Vinicius Silva[1], Ewerton Silva[1], Diogo Nascimento[1], Renzo Negrini[2], Letícia Mendes[1], Arthur Tavares[1], Mailson Oliveira[3]

[1] School of Engineering and Agricultural Sciences, Federal University of Alagoas, Brazil

[2] Precision Agriculture Center, University of Minnesota, Saint Paul, MN, USA

[3] Institute of Agriculture and Natural Resources, University of Nebraska-Lincoln, Lincoln, Nebraska, USA

A paper from the Proceedings of the
17th International Conference on Precision Agriculture and 11th Brazilian Congress on Precision and Digital Agriculture
13-15 July 2026
Porto Alegre, Brazil

Abstract.

Accurate, spatially explicit yield mapping underpins many precision agriculture decisions (e.g., variable-rate inputs and zone management), yet reliable yield-monitor data are not always available and can be difficult to standardize across operations. Satellite-based yield models are often built from hand-crafted vegetation indices or phenology metrics, which may limit their transferability across fields and seasons. Here we evaluate a pixel-level corn yield prediction workflow that uses Satellite Embedding V1 — 64-dimensional deep-learning-derived vector features computed from satellite imagery — to represent full-season crop conditions with minimal manual feature engineering. The study used 2023 and 2024 irrigated corn data from six commercial fields in Nebraska, USA, spanning multiple hybrids. Multi-source spatial layers were filtered and interpolated to a common pixel resolution to align predictors with observed grain yield. The H2O AutoML framework was applied to compare several regression algorithms (GBM, Random Forest, XGBoost, Deep Learning, GLM, and stacked ensembles) using 5-fold cross-validation. The best-performing model was a Gradient Boosting Machine (GBM), achieving cross-validated $R^2 = 0.95$, RMSE = 0.25 t ha^{-1}, MAE = 0.19 t ha^{-1}, and MSE = 0.06 (t ha^{-1})^2. Predicted yields closely tracked observed yields along the 1:1 line across the within-field yield range (~10–17 t ha^{-1}), indicating low bias and strong pixel-scale calibration. Because embedding features are high-dimensional and not directly interpretable, explainable AI methods were integrated to support agronomic inference. Global permutation importance identified a small subset of embedding dimensions as dominant model drivers (notably 23-24_utm, 23-24_ut35, and 23-24_ut54). SHAP (SHapley Additive exPlanations) values, computed via H2O's native implementation, provided a theoretically grounded decomposition based on cooperative game theory, satisfying local accuracy, missingness, and consistency. Beyond accurate prediction, the embedding representation offers a scalable pathway to delineate management zones by leveraging similarity in the embedding vector space, potentially enabling zone transfer to fields without yield-monitor data. This work demonstrates that Satellite Embedding V1

features coupled with AutoML-selected GBM models can generate accurate, interpretable, pixel-level yield predictions in irrigated corn and provides a foundation for scalable decision support where yield observations are incomplete or unavailable.

Keywords.
Corn yield prediction, satellite embeddings, AutoML, explainable AI.

Introduction

Length Yield formation arises from complex interactions among genotype, environment, and management that unfold continuously throughout the crop growing season (Lobell; Burke, 2010). These interactions generate high spatial and temporal variability within and across fields, making it difficult to produce reliable, spatially explicit yield estimates from conventional approaches alone (Basso et al., 2019). Despite decades of progress in crop modeling and remote sensing, the precise quantification of within-field yield heterogeneity remains limited, particularly in commercial production environments where ground-truth data collection is resource-intensive and often incomplete (Maestrini; Basso, 2018). Overcoming these constraints is essential for enabling site-specific crop management strategies, including variable-rate fertilization, irrigation management, and management zone delineation, all of which depend on spatially resolved yield information (Mulla, 2013).

Vegetation indices derived from multispectral imagery, such as the Normalized Difference Vegetation Index (NDVI), the Enhanced Vegetation Index (EVI), and the Leaf Area Index (LAI), have been widely used as proxies for crop biomass accumulation and photosynthetically active radiation interception, both closely related to final grain yield (Sakamoto et al., 2014). Phenology-based metrics extracted from vegetation index time series have also been shown to capture critical developmental stages — green-up, peak biomass, and senescence — that are strongly correlated with final productivity (Zheng et al., 2020). However, traditional vegetation index approaches suffer from sensitivity to atmospheric interference, soil background effects, and saturation under high biomass, as well as poor transferability across environments, years, and varieties (Xin et al., 2013). These limitations motivate the search for more robust, generalizable feature representations from satellite data.

Rather than relying on hand-crafted spectral indices, foundation models for Earth observation learn compact, task-agnostic representations by training on large volumes of multi-source, multi-temporal satellite imagery without domain-specific labeling (Rolf et al., 2021; Jakubik et al., 2023). Among the most recently developed approaches, AlphaEarth Foundations (AEF), made available as Satellite Embedding V1 in Google Earth Engine, produces 64-dimensional embedding vectors at 10-m spatial resolution for every land pixel annually from 2017 to 2024 (Brown et al., 2025). These embeddings are generated by a spatiotemporal architecture trained on over three billion image frames from Sentinel-1, Sentinel-2, Landsat 8/9, PALSAR-2, ERA5-Land climate reanalysis, GEDI LiDAR, and GRACE gravimetry. They consistently outperform both designed and learned featurization across mapping evaluations covering land use and land cover classification, crop type mapping, biophysical variable estimation, and change detection (Brown et al., 2025).

Early studies showed that ensemble methods such as Random Forests and Gradient Boosting Machines can effectively model nonlinear relationships between vegetation indices and crop yield when trained on sufficiently large datasets (Nevavuori et al., 2019). More recent work has leveraged deep learning architectures, including CNNs and RNNs, to extract complex temporal patterns from satellite image time series for yield forecasting at county, regional, and national scales (Khaki; Wang, 2019). At the field and sub-field scales, several studies have explored high-resolution satellite imagery combined with machine learning to delineate management zones and predict within-field yield variability (Shahhosseini et al., 2021). Automated machine learning (AutoML) frameworks such as H2O AutoML further facilitate model selection and hyperparameter tuning by systematically comparing multiple algorithms and building stacked ensemble models without extensive manual effort (Ledell; Poirier, 2020).

Existing satellite-based yield prediction models typically require selection and optimization of vegetation indices and phenological metrics for each target region and crop type, limiting transferability to new fields and seasons (Martos et al., 2021). Furthermore, most published approaches focus on regional or national scales and do not provide the sub-field spatial resolution

needed for variable-rate management decisions (Lobell, 2013).

