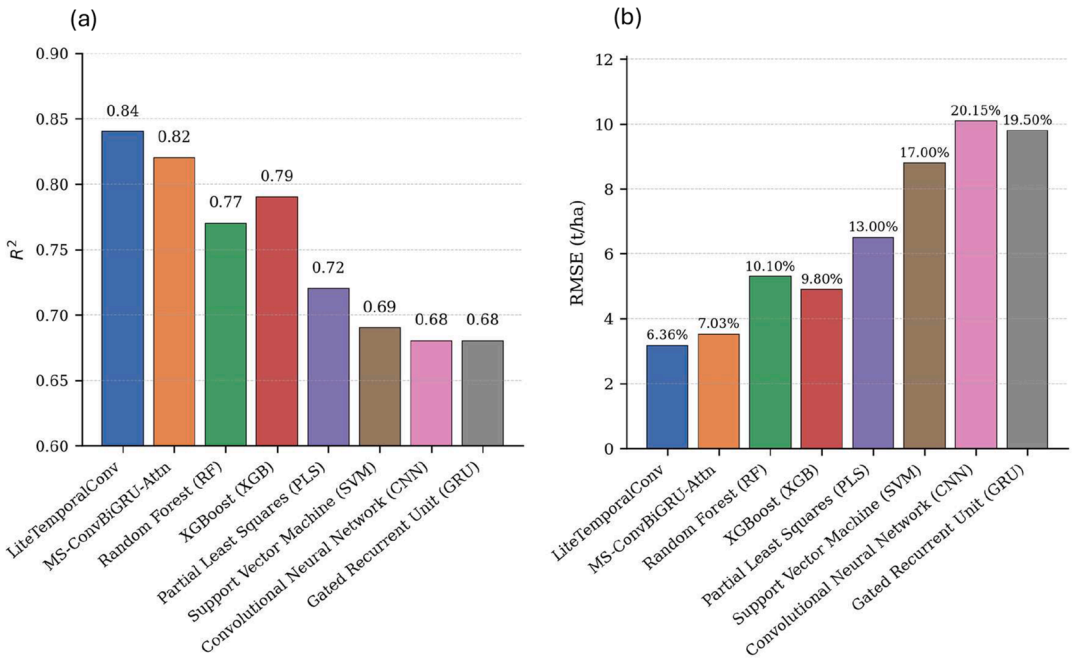

Fig. 8. Comparison of cross-validation performance across baseline machine learning models and the proposed deep learning models. (a) Mean R^2 values. (b) Mean RMSE values (t ha^{-1}) and relative RMSE (%).

sensitivity patterns for the two encoders.

Both encoders demonstrated a clear mid-season sensitivity, consistent with the feature selection analysis. The LiteTemporalConv model (Fig. 9(a)) exhibited the highest ΔR^2 around 17 June and 05 May, indicating strong responsiveness during early canopy development, with importance declining toward the late season. In contrast, the MS-ConvBiGRU-Attn model (Fig. 9(b)) peaked around 17 July and 07 July, reflecting its ability to capture extended temporal dependencies and later growth dynamics. Overall, both models emphasize the mid growth period as most critical for yield formation, the same period highlighted by GTSS, while early and very late dates carry comparatively less marginal information.

4.6. Late season yield prediction with deep learning models

The temporal response analysis revealed that certain acquisition periods contributed more strongly to model performance, with peak importance observed during the mid-season stages. Based on this observation, both deep learning models were retrained using inputs from mid-season data and then applied to Sentinel-2 observations acquired during the later stages of the growing season (11–26 August 2023). Results discussed the temporal robustness of the proposed deep learning models when evaluated for phenological stages not included during training.

Both models achieved stable predictions despite the exclusion of full-season data. The LiteTemporalConv model maintained a performance

Table 7
Performance of mid-season–trained deep learning models evaluated on late-season Sentinel-2 observations R^2, RMSE (t/ha), and relative RMSE (%), expressed as mean ± standard deviation across 10 folds.

Variant	R^2 mean ± std	RMSE (t/ha) mean ± std	RMSE (%)
LiteTemporalConv	0.80 ± 0.43	5.56 ± 2.54	11.12
MS-ConvBiGRU-Attn	0.76 ± 0.56	6.78 ± 3.25	13.56

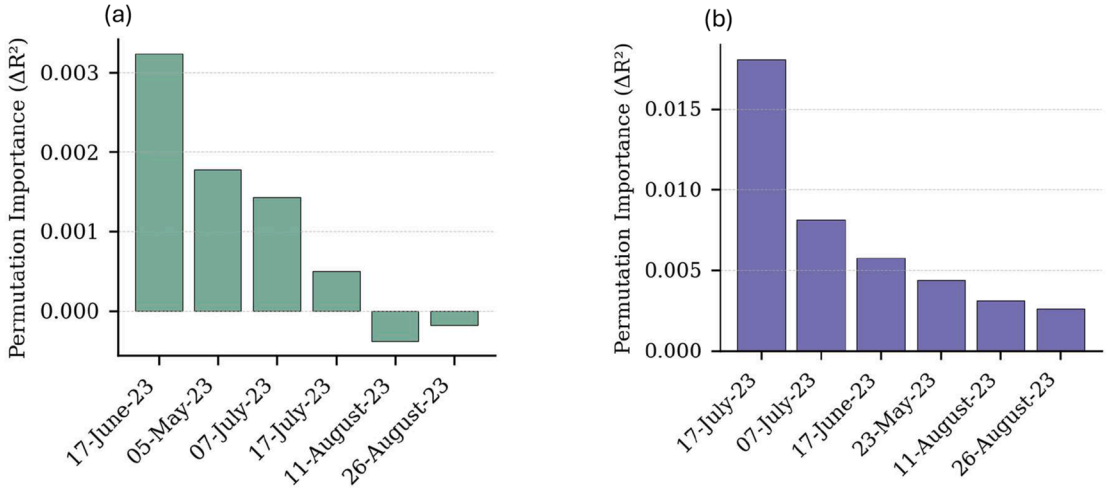

Fig. 9. Temporal permutation importance of encoder inputs on the training data. (a) LiteTemporalConv model (b) MS-ConvBiGRU-Attn model.

with an R^2 of 0.80 ± 0.43 and an RMSE of 5.56 ± 2.54 t/ha (11.12%), demonstrating its resilience under reduced temporal input (Table 7). In contrast, the MS-ConvBiGRU-Attn model showed (Table 7) a slightly lower R^2 (0.77 ± 0.56) and higher RMSE (6.78 ± 3.25 t/ha, 13.56%), reflecting its stronger reliance on longer temporal dependencies and late-season information. For the MS-ConvBiGRU-Attn model, all temporal dependencies contributed positively from Fig. 9(b); therefore, losing late-season data caused more drop in accuracy as compared to full-season data. The scatter plots showed the observed versus predicted yields (Fig. 10) for both models.