The primary objective of this study was to evaluate a pixel-level corn yield prediction workflow using Satellite Embedding V1 features as input to an H2O AutoML regression model, assess its predictive accuracy through cross-validation, and demonstrate its potential for generating spatially explicit yield maps.

Material and Methods

Study Area and Growing Seasons

The study was conducted in six commercial irrigated corn (Zea mays L.) fields in Nebraska, United States, spanning multiple hybrids. It covered two consecutive growing seasons — 2023 and 2024 — to capture interannual variability in weather conditions, management practices, and field-level yield performance. Planting and harvest followed the regional agronomic calendar, with sowing typically between late April and early June and harvest between September and November.

Study Area and Growing Seasons

The methodological framework integrates three main components: (i) dense spectral embedding data derived from satellite imagery via Google Earth Engine; (ii) georeferenced corn yield observations from yield monitors installed on GNSS/GPS-equipped combines; and (iii) an AutoML pipeline for model selection, hyperparameter optimization, and predictive performance evaluation. Figure 1 presents a schematic overview of the complete workflow.

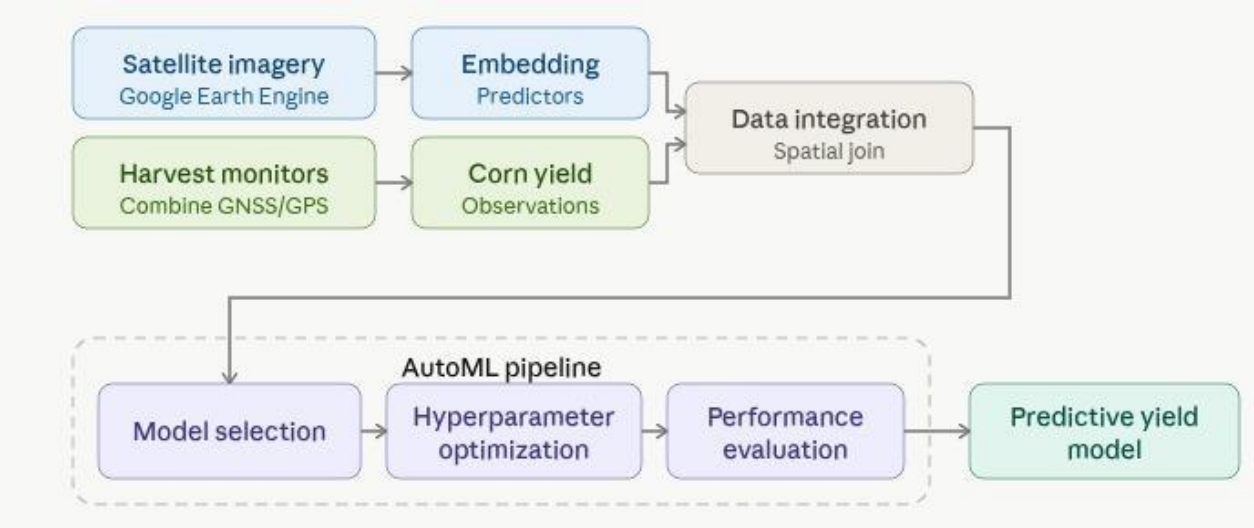

Fig. 1 Workflow diagram

Remote Sensing Data: Google Satellite Embedding V1

Temporal Acquisition Strategy and Quality Control

Imagery was acquired continuously over the complete corn development cycle in both growing seasons, from emergence to physiological maturity.

Dataset Dimensions

After acquisition and quality filtering, the satellite embedding dataset totaled 18,098 georeferenced observations, each represented by a 64-dimensional vector. These observations constituted the candidate pool for spatial matching with field-collected yield reference points.

Yield Reference Data from Yield Monitors

Monitoring Technology and Collection Protocol

A total of 18,098 raw observations of geographic coordinates (latitude and longitude in decimal degrees, WGS84 datum), instantaneous yield estimate, grain moisture content at harvest (%), machine travel speed (*km h*$^{-1}$), and effective header cutting width (*m*) were recorded across all fields harvested during the 2023 and 2024 growing seasons. Raw yield-monitor data are subject to systematic and random errors inherent to the acquisition process, requiring a pre-processing step before any analysis.

Data Quality Control with MapFilter 2.0.1

Raw yield-monitor data were processed with MapFilter 2.0.1, a free software developed specifically to analyze and remove inconsistent records in high-density agricultural datasets, including crop monitoring data, plant sensors, and soil sensors. After applying the global and local spatial filters, 10,011 yield observations were retained — georeferenced records of high spatial reliability expressed in *kg ha*$^{-1}$. These observations constituted the response-variable dataset for model training and evaluation.

Geospatial Integration and Feature Matrix Construction

The cleaned yield points were spatially intersected with the Google Satellite Embedding V1 raster through a point-in-pixel spatial join, assigning each observation the 64-dimensional embedding vector of the 10-m pixel at its geographic coordinate. The operation was performed in QGIS, ensuring consistency between the satellite data coordinate reference system (EPSG:4326 / WGS84) and the GNSS coordinates of the yield points.

The resulting integrated dataset constituted a feature matrix $X \in \mathbb{R}^{nxp}$ and a response vector $y \in \mathbb{R}^{n}$, where *n* is the number of observations and $p = 64$ corresponds to the embedding dimensions. The response variable y represented corn grain yield in *kg ha*$^{-1}$, corrected to 15.5% moisture. Observations with missing values (NA/NaN) — attributable to satellite coverage gaps from persistent clouds or temporal misalignment — were removed by listwise deletion. The final analytical set comprised $n = 10,011$ complete cases and $p = 64$ predictor variables.

Automated Machine Learning with H2O AutoML

Framework and Computational Environment

Model training and selection were conducted with H2O AutoML (H2O.ai Inc., v3.x, 2024), an open-source AutoML framework implemented via the H2O Python API. H2O AutoML automates the construction, evaluation, and ranking of a broad portfolio of supervised algorithms, performing implicit hyperparameter tuning through Cartesian or random grid search and building Stacked Ensemble models from the trained base learners. Processing was carried out on Google Colaboratory with a TPU v5e accelerator (v5e-1, single-chip) and 12 GB RAM, with a local H2O cluster initialized to run AutoML.