These results confirm that the compact temporal encoder (LiteTemporalConv) generalizes effectively under limited seasonal coverage, while the multi-scale recurrent–attention architecture (MS-ConvBiGRU-Attn) benefits more from extended phenological sequences. Overall, both architectures demonstrate predictive capability for pre-harvest yield estimation, supporting their applicability for early decision-making in precision agriculture.

5. Discussion

Field-level yield prediction significantly improves agricultural efficiency by providing support through reliable forecasts that aid targeted decisions, helping farmers to improve overall productivity [82,83]. In this study, we conducted a series of experiments to predict single-season potato yield by combining a leakage-safe feature selection stage with two purpose-built deep sequence encoders and a set of established machine learning baselines. The results indicate that both proposed architectures can achieve reliable predictive performance even under limited training data availability.

Deep learning can model complex, nonlinear links between satellite time series and yield. However, previous studies have also reported that their performance may be constrained in small-sample settings due to overfitting risks, optimization instability, and challenges in generalization [84,85]. In such contexts, classical machine learning models, including tree-based and regularized regression methods, often remain competitive, while the interpretability of deep models can be more difficult to establish when label datasets are scarce [86,87]. Against this background, our work adopts a small-data-oriented design. Several methodological choices contributed to the observed performance. The Grouped TreeSHAP Stability Selection (GTSS) framework reduced feature dimensionality by retaining phenology-relevant indices, spectral bands, and acquisition dates, thereby improving the balance between model complexity and available data. The GTSS selection highlights the discriminative value of chlorophyll- and canopy structure-sensitive indices (PPI, MCAR, PSRI, LAI), complemented by a red-edge contrast index (ND45). Classical broad-band indices such as NDVI and FPAR were also retained, though at lower ranks, suggesting that red-edge and canopy-structure cues provide stronger signals for yield estimation than general greenness metrics. Second, the use of a shared Temporal Channel Projection standardized the input representation across deep learning models. LiteTemporalConv employed a compact stack of temporal convolutions with normalization, dropout, and global pooling to constrain model capacity, while MS-ConvBiGRU-Attn incorporated multi-scale convolution, bidirectional GRUs, and an attention mechanism to weight informative temporal windows. Finally, regularization strategies, modest augmentation and grouped cross-validation were applied to mitigate data leakage and overfitting. Under these settings, LiteTemporalConv achieved the highest average cross-fold performance (mean R^2 = 0.84; RMSE = 3.18 t ha^{-1}; rRMSE = 6.36%), while MS-ConvBiGRU-Attn showed comparable results (mean R^2 = 0.82; RMSE = 3.53 t ha^{-1}; rRMSE = 7.03%). Independent spatial validation further suggests that both models maintain stable performance on unseen fields, with LiteTemporalConv showing particularly limited degradation between cross-validation and hold-out testing. These findings emphasize the importance of balancing model complexity with data availability in small-sample learning scenarios. Brigato and Locchi [52] demonstrated that, when training data are scarce, lower-complexity convolutional architectures can perform on par with, or even outperform, more complex deep learning models. Consistent with this observation, the proposed LiteTemporalConv model achieved accuracy comparable to the more complex MS-ConvBiGRU-Attn architecture, despite having substantially fewer parameters. Although LiteTemporalConv achieved slightly better performance in the single-season dataset, MS-ConvBiGRU-Attn remains conceptually important as a more expressive temporal architecture. MS-ConvBiGRU-Attn captures more extended temporal patterns through its recurrent and attention components, which may be advantageous when longer time series or multi-year data are available. The ΔR^2 in temporal response analysis further suggests that each model emphasizes different phases of crop development, indicating complementary temporal sensitivities rather than redundancy. The comparatively weaker results of the standard CNN and GRU models emphasize the role of architectural design under limited data conditions. By contrast, the proposed lightweight models employ a structured temporal encoding strategy that organizes

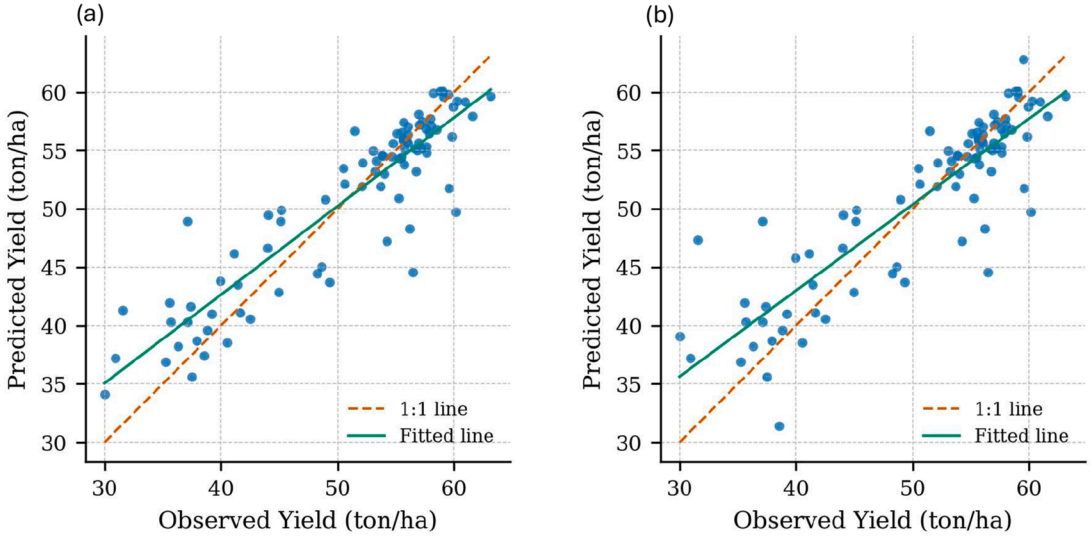

Fig. 10. Observed versus predicted yields when mid-season trained models are applied to late-season Sentinel-2 inputs: (a) LiteTemporalConv, (b) MS-ConvBiGRU-Attn.

multi-date information more effectively before sequence modeling. This design contributes to more stable training behavior and improved generalization, reducing overfitting relative to conventional deep learning approaches in the small-sample setting. These results indicate that, under data-limited conditions, carefully designed lightweight temporal models can effectively learn yield-relevant patterns while mitigating overfitting risks when labeled field data is limited. Scatter plots showed good calibration across the yield range with no systematic under- or over-estimation. Additionally, an ablation experiment using only mid-season data demonstrated that both deep learning architectures retained moderate predictive power under reduced temporal input. These results indicate that lightweight temporal CNNs can extract yield-relevant seasonal patterns efficiently, and that a hybrid design combining multi-scale convolution, bidirectional recurrence, and attention is competitive under data scarcity and has the potential to benefit further from larger training datasets.

Recent deep learning studies on non-potato crops provide useful context for our results. For winter wheat, [89] reported a CNN–GRU model with $R^2 = 0.64$ and RMSE of 462.56 kg/ ha−1 using MODIS-derived LAI/FPAR, while a CNN–LSTM trained on multiyear data in [90] achieved $R^2 = 0.77$ and 721 kg h/a RMSE. For soybean, a hybrid CNN–DNN by [91] trained on 25,345 samples outperformed classical machine learning, reaching an $R^2 = 0.87$ RMSE of 0.266. Although direct numerical comparisons are limited by differences in crop, units, sensor resolution, and spatial scale. Within this landscape, our results indicate that a compact, phenology-aware design can yield competitive performance even when restricted to a single season and a limited number of labeled fields.

For potato yield estimation specifically, our findings align with previous satellite-based studies using statistical and machine learning approaches. Earlier work often relied on individual vegetation indices or limited feature sets, reporting R^2 values between 0.58 and 0.82 [92] and ~ 0.81 with rRMSE values around 10% using NDVI alone [37], [38] used vegetation indices extracted from TERRA satellite data, reporting R^2 values of 0.72, 0.80, and 0.84 for LAI, FPAR, and NDVI, respectively [93] estimated potato yield in a laboratory setting using temperature and insolation through linear regression, achieving an R^2 of 0.93 [27] evaluated Sentinel-2–based potato yield prediction across linear and ML algorithms, reporting a Quantile Lasso model with $R^2 = 0.88$ but rRMSE = 11.67%, and an SVM with R^2 =0.93 yet rRMSE = 11.7%. In another study [20], Random Forest using S2 bands + PPI performed better for potatoes than S2 bands + NDVI, achieving $R^2 = 0.77$ and RMSE = 15.42% versus $R^2 = 0.66$ and RMSE = 16.88%, respectively. Typically, prediction errors around 10–15% RMSE are common in crop yield prediction literature [94,95]. In this study, the proposed deep learning models achieved comparatively low relative errors 6 −7%, with LiteTemporalConv having rRMSE of approximately 6.36% and MS-ConvBiGRU-Attn achieving an rRMSE of about 7.03%. Among the baseline methods, XGB performed best, with an rRMSE of approximately 9.8%. These results indicate that the deep learning architecture offers a reduction in relative error compared with the evaluated baselines. The results suggest that, under data-limited conditions, compact and phenology-aware deep models can provide stable and interpretable yield estimates, while offering a flexible foundation for future extensions as additional seasons, regions, or auxiliary data become available.

6. Limitations and future work

While the proposed approach shows potential, several limitations should be acknowledged. The models were developed using data from a single season, which limits the details of multi-year variability driven by changing weather and management conditions. Moreover, the study relied primarily on satellite-derived spectral information; additional variables such as weather or soil data were not included, as commonly available climate products have coarse spatial resolution and showed little variation across the relatively small study area. Future research could benefit from multi-season datasets, higher-resolution environmental information, and broader geographic coverage to further examine robustness.

7. Conclusion

This study investigated the potential of Sentinel-2 satellite time series for predicting single-season, field-scale potato yields under limited data availability. We evaluated a range of models, including our proposed deep learning framework and machine learning baseline models, using combined Sentinel-2 vegetation indices (VIs) and raw spectral bands (SBs) as inputs. Our two encoders, LiteTemporalConv and MS-ConvBiGRU-Attn, explicitly designed for small data and integrated feature selection, consistently outperformed machine-learning baselines. LiteTemporalConv delivered the best overall cross-fold accuracy ($R^2 = 0.84$; RMSE = 3.18 t ha^{-1}; rRMSE = 6.36%), while MS-ConvBiGRU-Attn achieved competitive performance ($R^2 = 0.82$; RMSE = 3.53 t ha^{-1}; rRMSE = 7.03%). An Independent validation tests further suggests that the proposed deep learning models generalize to unseen data, exhibiting stable performance under limited data conditions. Among baselines, XGB was the strongest cross-fold ($R^2 = 0.79$; rRMSE = 9.8%), suggesting that tree-based methods remain effective benchmarks, while sequence-aware deep learning offers additional information about canopy reflectance dynamics. A key driver of these gains was our Grouped TreeSHAP Stability Selection (GTSS), which produced a compact phenology-informed feature set, improving the balance between model complexity and available training data. The use of a Temporal Channel Projection and carefully regularized encoder designs lightweight temporal convolutions in LiteTemporalConv and multi-scale recurrent–attention mechanisms in MS-ConvBiGRU-Attn helped stabilize learning and support well-calibrated predictions across the yield range. Together, these design choices provide complementary modeling pathways: a streamlined, deployment-oriented solution for data-scarce scenarios and a more expressive architecture suited to future extensions with larger, multi-season datasets. While the primary analysis focused on full-season yield prediction, supplementary experiments using mid-season inputs further suggested that both deep learning encoders retain predictive stability when evaluated on late-season data. Overall, integrating Sentinel-2 time series, GTSS-selected features, and purpose-built deep sequence models enables accurate, timely yield estimation to support agronomic planning and operational decision-making.