AutoML Configuration

The AutoML run was configured with a maximum runtime of 7,200 seconds or at most 100 individual models — whichever was reached first. A fixed random seed (seed = 1) was set to ensure full reproducibility. The task was formulated as regression with a continuous response variable. All 64 embedding features were supplied as predictors without prior selection or dimensionality reduction. Internal model performance was assessed by 5-fold cross-validation, with cross-validated RMSE as the primary selection criterion.

Selected Model

The best model was selected as the final model based on the lowest mean cross-validated RMSE.

GBM is an additive ensemble method that sequentially fits decision trees to the negative gradient of the loss function — Mean Squared Error (MSE) for regression — producing a final predictor that approximates the functional relationship between spectral embeddings and yield as a sum of T weak learners:

$$F(x) = F_0 + \Sigma_{t=1} T \nu \cdot h_t(x) \quad (1)$$

where F_0 is the initial constant prediction, $h_t(x)$ is the t-th regression tree, and v is the learning rate (learn_rate), a shrinkage parameter controlling each tree's individual contribution.

The uniform depth of 13 levels across all 74 trees indicates that AutoML identified a high-representational-capacity configuration as optimal for this dataset, allowing each tree to capture high-order interactions among the embedding dimensions. The mean of *10,408 leaves per tree* reflects the complexity of the partitions learned in the 64-feature space. The number of trees ($ntrees = 74$) was determined by the early-stopping criterion applied to the cross-validation learning curve, halting when no significant RMSE improvement was observed over consecutive rounds. Models were ranked from lowest to highest mean cross-validated RMSE (cv_rmse).

Model Evaluation

Cross-Validation Strategy

Generalization performance was estimated by 5-fold cross-validation ($k = 5$). The analytical set ($n = 10,011$) was randomly partitioned into 5 non-overlapping subsets of approximately equal size. In each iteration, the model was trained on the four remaining folds ($n\_train \approx 8,009$) and evaluated on the held-out fold ($n\_val \approx 2,002$). Metrics were computed for each fold and then averaged, yielding a stable, low-variance estimate of predictive accuracy.

Performance Metrics

The following quantitative metrics were used to characterize model predictive performance:

$$RMSE = \sqrt{\left[\left(\frac{1}{n}\right) \cdot \Sigma_i (y_i - \hat{y}_i)^2\right]} \quad (2)$$

$$MAE = \left(\frac{1}{n}\right) \cdot \Sigma_i \left|y_i - \hat{y}_i\right| \quad (3)$$

$$R^2 = 1 - \left[\frac{\Sigma_i (y_i - \hat{y}_i)^2}{\Sigma_i (y_i - \bar{y})^2}\right] \quad (4)$$

where y_i is the observed value, $\hat{y}_i$ is the predicted value, $\bar{y}$ is the mean of observed values, and n is the number of observations. RMSE and MAE are expressed in *t ha⁻¹* and quantify mean prediction error; RMSE penalizes larger deviations more strongly due to the squared term, whereas MAE is more robust to heteroscedasticity. *R²* is dimensionless and expresses the proportion of yield variance explained by the model. Mean Squared Error:

$$MSE = \left(\frac{1}{n}\right) \Sigma_i (y_i - \hat{y}_i)^2 \quad (5)$$

is expressed in (t ha⁻¹)² and equals *RMSE²*. The mean error:

$$ME = \left(\frac{1}{n}\right) \Sigma_i (\hat{y}_i - y_i) \quad (6)$$

was also computed as a global bias measure; values near zero indicate the absence of systematic

over- or under-prediction. Error metrics produced by H2O in *kg ha⁻¹* were converted to *t ha⁻¹* by dividing linear metrics (RMSE, MAE, ME) by 1,000 and squared metrics (MSE, MRD) by 1,000².

Feature Importance via SHAP Analysis

The relative contribution of each of the 64 embedding dimensions to GBM predictions was assessed using SHAP (SHapley Additive exPlanations) values, computed via the native H2O implementation (Lundberg; Lee, 2017). SHAP values provide a decomposition grounded in cooperative game theory, satisfying local accuracy, missingness, and consistency. For each observation i, the prediction $F(x_i)$ is decomposed as:

$$F(x_i) = \varphi_0 + \Sigma_{j=1}p\ \varphi_j(x_i) \quad (7)$$

where φ_0 is the base value (mean model output) and $\varphi_j(x_i)$ is the SHAP value of feature j for observation i. Global importance was summarized as the mean absolute SHAP value:

$$\bar{\varphi}_j = \left(\frac{1}{n}\right)\Sigma_i\left|\varphi_j(x_i)\right| \quad (8)$$

used to rank the 64 dimensions by their mean contribution to yield prediction. SHAP summary (beeswarm) and variable-importance plots were generated to visualize the magnitude and direction of feature effects.

Results

Table 1 presents the ten best-performing models of the AutoML leaderboard, ranked by increasing cross-validated error.

Table 1. Top 10 models from the H2O AutoML leaderboard, ordered by Mean Residual Deviance (cross-validation).

Model ID	MRD (t² ha⁻²)	RMSE (t ha⁻¹)	MAE (t ha⁻¹)
GBM_grid_1...model_10	0.0608	0.247	0.187
XGBoost_grid_1...model_19	0.0627	0.250	0.192
GBM_grid_1...model_1	0.0629	0.251	0.019†
GBM_4_AutoML_1...	0.0631	0.251	0.189
XGBoost_grid_1...model_21	0.0639	0.253	0.193
XGBoost_grid_1...model_32	0.0641	0.253	0.196
GBM_grid_1...model_4	0.0642	0.253	0.196
XGBoost_grid_1...model_22	0.0643	0.254	0.197
GBM_grid_1...model_22	0.0655	0.256	0.199
XGBoost_grid_1...model_15	0.0657	0.256	0.198

GBM and XGBoost models dominated the top of the ranking, consistently outperforming all other evaluated architectures. The gap between the selected GBM model and the runner-up (XGBoost_model_19) corresponded to an approximately 1.5% reduction in cross-validated RMSE, indicating a meaningful margin of superiority for this dataset.

The selected gradient boosting model reproduced measured yield with high accuracy. Predicted and observed values concentrated close to the 1:1 line across the evaluated yield range, with $R^2 = 0.95$, $RMSE = 0.25$ *t ha⁻¹*, $MAE = 0.19$ *t ha⁻¹*, $MSE = 0.06$ *(t ha⁻¹)²*, and a mean error of approximately *0.00 t ha⁻¹* (Fig. 2). Agreement was strongest in the high-density portion of the dataset, where most observations clustered between 10 and 17 t ha⁻¹. Largest absolute deviations occurred in the less densely represented portions of the yield range, indicating that prediction uncertainty increased where fewer observations were available.