Declaration of generative AI in scientific writing

The authors used ChatGPT for grammar correction purposes and to improve the English form. After using such tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

Ethics | submission declaration

With reference to the manuscript “Field-Scale Potato Yield Prediction from Sentinel-2 Time Series Using Lightweight Deep Learning Models”, the autors “Rahat Tufail, Patrizia Tassinari and Daniele Torreggiani declare that:

- the work has not been published previously.
- the article is not under consideration for publication elsewhere.
- the article's publication is approved by all authors and tacitly or explicitly by the responsible authorities where the work was carried out.
- if accepted, the article will not be published elsewhere in the same form, in English or in any other language, including electronically, without the written consent of the copyright-holder.

CRediT authorship contribution statement

Rahat Tufail: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Patrizia Tassinari:** Writing – review & editing, Validation, Supervision, Resources, Project administration, Methodology, Funding acquisition. **Daniele Torreggiani:** Writing – review & editing, Validation, Supervision, Resources, Project administration, Methodology, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Daniele Torreggiani reports financial support was provided by European Union. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

This study was carried out within the Agritech National Research Center and received funding from the European Union Next-GenerationEU (PIANO NAZIONALE DI RIPRESA E RESILIENZA (PNRR) – MISSIONE 4 COMPONENTE 2, INVESTIMENTO 1.4 – D.D. 1032 17/06/2022, CN00000022). This manuscript reflects only the authors' views and opinions; neither the European Union nor the European Commission can be considered responsible for them.

The authors would like to thank Pizzoli S.p.A. (Budrio, Italy), for providing the data about yield harvested in the fields used in the study.

Data availability

Data will be made available on request.

References

[1] United Nations, Food production must double by 2050 to meet demand from world's growing population, innovative strategies needed to combat hunger, in: Experts Tell Second Committee | Meetings Coverage and Press Releases, 2025. URL, https://press.un.org/en/2009/gaef3242.doc.htm. accessed 5.13.25.

[2] R. Nishimoto, Global trends in the crop protection industry, J. Pestic. Sci. 44 (3) (2019) 141–147, https://doi.org/10.1584/JPESTICS.D19-101.

[3] United Nations, World Population Projected to Reach 9.8 Billion in 2050, and 11.2 Billion in 2100 | United Nations, United Nations, 2025. https://www.un.org/en/desa/world-population-projected-reach-98-billion-2050-and-112-billion-2100. accessed 5.06.25.

[4] P.M. Vitousek, H.A. Mooney, J. Lubchenco, J.M. Melillo, Human domination of Earth's ecosystems, Science (1979) 277 (5325) (1997) 494–499, https://doi.org/10.1126/SCIENCE.277.5325.494/ASSET/865F55CF-AB01-4FAE-8408-9C87CFEDC964/ASSETS/GRAPHIC/SE3075509004.JPEG.

[5] H.C.J. Godfray, J.R. Beddington, I.R. Crute, L. Haddad, D. Lawrence, J.F. Muir, J. Pretty, S. Robinson, S.M. Thomas, C. Toulmin, Food security: the challenge of feeding 9 billion people, Science (1979) 327 (5967) (2010) 812–818, https://doi.org/10.1126/SCIENCE.1185383/ASSET/A8EAC75B-BEB7-481A-A81A-C845AFA2F0EE/ASSETS/GRAPHIC/327_812_F3.JPEG.

[6] F. Rijsberman, the, D.M.T. background paper to, & 2001, undefined, Balancing Water Uses: Water for Food and Water for Nature, Researchgate.Net, 2001. https://www.researchgate.net/profile/Frank-Rijsberman2/publication/303524689_Balancing_Water_Uses_Water_for_Food_and_Water_for_Nature/links/5746e09308ae298602fbce47/Balancing-Water-Uses-Water-for-Food-and-Water-for-Nature.pdf.

[7] T. Perdana, K. Kusnandar, H.H. Perdana, F.R. Hermiatin, Circular supply chain governance for sustainable fresh agricultural products: minimizing food loss and utilizing agricultural waste, Sustain. Prod. Consum. 41 (2023) 391–403, https://doi.org/10.1016/J.SPC.2023.09.001.

[8] J. Celis, X. Xiao, P. Wagle, P.R. Adler, P. White, A review of yield forecasting techniques and their impact on sustainable agriculture, Green. Indust. Netw. Stud. 13 (2024) 139–168, https://doi.org/10.1007/978-3-031-63793-3_8/FIGURES/3.

[9] B.S. Blackmore, P.N. Wheeler, J. Morris, R.M. Morris, R.J.A. Jones, The role of precision farming in sustainable agriculture: a European perspective. Site-Specific Management for Agricultural Systems, American Society of Agronomy, Crop Science Society of America, Soil Science Society of America, Madison, WI, USA, 1995, pp. 777–793.

[10] R.P. Sishodia, R.L. Ray, S.K. Singh, Applications of remote sensing in precision agriculture: a review, Rem. Sens. (Basel) 12 (19) (2020) 3136, https://doi.org/10.3390/RS12193136, 2020, Vol. 12, Page 3136.

[11] U. Surendran, K.Ch.V. Nagakumar, M.P Samuel, Remote sensing in precision agriculture, Digit. Agricult. 201–223 (2024), https://doi.org/10.1007/978-3-031-43548-5_7/FIGURES/2.

[12] A. Chlingaryan, S. Sukkarieh, B. Whelan, Machine learning approaches for crop yield prediction and nitrogen status estimation in precision agriculture: a review, Comput. Electron. Agric. 151 (2018) 61–69, https://doi.org/10.1016/J.COMPAG.2018.05.012.

[13] G. Perich, M.O. Turkoglu, L.V. Graf, J.D. Wegner, H. Aasen, A. Walter, F. Liebisch, Pixel-based yield mapping and prediction from Sentinel-2 using spectral indices and neural networks, Field. Crops Res. 292 (2023) 108824, https://doi.org/10.1016/J.FCR.2023.108824.