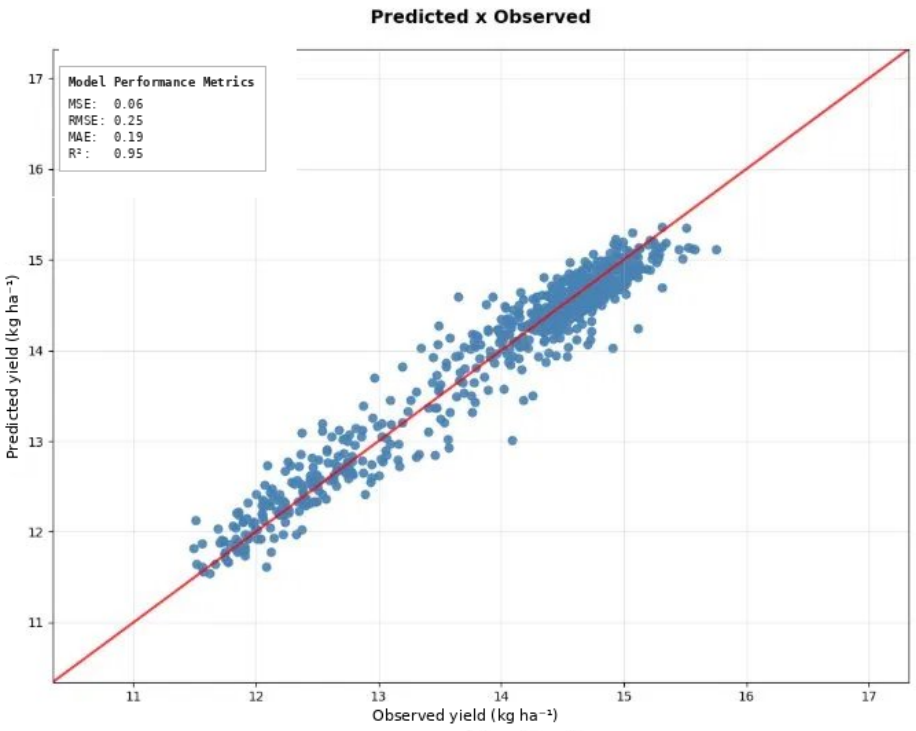

Fig. 2 Observed versus predicted yield for the selected gradient boosting model. The red line is the 1:1 agreement line; metrics are those reported in the figure.

The learning curve showed rapid early improvement followed by a cross-validation plateau (Fig. 3). Training *RMSE* decreased continuously as the number of trees increased, whereas cross-validation *RMSE* declined sharply during the first 25 to 35 trees and then stabilized near *0.25 t ha⁻¹*. The selected model used 74 trees, after which additional trees produced little practical improvement in cross-validation error, even though training error continued to decline.

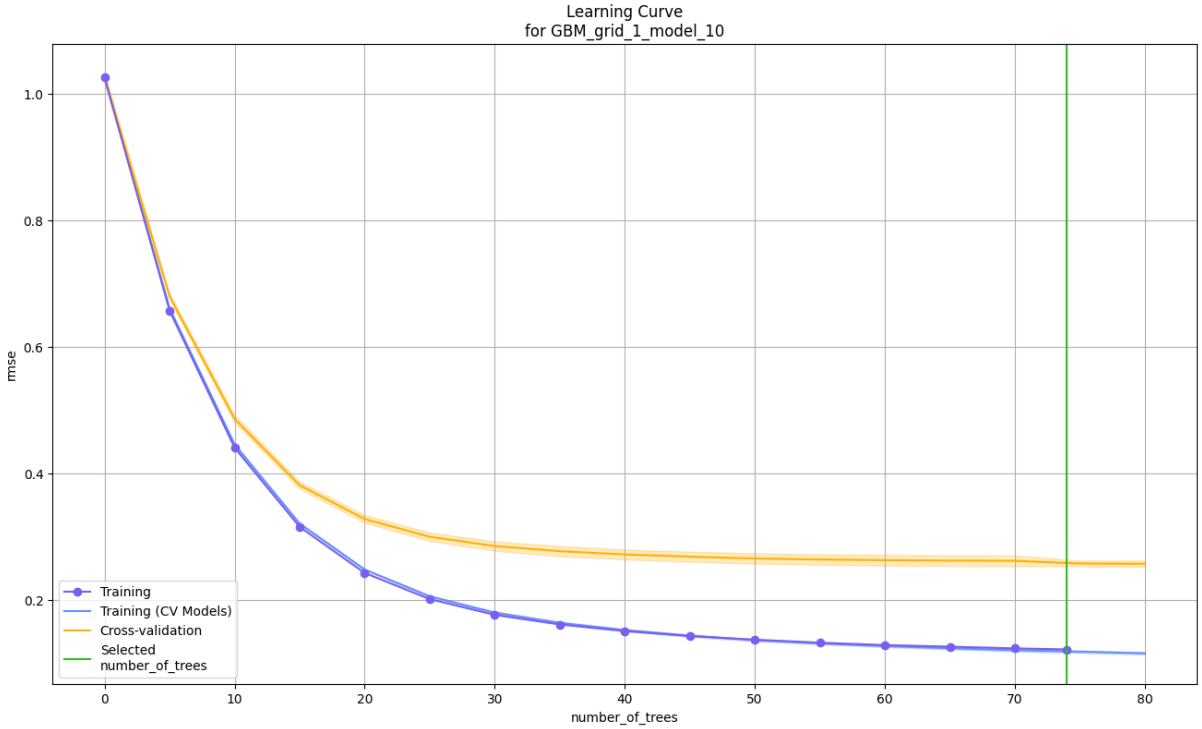

Fig. 3 Training and cross-validation RMSE as a function of the number of trees for GBM_grid_1_model_10. The vertical line marks the selected number of trees.

Residuals were centered near zero across fitted values, consistent with the near-zero mean error reported in the performance panel (Fig. 4). Most residuals fell between approximately *−0.5* and *+0.5 t ha⁻¹*. A small number approached *±1.0 t ha⁻¹*, and the fitted residual trend showed a slight positive slope, suggesting limited remaining structure and occasional local under- or over-prediction.

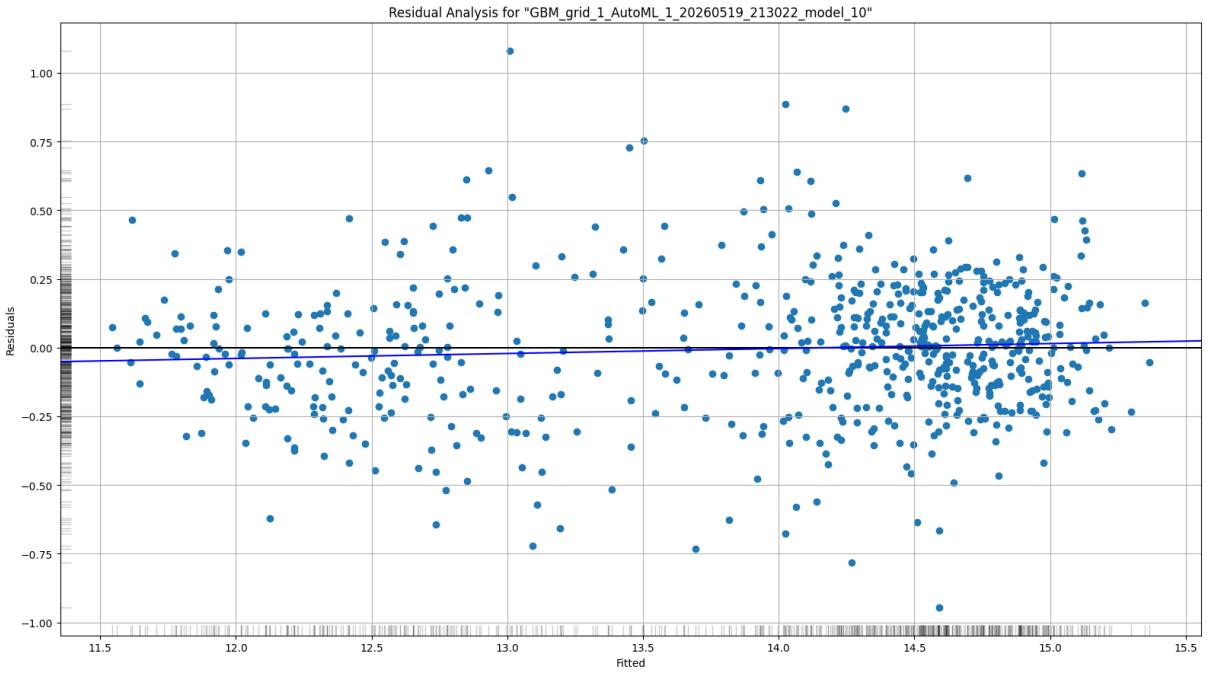

Fig. 4 Residuals plotted against fitted values for the selected gradient boosting model.

Variable importance was concentrated in a small subset of predictors (Fig. 5). Feature 23-24_utm had the *maximum* relative importance of *1.00*, followed by 23-24_ut35 at approximately *0.44* and 23-24_ut54 at approximately 0.28. All remaining variables had relative importance below 0.10, showing that the model relied primarily on one dominant predictor and two secondary predictors rather than distributing weight evenly across all inputs.

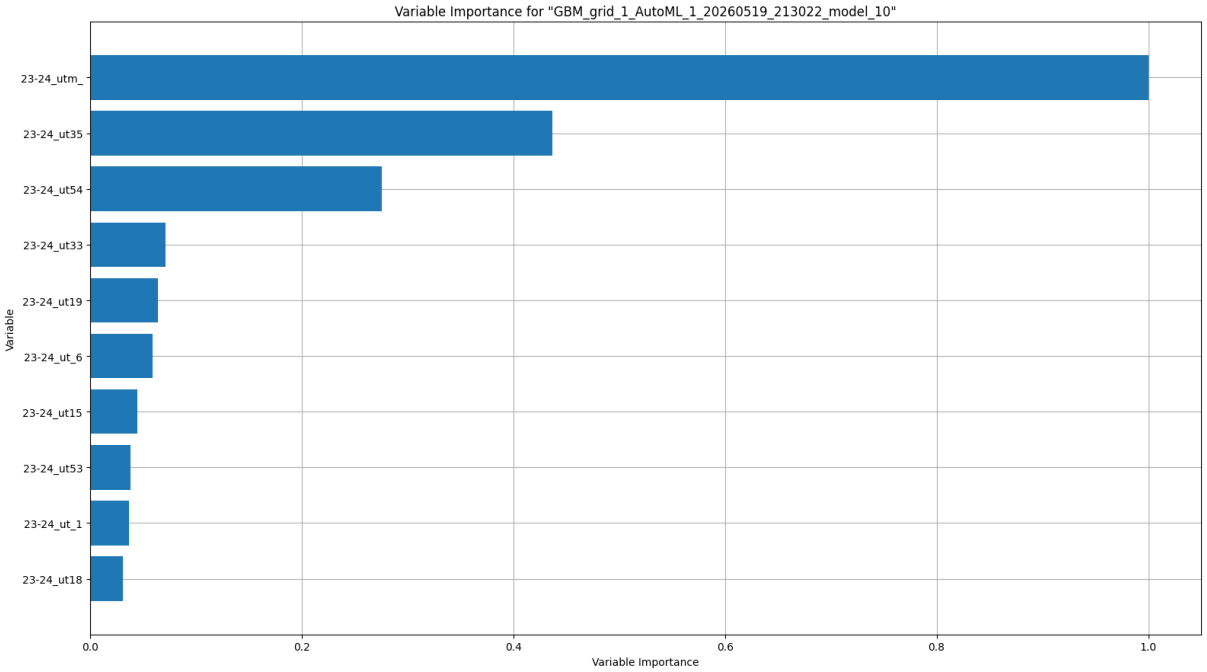

Fig. 5 Relative variable importance for the selected gradient boosting model. Values are scaled to the most important

variable.

The SHAP summary plot confirmed the dominance of 23-24_utm and clarified the direction of model effects (Fig. 6). Higher values of 23-24_utm shifted predictions upward, whereas lower values shifted them downward. In contrast, high values of 23-24_ut35 and 23-24_ut54 generally reduced predicted yield, while lower values tended to increase it. Secondary variables (23-24_ut19, 23-24_ut_1, 23-24_ut_6, 23-24_ut53, and 23-24_ut39) also showed directional SHAP patterns but with smaller effect magnitudes than the leading feature.