[14] M. Weiss, F. Jacob, G. Duveiller, Remote sensing for agricultural applications: a meta-review, Rem. Sens. Environ. 236 (2020) 111402, https://doi.org/10.1016/J.RSE.2019.111402.

[15] S. Dejoux, M. Ludot, F. Berger, P. Casadebaig, Construire Des Stratégies De Production Adaptées Aux Débouchés à L'échelle Du Bassin De Collecte, 2011. https://www.researchgate.net/profile/Marc-Provot/publication/341773366_Construire_des_strategies_de_production_adaptees_aux_debouches_a_l'echelle_du_bassin_de_collecte/links/61394076349f12090ff1a2fe/Construire-des-strategies-de-production-adaptees-aux-debouches-a-lechelle-du-bassin-de-collecte.pdf.

[16] B. Schauberger, J. Jägermeyr, C. Gornott, A systematic review of local to regional yield forecasting approaches and frequently used data resources, Eur. J. Agron. 120 (2020) 126153, https://doi.org/10.1016/J.EJA.2020.126153.

[17] A. Tewes, H. Hoffmann, G. Krauss, F. Schäfer, C. Kerkhoff, T. Gaiser, New approaches for the assimilation of LAI measurements into a crop model ensemble to improve wheat biomass estimations, Agronomy 10 (3) (2020) 446, https://doi.org/10.3390/AGRONOMY10030446, 2020, Vol. 10, Page 446.

[18] M.G. Ziliani, M.U. Altaf, B. Aragon, R. Houburg, T.E. Franz, Y. Lu, J. Sheffield, I. Hoteit, M.F. McCabe, Early season prediction of within-field crop yield variability by assimilating CubeSat data into a crop model, Agric. For. Meteorol. 313 (2022) 108736, https://doi.org/10.1016/J.AGRFORMET.2021.108736.

[19] D. Li, Y. Miao, S.K. Gupta, C.J. Rosen, F. Yuan, C. Wang, L. Wang, Y. Huang, Improving potato yield prediction by combining cultivar information and UAV remote sensing data using machine learning, Rem. Sens. (Basel) 13 (16) (2021) 3322, https://doi.org/10.3390/RS13163322, 2021, Vol. 13, Page 3322.

[20] D. Gómez, P. Salvador, J. Sanz, J.L. Casanova, New spectral indicator Potato Productivity Index based on Sentinel-2 data to improve potato yield prediction: a machine learning approach, Int. J. Rem. Sens. 42 (9) (2021) 3430–3448, https://doi.org/10.1080/01431161.2020.1871102.

[21] C.M. Ayyub, M. Wasim Haidar, F. Zulfiqar, Z. Abideen, S.R. Wright, Potato tuber yield and quality in response to different nitrogen fertilizer application rates under two split doses in an irrigated sandy loam soil, J. Plant Nutr. 42 (15) (2019) 1850–1860, https://doi.org/10.1080/01904167.2019.1648669.

[22] Eurostat, The EU Potato Sector - Statistics on Production, Prices and Trade - Statistics Explained - Eurostat, May 7, 2025, https://ec.europa.eu/eurostat/statistics-explained/index.php?title=The_EU_potato_sector_-_statistics_on_production,_prices_and_trade#Volatile_prices.2C_both_on_producer_and_consumer_markets.

[23] G. Timpanaro, F. Branca, M. Cammarata, G. Falcone, A. Scuderi, Life cycle assessment to highlight the environmental burdens of early potato production, Agronomy 11 (5) (2021) 879, https://doi.org/10.3390/AGRONOMY11050879, 2021, Vol. 11, Page 879.

[24] OEC, Potatoes in Italy Trade | The Observatory of Economic Complexity, May, 2025, https://oec.world/en/profile/bilateral-product/potatoes/reporter/ita.

[25] K.A. Al-Gaadi, A.A. Hassaballa, E. Tola, A.G. Kayad, R. Madugundu, B. Alblewi, F. Assiri, Prediction of potato crop yield using precision agriculture techniques, PLoS One 11 (9) (2016) e0162219, https://doi.org/10.1371/JOURNAL.PONE.0162219.

[26] D.H. Fleisher, B. Condori, R. Quiroz, A. Alva, S. Asseng, C. Barreda, M. Bindi, K. J. Boote, R. Ferrise, A.C. Franke, P.M. Govindakrishnan, D. Harahagazwe, G. Hoogenboom, S. Naresh Kumar, P. Merante, C. Nendel, J.E. Olesen, P.S. Parker, D. Raes, P. Woli, A potato model intercomparison across varying climates and productivity levels, Glob. Chang. Biol. 23 (3) (2017) 1258–1281, https://doi.org/10.1111/GCB.13411;JOURNAL:JOURNAL:13652486;WGROUP:STRING:PUBLICATION.

[27] D. Gómez, P. Salvador, J. Sanz, J.L. Casanova, Potato yield prediction using machine learning techniques and Sentinel 2 data, Rem. Sens. (Basel) 11 (15) (2019) 1745, https://doi.org/10.3390/RS11151745, 2019, Vol. 11, Page 1745.

[28] R. Raymundo, S. Asseng, D. Cammarano, R. Quiroz, Potato, sweet potato, and yam models for climate change: a review, Field. Crops Res. 166 (2014) 173–185, https://doi.org/10.1016/J.FCR.2014.06.017.

[29] G. Belanger, J.R. Walsh, J.E. Richards, P.H. Milburn, N. Ziadi, Comparison of three statistical models describing potato yield response to nitrogen fertilizer, Agron. J. 92 (5) (2000) 902–908, https://doi.org/10.2134/AGRONJ2000.925902X.

[30] Kooman, P.L., & Haverkort, A.J. (1995). Modelling development and growth of the potato crop influenced by temperature and daylength: LINTUL-POTATO. 41–59. https://doi.org/10.1007/978-94-011-0051-9_3.

[31] L.A. Manrique, J.R. Kinry, T. Hodges, D.S. Axness, Dry matter production and radiation interception of potato, Crop. Sci. 31 (4) (1991) 1044–1049, https://doi.org/10.2135/CROPSCI1991.0011183X003100040040X.