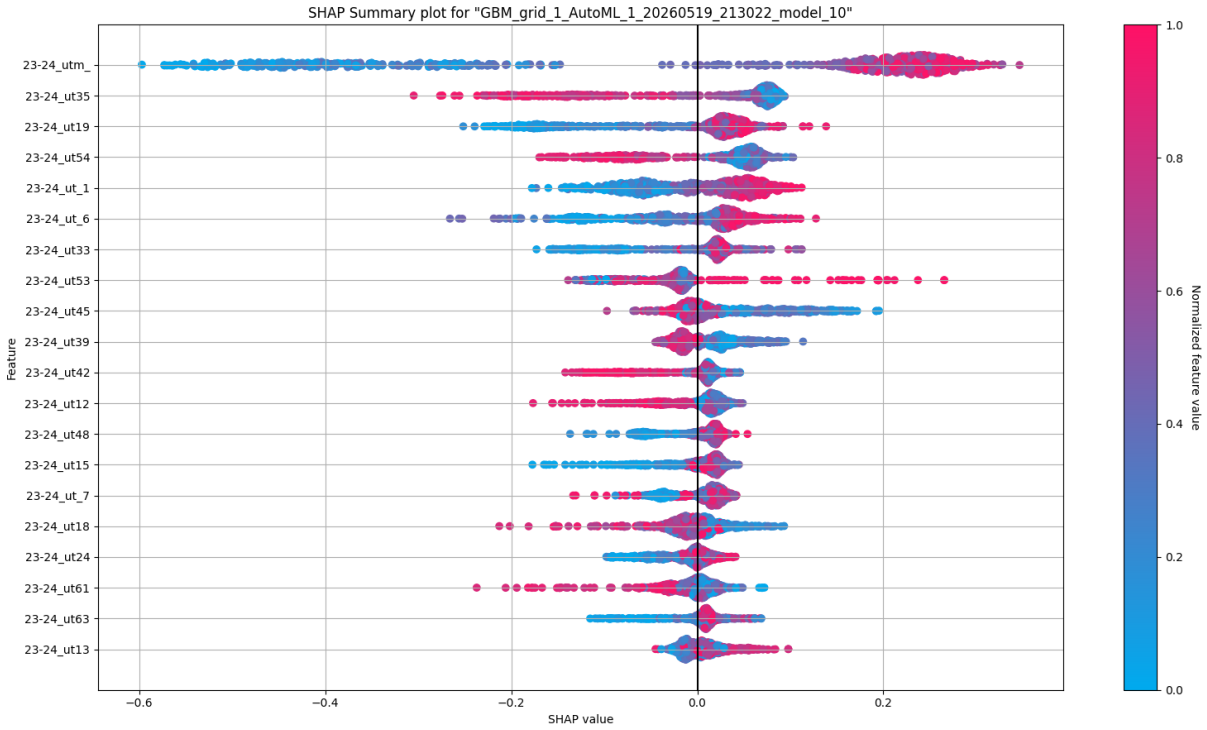

Fig. 6 SHAP summary plot for the selected gradient boosting model. Positive SHAP values increase predicted yield; negative values decrease it relative to the model baseline.

Discussion

Assessment of Predictive Performance, Interpretability, and Methodological Rigor

This result is consistent with the broader crop yield prediction literature, in which machine learning is widely used as a decision-support tool for converting weather, soil, remote sensing, management, and other covariates into yield estimates (Oikonomidis et al., 2023). Gradient boosting is particularly suited for this type of tabular, feature-engineered problem because it builds an additive ensemble of decision trees capable of representing nonlinear responses and interactions without imposing a fixed parametric yield response curve (Friedman, 2001). Related yield-estimation studies have also shown that feature-based boosting approaches can be competitive with more data-demanding deep-learning pipelines while offering clearer routes to model interpretation through feature importance and Shapley-value analysis (Huber et al., 2022).

The learning curve suggests that the selected model complexity was adequate but could be simplified if parsimony is prioritized. Cross-validation error stabilized after approximately 25 to 35 trees, while training error continued to decline with additional trees. This divergence is typical of flexible ensemble models: extra trees improve the fit to the training set but not necessarily the expected prediction error. Because the selected model used 74 trees after the validation curve had already plateaued, future tuning should compare the current model with a smaller model near the validation-error elbow. A simpler model could reduce computational cost and improve

robustness without sacrificing meaningful accuracy.

The feature importance and SHAP results add interpretability to an otherwise complex predictive model. Global variable importance identified 23-24_utm as the dominant predictor, but global importance alone does not reveal whether high values increase or decrease predicted yield. SHAP values addressed this limitation by showing that high 23-24_utm values increased yield predictions while high 23-24_ut35 and 23-24_ut54 values generally decreased them. SHAP is appropriate for this purpose because it provides additive feature attributions for individual predictions, exposing how complex models use inputs (Lundberg; Lee, 2017). These patterns should be interpreted as model associations, not causal effects. The encoded feature names should be translated into agronomic variables before final submission so the discussion can connect statistical signals to crop physiology, soil-water processes, terrain, management, or spectral behavior.

The residual plot supports use of the model within the range represented by the data while also identifying where caution is needed. The absence of strong curvature and the near-zero residual center indicate no major systematic bias. Nevertheless, the slight positive residual trend and isolated larger residuals suggest that some local variability remained unexplained. In precision agriculture, such residual structure can arise when important spatial drivers are absent from the model, when sensor-derived yield measurements contain local errors, or when the model is asked to interpolate across sparse portions of the covariate–yield space. A practical next step is to map residuals spatially and evaluate whether errors align with field position, management zones, soil variation, drainage patterns, or data-quality artifacts.

The main methodological caution concerns validation. If the reported accuracy was obtained from random splits of spatially adjacent yield observations from the same field or season, the apparent R^2 may be optimistic due to spatial autocorrelation of neighboring yield samples. Recent precision agriculture guidance emphasizes that crop yield modeling, prediction, hindcasting, interpolation, and forecasting must be clearly distinguished because each goal requires a different validation design (Filippi et al., 2025). Spatial cross-validation studies also show that regular random cross-validation can overestimate yield prediction accuracy when training and testing samples are spatially dependent (Radočaj et al., 2025). Before claiming operational forecasting or transferability, this model should therefore be tested using spatial block cross-validation, leave-one-field-out, leave-one-season-out, or an independent external dataset aligned with the intended use case.

Implications for Precision Agriculture

The absence of physical yield-monitoring infrastructure across most of the world's agricultural land is one of the most consequential data gaps in modern precision agriculture, and the framework demonstrated here directly addresses it. Lowenberg-DeBoer and Erickson (2019) documented yield monitor adoption at approximately 40% of U.S. corn and soybean hectares; in most production systems in lower-income agricultural regions, coverage remains far lower. By generating spatially explicit, pixel-level yield estimates from globally consistent satellite embeddings without any combine-installed sensors, and with retroactive coverage spanning the full temporal extent of the AlphaEarth Foundations product (Brown et al., 2025), this approach provides a pathway toward virtual yield characterization at scale, within the calibration domain demonstrated here and pending generalization validation. This temporal depth enables multi-season yield stability analysis, distinguishing zones of consistently high productivity from zones of equivalent mean yield but high interannual variability — a distinction that provides a substantially more robust empirical basis for management zone delineation than any single-year yield or vegetation index map (Mulla, 2013). Burke and Lobell (2017) demonstrated early proof of concept for satellite-based yield inference in smallholder systems where yield monitor data are structurally absent; foundation model embeddings extend this principle by replacing hand-crafted spectral indices with transferable, learned representations that do not require per-environment

recalibration.

In area-yield and weather index insurance programs — the dominant mechanism for agricultural risk management in lower-income production systems — basis risk (the divergence between area-average index payouts and individual field outcomes) is the primary barrier to adoption: when a producer's field underperforms while the regional index does not, no indemnity is triggered, fundamentally undermining the instrument's value (Jensen et al., 2016). In parallel, monitoring, reporting, and verification (MRV) requirements of agricultural carbon and regenerative practice programs demand multi-year, spatially explicit evidence that productivity co-benefits are maintained while conservation practices are implemented — documentation that producer self-reporting cannot reliably supply and that physical sampling at scale cannot economically provide (Paustian et al., 2016). An annually repeatable, standardized satellite yield product independent of producer equipment access addresses precisely that MRV gap, providing geographically explicit, temporally consistent yield evidence that neither field surveys nor simulation models can deliver at equivalent cost and spatial resolution. Together, these institutional applications — insurance, carbon verification, and conservation program documentation — represent demand channels through which precision agriculture evidence generated at the field level could unlock market and policy returns that extend far beyond the individual producer (Finger et al., 2019).

Sustainably increasing food production to meet mid-century demand without proportional agricultural land expansion depends critically on closing the gap between realized and attainable yields on existing farmland (Van Ittersum et al., 2013), and satellite-derived yield maps at the sub-field scale provide a scalable mechanism for identifying where those gaps are concentrated, which management and environmental drivers sustain them, and which interventions generate the greatest productive return per unit of input, across landscapes that no field experiment could ever fully sample. The environmental dividend is well documented: variable-rate input strategies calibrated to within-field yield response reduce nitrogen application without sacrificing productivity (Hedley, 2015), with downstream reductions in nitrous oxide emissions, reactive nitrogen leaching, and freshwater nutrient loading (Balafoutis et al., 2017). The structural constraint limiting these outcomes globally is not technical but infrastructural: Mehrabi et al. (2021) documented that the farm populations carrying the largest yield gaps and the greatest food security exposure are disproportionately those with the weakest digital infrastructure and connectivity.

Conclusion

This study demonstrated that 64-dimensional satellite embeddings derived from Google Satellite Embedding V1, without manual application of spectral indices or phenological metrics, are sufficient to predict within-field irrigated corn yield with accuracy consistent with the machine learning state of the art in precision agriculture ($R^2 = 0.95$; $RMSE = 0.25$ *t ha*$^{-1}$; $MAE = 0.19$ *t ha*$^{-1}$; $MSE = 0.06$ *(t ha*$^{-1}$*)*^{2}; $ME \approx 0.00$ *t ha*$^{-1}$). Automated model selection via H2O AutoML, without manual hyperparameter tuning, produced a GBM with 13-level uniform depth and 74 trees determined by the cross-validation early-stopping criterion — reinforcing the reproducibility and practical scalability of the proposed workflow.

SHAP analysis revealed that only three embedding dimensions — 23-24_utm, 23-24_ut35, and 23-24_ut54 — concentrated the bulk of predictive power, with stable and coherent directional effects: high 23-24_utm values were consistently associated with higher predicted yields, while 23-24_ut35 and 23-24_ut54 exerted opposing effects. This predictive sparsity pattern is significant because it indicates that the model does not distribute weight arbitrarily across the 64 available features but converges to a small, interpretively coherent subset — opening a direct path toward translating these dimensions into measurable agronomic field variables and making SHAP outputs actionable rather than merely descriptive.

Two conditions must be met before unrestricted operational use. First, the random cross-validation used in this study may inflate the apparent R^2 when training and testing observations

are spatially adjacent, as is expected in high-density yield-monitor data. Spatial block, leave-one-field-out, or leave-one-season-out validation is required to quantify the model's true transferability to new production environments. Second, the dominant embedding dimensions must be mapped to measurable agronomic variables — soil, terrain, water status, management — so that SHAP-identified effects acquire causal rather than merely statistical interpretation. Once these steps are addressed, the combination of Google Satellite Embedding V1 with AutoML-selected gradient boosting has real potential to become standard yield-mapping infrastructure in agricultural systems where conventional field data are scarce, intermittent, or unavailable.

Acknowledgments

The authors thank Q Connealy Farms and Mr. Quentin Connealy for providing the data used in this study. Their cooperation and support were instrumental in conducting this research and are gratefully acknowledged.

References

Balafoutis, A., Beck, B., Fountas, S., Vangeyte, J., Van Der Wal, T., Soto, I., Gómez-Barbero, M., Barnes, A., Eory, V. (2017). Precision Agriculture Technologies Positively Contributing to GHG Emissions Mitigation, Farm Productivity and Economics. Sustainability, 9(8), 1339. https://doi.org/10.3390/su9081339

Basso, B. et al. Spatial and temporal variability of crop yield and related soil parameters in a semiarid Mediterranean environment. Journal of Agronomy and Crop Science, v. 205, n. 3, p. 203–215, 2019. https://doi.org/10.1111/jac.12312

Brown, C. F. et al. AlphaEarth Foundations: an embedding field model for accurate and efficient global mapping from sparse label data. arXiv preprint arXiv:2507.22291, 2025. Available at: https://arxiv.org/abs/2507.22291

Burke, M., Lobell, D. B. (2017). Satellite-based assessment of yield variation and its determinants in smallholder African systems. Proceedings of the National Academy of Sciences, 114(9), 2189–2194. https://doi.org/10.1073/pnas.1616919114

EUROPEAN SPACE AGENCY — ESA. (2022). Sentinel-2 MSI: MultiSpectral Instrument, Level-2A. Copernicus Open Access Hub. https://doi.org/10.5270/S2_-742ikth