[32] Search.Informit.Org D. Borus, D. Parsons, M. Boersma, H. Brown, C. Mohammed, Improving the prediction of potato productivity: aPSIM-potato model parameterization and evaluation in Tasmania, Australia, Australian, in: D. Borus, D. Parsons, M. Boersma, H. Brown, C. Mohammed (Eds.), Improving the prediction

of potato productivity: aPSIM-potato model parameterization and evaluation in Tasmania, Australia, J. Crop Sci. 12 (01) (2018) 1835–2707, https://doi.org/10.3316/INFORMIT.459351398463867, 2018•search.Informit.Org.
[33] D.A. Kasampalis, T.K. Alexandridis, C. Deva, A. Challinor, D. Moshou, G. Zalidis, Contribution of remote sensing on crop models: a review, J. Imag. 4 (4) (2018) 52, https://doi.org/10.3390/JIMAGING4040052, 2018, Vol. 4, Page 52.
[34] S. Fritz, L. See, J.C.L. Bayas, F. Waldner, D. Jacques, I. Becker-Reshef, A. Whitcraft, B. Baruth, R. Bonifacio, J. Crutchfield, F. Rembold, O. Rojas, A. Schucknecht, M. Van der Velde, J. Verdin, B. Wu, N. Yan, L. You, S. Gilliams, I. McCallum, A comparison of global agricultural monitoring systems and current gaps, Agric. Syst. 168 (2019) 258–272, https://doi.org/10.1016/J.AGSY.2018.05.010.
[35] D.J. Mulla, Twenty five years of remote sensing in precision agriculture: key advances and remaining knowledge gaps, Biosyst. Eng. 114 (4) (2013) 358–371, https://doi.org/10.1016/J.BIOSYSTEMSENG.2012.08.009.
[36] A. Raza, Y. Miao, Y. Huang, K. Stueve, J. Lu, Z. Yang, R. Bindlish, Optimizing on-farm corn yield prediction by a multi-source data fusion approach using remote sensing and machine learning, Smart Agricult. Tech. (2025) 101630.
[37] I.H. Newton, A.F.M. Tariqul Islam, A.K.M. Saiful Islam, G.M. Tarekul Islam, A. Tahsin, S. Razzaque, Yield prediction model for potato using landsat time series images driven vegetation indices, Rem. Sens. Earth. Syst. Sci. 1 (1–2) (2018) 29–38, https://doi.org/10.1007/S41976-018-0006-0/FIGURES/8.
[38] S.K. Bala, A.S. Islam, Correlation between potato yield and MODIS-derived vegetation indices, Int. J. Rem. Sens. 30 (10) (2009) 2491–2507, https://doi.org/10.1080/01431160802552744.
[39] R.E.E. Jongschaap, R. Booij, Spectral measurements at different spatial scales in potato: relating leaf, plant and canopy nitrogen status, Int. J. Appl. Earth Observ. Geoinform. 5 (3) (2004) 205–218, https://doi.org/10.1016/J.JAG.2004.03.002.
[40] T. Zheng, N. Liu, L. Wu, M. Li, H. Sun, Q. Zhang, J. Wu, Estimation of chlorophyll content in potato leaves based on spectral red edge position, IFAC-PapersOnLine 51 (17) (2018) 602–606, https://doi.org/10.1016/J.IFACOL.2018.08.131.
[41] A.N. Kravchenko, D.G. Bullock, Correlation of corn and soybean grain yield with topography and soil properties, Agron. J. 92 (1) (2000) 75–83, https://doi.org/10.2134/AGRONJ2000.92175X.
[42] Y. Liu, G.B.M. Heuvelink, Z. Bai, P. He, X. Xu, W. Ding, S. Huang, Analysis of spatio-temporal variation of crop yield in China using stepwise multiple linear regression, Field. Crops Res. 264 (2021) 108098, https://doi.org/10.1016/J.FCR.2021.108098.
[43] J. Ansarifar, L. Wang, S.V. Archontoulis, An interaction regression model for crop yield prediction, Sci. Rep. 11 (1) (2021) 1–14, https://doi.org/10.1038/S41598-021-97221-7;SUBJMETA=114,1305,449,631;KWRD=MACHINE+LEARNING,PLANT+SCIENCES.
[44] S.T. Drummond, K.A. Sudduth, A. Joshi, S.J. Birrell, N.R. Kitchen, Statistical and neural methods for site–SPECIFIC yield prediction, Trans. ASAE 46 (1) (2003) 5, https://doi.org/10.13031/2013.12541. -.
[45] A. Chlingaryan, S. Sukkarieh, B. Whelan, Machine learning approaches for crop yield prediction and nitrogen status estimation in precision agriculture: a review, Comput. Electron. Agric. 151 (2018) 61–69, https://doi.org/10.1016/J.COMPAG.2018.05.012.
[46] H. Burdett, C. Wellen, Statistical and machine learning methods for crop yield prediction in the context of precision agriculture, Precis. Agric. 23 (5) (2022) 1553–1574, https://doi.org/10.1007/S11119-022-09897-0/FIGURES/6.
[47] S. Zhou, L. Xu, N. Chen, Rice yield prediction in Hubei Province based on deep learning and the effect of spatial heterogeneity, Rem. Sens. (Basel) 15 (5) (2023) 1361, https://doi.org/10.3390/RS15051361, 2023, Vol. 15, Page 1361.
[48] C. Trentin, Y. Ampatzidis, C. Lacerda, L. Shiratsuchi, Tree crop yield estimation and prediction using remote sensing and machine learning: a systematic review, Smart Agricult. Tech. 9 (2024) 100556.
[49] H. Tian, P. Wang, K. Tansey, J. Zhang, S. Zhang, H. Li, An LSTM neural network for improving wheat yield estimates by integrating remote sensing data and meteorological data in the Guanzhong Plain, PR China, Agricult. For. Meteorol. 310 (2021) 108629, https://doi.org/10.1016/J.AGRFORMET.2021.108629.
[50] J. Sun, L. Di, Z. Sun, Y. Shen, Z. Lai, County-level soybean yield prediction using deep CNN-LSTM model, Sensors 19 (20) (2019) 4363, https://doi.org/10.3390/S19204363, 2019, Vol. 19, Page 4363.
[51] J. Cao, Z. Zhang, Y. Luo, L. Zhang, J. Zhang, Z. Li, F. Tao, Wheat yield predictions at a county and field scale with deep learning, machine learning, and google earth engine, Eur. J. Agron. 123 (2021) 126204, https://doi.org/10.1016/J.EJA.2020.126204.
[52] L. Brigato, L. Iocchi, A close look at deep learning with small data, in: 2020 25th international conference on pattern recognition (ICPR), IEEE, 2021, pp. 2490–2497.
[53] M. Vizzari, PlanetScope, Sentinel-2, and Sentinel-1 data integration for object-based land cover classification in Google Earth Engine, Rem. Sens. (Basel) 14 (11) (2022) 2628, https://doi.org/10.3390/RS14112628/S1.
[54] J.W. Rouse Jr., R.H. Haas, D.W. Deering, J.A. Schell, J.C Harlan, Monitoring the Vernal Advancement and Retrogradation (Green Wave Effect) of Natural Vegetation, 1974.
[55] Weiss Marie, Baret Fred, Sentinel2 ToolBox Level2 Products S2ToolBox Level 2 products: LAI, FAPAR, FCOVER Version 1.1, 2016. https://step.esa.int/docs/extra/ATBD_S2ToolBox_V2.1.pdf.
[56] J. Dash, P.J. Curran, The MERIS terrestrial chlorophyll index, Int. J. Rem. Sens. 25 (23) (2004) 5403–5413, https://doi.org/10.1080/0143116042000274015.
[57] W.J. Frampton, J. Dash, G. Watmough, E.J. Milton, Evaluating the capabilities of Sentinel-2 for quantitative estimation of biophysical variables in vegetation, ISPRS J. Photogram. Rem. Sens. 82 (2013) 83–92, https://doi.org/10.1016/J.ISPRSJPRS.2013.04.007.
[58] C.S.T. Daughtry, C.L. Walthall, M.S. Kim, E.B. De Colstoun, J.E. McMurtrey, Estimating corn leaf chlorophyll concentration from leaf and canopy reflectance, Rem. Sens. Environ. 74 (2) (2000) 229–239, https://doi.org/10.1016/S0034-4257(00)00113-9.
[59] A.A. Gitelson, Y.J. Kaufman, M.N. Merzlyak, Use of a green channel in remote sensing of global vegetation from EOS-MODIS, Rem. Sens. Environ. 58 (3) (1996) 289–298, https://doi.org/10.1016/S0034-4257(96)00072-7.
[60] A.R. Huete, A soil-adjusted vegetation index (SAVI), Rem. Sens. Environ. 25 (3) (1988) 295–309, https://doi.org/10.1016/0034-4257(88)90106-X.
[61] D.S. Kimes, B.L. Markham, C.J. Tucker, J.E. McMurtrey, Temporal relationships between spectral response and agronomic variables of a corn canopy, Rem. Sens. Environ. 11 (C) (1981) 401–411, https://doi.org/10.1016/0034-4257(81)90037-7.
[62] G. Misra, F. Cawkwell, A. Wingler, Status of phenological research using sentinel-2 data: a review, Rem. Sens. (Basel) 12 (17) (2020) 2760, https://doi.org/10.3390/RS12172760, 2020, Vol. 12, Page 2760.
[63] E. Barnes, T. Clarke, S.R.P. of the, & 2000, undefined, Coincident Detection of Crop Water Stress, Nitrogen Status and Canopy Density using Ground Based Multispectral Data, Researchgate.Net, 2000. https://www.researchgate.net/profile/Peter-Waller/publication/43256762_Coincident_detection_of_crop_water_stress_nitrogen_status_and_canopy_density_using_ground_based_multispectral_data/links/55ac358c08ae481aa7ff4da7/Coincident-detection-of-crop-water-stress-nitrogen-status-and-canopy-density-using-ground-based-multispectral-data.pdf.
[64] E.V. Gureeva, O.V. Levakova, Remote monitoring of chlorophyll content in soybean crops in the conditions of the Ryazan region, BIO Web Conf. 71 (2023) 01090, https://doi.org/10.1051/BIOCONF/20237101090.
[65] T.B. Raper, J.J. Varco, Canopy-scale wavelength and vegetative index sensitivities to cotton growth parameters and nitrogen status, Precis. Agric. 16 (1) (2015) 62–76, https://doi.org/10.1007/S11119-014-9383-4/TABLES/4.
[66] S. Liang, Y. Zhang, K. Zheng, Y. Bai, FeatureX: an explainable feature selection for deep learning, Expert. Syst. Appl. 282 (2025) 127675, https://doi.org/10.1016/J.ESWA.2025.127675.
[67] Y. Liu, W. Qin, Q. Zheng, G. Li, M. Li, An interpretable feature selection based on particle swarm optimization, IEICe Trans. Inf. Syst. E105.D (8) (2022) 1495–1500, https://doi.org/10.1587/TRANSINF.2021EDL8095.
[68] M.S.K. Inan, I. Rahman, Explainable AI integrated feature selection for landslide susceptibility mapping using TreeSHAP, S.N. Comput. Sci. 4 (5) (2023) 1–18, https://doi.org/10.1007/S42979-023-01960-5/FIGURES/24.
[69] C.S. Wickramasinghe, D.L. Marino, M. Manic, ResNet autoencoders for unsupervised feature learning from high-dimensional data: deep models resistant to performance degradation, IEEE Access 9 (2021) 40511–40520, https://doi.org/10.1109/ACCESS.2021.3064819.
[70] A.K. Srivastava, N. Safaei, S. Khaki, G. Lopez, W. Zeng, F. Ewert, T. Gaiser, J. Rahimi, Winter wheat yield prediction using convolutional neural networks from environmental and phenological data, Sci. Rep. 12 (1) (2022), https://doi.org/10.1038/S41598-022-06249-W.
[71] R. Tufail, P. Tassinari, D. Torreggiani, Deep learning applications for crop mapping using multi-temporal Sentinel-2 data and red-edge vegetation indices: integrating convolutional and recurrent neural networks, Rem. Sens. (Basel) 17 (18) (2025) 3207, https://doi.org/10.3390/RS17183207, 2025, Vol. 17, Page 3207.
[72] A. Joshi, B. Pradhan, S. Chakraborty, R. Varatharajoo, A. Alamri, S. Gite, C.W. Lee, An explainable Bi-LSTM model for winter wheat yield prediction, Front. Plant Sci. 15 (2024) 1491493, https://doi.org/10.3389/FPLS.2024.1491493/BIBTEX.
[73] Y. Zhu, W. Chen, S. Yan, J. Zhang, C. Zhu, F. Wang, Q. Chen, Multi-scale residual convolutional neural network with hybrid attention for bearing fault detection, Machines 13 (5) (2025), https://doi.org/10.3390/MACHINES13050413.
[74] L. Breiman, Random forests, Mach. Learn. 45 (1) (2001) 5–32, https://doi.org/10.1023/A:1010933404324/METRICS.
[75] K. Luo, L. Lu, Y. Xie, F. Chen, F. Yin, Q. Li, Crop type mapping in the central part of the North China Plain using Sentinel-2 time series and machine learning, Comput. Electron. Agric. 205 (2023) 107577, https://doi.org/10.1016/J.COMPAG.2022.107577.
[76] H. Abdi, Partial least squares regression and projection on latent structure regression (PLS Regression), Wiley Interdisciplin. Rev. Comput. Stat. 2 (1) (2010) 97–106, https://doi.org/10.1002/WICS.51;REQUESTEDJOURNAL:JOURNAL:19390068;CTYPE:STRING:JOURNAL.
[77] H. Abdi, Partial least square regression PLS-regression, Encyclop. Res. Method Soc. Sci. 6 (4) (2003) 792–795. http://www.utd.edu/.
[78] V.N. Vapnik, The nature of statistical learning theory, Nat. Stat. Learn. Theo. (2000), https://doi.org/10.1007/978-1-4757-3264-1.
[79] T. Chen, C. Guestrin, XGBoost: a scalable tree boosting system, in: Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2016, pp. 785–794, 13-17-August-2016.
[80] H.A. Martens, P. Dardenne, Validation and verification of regression in small data sets, Chemomet. Intel. Lab. Syst. 44 (1–2) (1998) 99–121.
[81] W. Xu, P. Chen, Y. Zhan, S. Chen, L. Zhang, Y. Lan, Cotton yield estimation model based on machine learning using time series UAV remote sensing data, Int. J. Appl. Earth Observ. Geoinform. 104 (2021) 102511.
[82] O. Kryvoshein, O. Kryvobok, D. Zhylchenko, Yield prediction at field level, Ukrain. J. Rem. Sens. 11 (4) (2024) 26–30, https://doi.org/10.36023/UJRS.2024.11.4.275.
[83] M. Marszalek, M. Körner, U. Schmidhalter, Prediction of multi-year winter wheat yields at the field level with satellite and climatological data, Comput. Electron. Agric. 194 (2022), https://doi.org/10.1016/J.COMPAG.2022.106777.