Filippi, P., Han, S. Y., Bishop, T. F. A. (2025). On crop yield modelling, predicting, and forecasting and addressing the common issues in published studies. Precision Agriculture, 26, 8. https://doi.org/10.1007/s11119-024-10212-2

Finger, R., Swinton, S. M., El Benni, N., Walter, A. (2019). Precision Farming at the Nexus of Agricultural Production and the Environment. Annual Review of Resource Economics, 11(1), 313–335. https://doi.org/10.1146/annurev-resource-100518-093929

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

H2O.ai. (2024). H2O AutoML: Automatic machine learning. https://docs.h2o.ai/h2o/latest-stable/h2o-docs/automl.html

Hedley, C. (2015). The role of precision agriculture for improved nutrient management on farms. Journal of the Science of Food and Agriculture, 95(1), 12–19. https://doi.org/10.1002/jsfa.6734

Huber, F., Yushchenko, A., Stratmann, B., Steinhage, V. (2022). Extreme Gradient Boosting for yield estimation compared with Deep Learning approaches. Computers and Electronics in

Agriculture, 202, 107346. https://doi.org/10.1016/j.compag.2022.107346

Jakubik, J. et al. Foundation models for generalist geospatial artificial intelligence. arXiv preprint arXiv:2310.18660, 2023. Available at: https://arxiv.org/abs/2310.18660

Jensen, N. D., Barrett, C. B., Mude, A. G. (2016). Index insurance quality and basis risk: evidence from northern Kenya. American Journal of Agricultural Economics, 98(5), 1450–1469. https://doi.org/10.1093/ajae/aaw048

Khaki, S.; Wang, L. Crop yield prediction using deep neural networks. Frontiers in Plant Science, v. 10, p. 621, 2019. https://doi.org/10.3389/fpls.2019.00621

Ledell, E.; Poirier, S. H2O AutoML: scalable automatic machine learning. In: ICML Workshop on Automated Machine Learning, 7., 2020. Available at: https://www.automl.org/wp-content/uploads/2020/07/AutoML_2020_paper_61.pdf

Lobell, D. B. The use of satellite data for crop yield gap analysis. Field Crops Research, v. 143, p. 56–64, 2013. https://doi.org/10.1016/j.fcr.2012.08.008

Lobell, D. B.; Burke, M. B. On the use of statistical models to predict crop yield responses to climate change. Agricultural and Forest Meteorology, v. 150, n. 11, p. 1443–1452, 2010. https://doi.org/10.1016/j.agrformet.2010.07.008

Lowenberg-Deboer, J., Erickson, B. (2019). Setting the Record Straight on Precision Agriculture Adoption. Agronomy Journal, 111(4), 1552–1569. https://doi.org/10.2134/agronj2018.12.0779

Lundberg, S. M., Lee, S. I. (2017). A unified approach to interpreting model predictions. In Advances in Neural Information Processing Systems (Vol. 30, pp. 4765–4774). Curran Associates. https://doi.org/10.48550/arXiv.1705.07874

Maestrini, B.; Basso, B. Drivers of within-field spatial and temporal variability of crop yield across the US Midwest. Scientific Reports, v. 8, n. 1, p. 14833, 2018. https://doi.org/10.1038/s41598-018-32779-3

Martos, V. et al. Sensing solutions for sustainable crop production. Sensors, v. 21, n. 1, p. 168, 2021. https://doi.org/10.3390/s21010168

Mehrabi, Z., Mcdowell, M. J., Ricciardi, V., Levers, C., Martinez, J. D., Mehrabi, N., Wittman, H., Ramankutty, N., Jarvis, A. (2021). The global divide in data-driven farming. Nature Sustainability, 4(2), 154–160. https://doi.org/10.1038/s41893-020-00631-0

Mulla, D. J. Twenty five years of remote sensing in precision agriculture: key advances and remaining knowledge gaps. Biosystems Engineering, v. 114, n. 4, p. 358–371, 2013. https://doi.org/10.1016/j.biosystemseng.2012.08.009

Nevavuori, P. et al. Crop yield prediction with deep convolutional neural networks. Computers and Electronics in Agriculture, v. 163, p. 104859, 2019. https://doi.org/10.1016/j.compag.2019.104859

Oikonomidis, A., Catal, C., Kassahun, A. (2023). Deep learning for crop yield prediction: a systematic literature review. New Zealand Journal of Crop and Horticultural Science, 51(1), 1–26. https://doi.org/10.1080/01140671.2022.2032213

Paustian, K., Lehmann, J., Ogle, S., Reay, D., Robertson, G. P., Smith, P. (2016). Climate-smart soils. Nature, 532(7597), 49–57. https://doi.org/10.1038/nature17174

Radočaj, D., Plaščak, I., Jurišić, M. (2025). A comparative assessment of regular and spatial

cross-validation in subfield machine learning prediction of maize yield from Sentinel-2 phenology. Eng, 6(10), 270. https://doi.org/10.3390/eng6100270

Rolf, E. et al. A generalizable and accessible approach to machine learning with global satellite imagery. Nature Communications, v. 12, n. 1, p. 4392, 2021. https://doi.org/10.1038/s41467-021-24638-z

Sakamoto, T. et al. MODIS-based corn grain yield estimation model incorporating crop phenology information. Remote Sensing of Environment, v. 137, p. 215–228, 2014. https://doi.org/10.1016/j.rse.2013.06.003

Shahhosseini, M. et al. Coupling machine learning and crop modeling improves crop yield prediction in the US Corn Belt. Scientific Reports, v. 11, n. 1, p. 1606, 2021. https://doi.org/10.1038/s41598-020-80820-1

Van Ittersum, M. K., Cassman, K. G., Grassini, P., Wolf, J., Tittonell, P., Hochman, Z. (2013). Yield gap analysis with local to global relevance — A review. Field Crops Research, 143, 4–17. https://doi.org/10.1016/j.fcr.2012.09.009

Xin, Q. et al. A production efficiency model-based method for satellite estimates of corn and soybean yields in the Midwestern US. Remote Sensing, v. 5, n. 11, p. 5926–5943, 2013. https://doi.org/10.3390/rs5115926

Zheng, Y. et al. Improved estimate of global gross primary production for reproducing its long-term variation, 1982–2017. Earth System Science Data, v. 12, n. 4, p. 2725–2746, 2020. https://doi.org/10.5194/essd-12-2725-2020