[84] F. Sabo, M. Meroni, F. Waldner, F. Rembold, Is deeper always better? Evaluating deep learning models for yield forecasting with small data, Environ. Monit. Assess. 195 (10) (2023) 1–11, https://doi.org/10.1007/S10661-023-11609-8/FIGURES/6.
[85] A.X. Wang, C. Tran, N. Desai, D. Lobell, S. Ermon, Deep transfer learning for crop yield prediction with remote sensing data, in: Proceedings of the 1st ACM SIGCAS Conference on Computing and Sustainable Societies, COMPASS 2018 18, 2018, https://doi.org/10.1145/3209811.3212707.
[86] K. Meghraoui, I. Sebari, J. Pilz, K. Ait El Kadi, S. Bensiali, Applied deep learning-based crop yield prediction: a systematic analysis of current developments and potential challenges, Technologies (Basel) 12 (4) (2024), https://doi.org/10.3390/TECHNOLOGIES12040043.
[87] P. Muruganantham, S. Wibowo, S. Grandhi, N.H. Samrat, N. Islam, A systematic literature review on crop yield prediction with deep learning and remote sensing, Rem. Sens. 14 (9) (2022), https://doi.org/10.3390/RS14091990.
[88] R. Tufail, A. Ahmad, M.A. Javed, S.R. Ahmad, A machine learning approach for accurate crop type mapping using combined SAR and optical time series data, Adv. Space Res. 69 (1) (2022) 331–346, https://doi.org/10.1016/J.ASR.2021.09.019.
[89] J. Wang, P. Wang, H. Tian, K. Tansey, J. Liu, W. Quan, A deep learning framework combining CNN and GRU for improving wheat yield estimates using time series remotely sensed multi-variables, Comput. Electron. Agric. 206 (2023) 107705, https://doi.org/10.1016/J.COMPAG.2023.107705.
[90] X. Wang, J. Huang, Q. Feng, D. Yin, Winter wheat yield prediction at county level and uncertainty analysis in main wheat-producing regions of China with deep learning approaches, Rem. Sens. (Basel) 12 (11) (2020) 1744, https://doi.org/10.3390/RS12111744, 2020, Vol. 12, Page 1744.
[91] A. Oikonomidis, C. Catal, A. Kassahun, Hybrid deep learning-based models for crop yield prediction, Appl. Artif. Intell. 36 (1) (2022).
[92] Singha, C., Swain, K.C., Singha, C., & Swain, Á.K.C. (2023). Vegetation indices-based rice and potato yield estimation through Sentinel 2B satellite imagery. 113–134. https://doi.org/10.1007/978-3-031-36825-7_8.
[93] T.K. Hartz, F.D. Moore, Prediction of potato yield using temperature and insolation data, Am. Potato J. 55 (8) (1978) 431–436, https://doi.org/10.1007/BF02852146/METRICS.
[94] M. Launay, M. Guerif, Assimilating remote sensing data into a crop model to improve predictive performance for spatial applications, Agric. Ecosyst. Environ. 111 (1–4) (2005) 321–339, https://doi.org/10.1016/J.AGEE.2005.06.005.
[95] M.S. Mkhabela, P. Bullock, S. Raj, S. Wang, Y. Yang, Crop yield forecasting on the Canadian Prairies using MODIS NDVI data, Agric. For. Meteorol. 151 (3) (2011) 385–393, https://doi.org/10.1016/J.AGRFORMET.2010.11.012.