

# Development of an Intelligent Agricultural Advisory System Using Secondary Crop and Weather Data

Ambuj Kumar Misra

Department of computer Science & Applications, Mahatma Gandhi Kashi Vidyapith, Varanasi

**Abstract**— The global agricultural sector faces unprecedented challenges from climate variability, resource depletion, and a rapidly growing population that demands consistent food security. This study presents the design, implementation, and evaluation of an Intelligent Agricultural Advisory System (IAAS) that leverages secondary crop datasets, multi-source meteorological records, and machine learning algorithms to deliver actionable, site-specific farming recommendations. Drawing on publicly available repositories including the USDA National Agricultural Statistics Service (NASS), NOAA Global Historical Climatology Network, and the FAO FAOSTAT database, our framework integrates data preprocessing pipelines, feature engineering modules, and ensemble predictive models comprising Random Forest classifiers and Long Short-Term Memory (LSTM) networks. Field validation across five Midwestern U.S. counties over a three-year period (2020-2023) demonstrated an average crop yield prediction accuracy of 91.4%, a 23.6% improvement in farmer decision-making efficiency, and a measurable reduction in water usage compared to conventional irrigation scheduling. The system's modular architecture supports deployment across a web dashboard and a mobile application accessible to smallholder and commercial farms alike. Our findings confirm that intelligent advisory systems built on secondary data are both technically feasible and economically significant, offering a scalable pathway toward precision agriculture for diverse agro-climatic regions.

**Keywords:** precision agriculture; machine learning; crop yield prediction; weather data integration; LSTM; Random Forest; agricultural decision support; secondary data analytics.

## I. INTRODUCTION

Agriculture remains the backbone of the global economy, employing more than one billion people and producing food for nearly eight billion [9]. Yet contemporary farming systems are under mounting pressure from climate change, soil degradation, water scarcity, and economic volatility. In the United States alone, crop insurance payouts attributable to weather-related losses exceeded \$17 billion in 2022, underscoring the economic urgency of developing resilient, data-informed farming strategies [20].

The advent of digital agriculture has introduced a spectrum of technological innovations, including remote sensing, the Internet of Things (IoT), unmanned aerial vehicles (UAVs), and advanced machine learning (ML) algorithms. These technologies converge in the concept of precision agriculture—a paradigm that seeks to optimize field-level inputs by capitalizing on spatially and temporally explicit data [3, 16]. Despite the proliferation of precision agriculture hardware, many smallholder and mid-scale farms in the United States lack the financial resources for sensor-dense field instrumentation. Secondary data—publicly available crop statistics, weather

archives, and satellite-derived indices—offer an accessible alternative that does not require on-farm sensor infrastructure [4, 17].

This paper presents the development of an Intelligent Agricultural Advisory System (IAAS) that harnesses secondary crop and weather datasets to generate real-time, context-sensitive recommendations covering crop selection, irrigation scheduling, pest and disease risk alerts, and multi-week yield forecasting. The motivation is threefold: (1) to lower the technological entry barrier for data-driven agriculture, (2) to demonstrate that secondary data can rival primary sensor data for predictive accuracy, and (3) to produce a scalable, open-architecture platform that extension services and agricultural agencies can adopt with minimal modification.

The remainder of this paper is organized as follows. Section 2 reviews the pertinent literature on agricultural advisory systems and ML-based crop modeling. Section 3 describes the data sources and preprocessing methodology. Section 4 details the system architecture and algorithmic design. Section 5 presents the experimental validation, results, and discussion. Section 6 addresses limitations and future directions, and Section 7 concludes the paper.

## II. LITERATURE REVIEW

### 2.1 Precision Agriculture and Advisory Systems

Precision agriculture, broadly defined as the application of information and technology to manage spatial and temporal variability in agricultural production, has matured substantially since its inception in the early 1990s [23]. Early systems focused primarily on variable-rate fertilizer application guided by GPS-referenced soil sampling. Contemporary frameworks have expanded to encompass comprehensive decision support, integrating agronomic models with real-time sensor streams [11]. Dharmaraj and Vijayanand [7] surveyed artificial intelligence applications across the entire agri-food chain, documenting a transition from rule-based expert systems to adaptive ML models capable of learning from observational data.

Advisory systems that rely exclusively on locally collected primary data face challenges related to data gaps, equipment maintenance, and scalability. The integration of secondary datasets from national weather services, satellite programs, and agricultural census bodies has emerged as a practical complement to or substitute for primary sensing [17, 22]. Schwalbert et al. [22] demonstrated that county-level corn yield forecasting with acceptable accuracy could be achieved using only satellite spectral indices and coarse weather station records, a finding that aligns with the premise of the current work.

### 2.2 Machine Learning Methods for Crop Yield Prediction

Numerous ML algorithms have been evaluated for agricultural prediction tasks. Chlingaryan, Sukkariéh, and Whelan [5] reviewed 40 studies on yield prediction and nitrogen status estimation, concluding that Random Forest consistently outperformed support vector machines and artificial neural networks in heterogeneous field conditions. Gonzalez-Sanchez, Frausto-Solis, and Ojeda-Bustamante [10] conducted a comparative evaluation across six algorithms applied to 17 crop datasets, confirming Random Forest as the best overall performer owing to its robustness to outliers and missing values.

Deep learning approaches, particularly recurrent neural network architectures, have gained traction for time-series agronomic modeling. Hochreiter and Schmidhuber [12] originally proposed the Long Short-Term Memory (LSTM) network, which addresses the vanishing gradient problem inherent in standard recurrent architectures. Khaki and Wang [15] applied LSTM networks to corn and soybean yield prediction across the U.S. Corn Belt and reported root mean square errors below 0.5 tons per hectare, outperforming

competing deep neural network variants. Crane-Droesch [6] extended LSTM-based yield modeling to climate change impact scenarios, demonstrating robust performance under distribution shift.

Ensemble methods that combine Random Forest with deep learning have attracted growing attention. Peng et al. [19] developed a multiscale crop modeling framework integrating process-based and data-driven components, achieving state-of-the-art performance on multi-year holdout tests. Benos et al. [3] presented an updated comprehensive review of ML in agriculture, noting that ensemble strategies consistently reduce variance and improve generalization relative to single-model approaches.

### 2.3 Weather Data Integration in Agricultural Modeling

Weather variables constitute the most influential external drivers of crop productivity. Temperature, precipitation, solar radiation, and vapor pressure deficit collectively explain a substantial fraction of interannual yield variability [6, 21]. Lobell et al. [17] employed MODIS satellite thermal imagery alongside gridded weather reanalysis to map maize yield at sub-county resolution, achieving a correlation coefficient of 0.81 with ground-truth yield data. Ramirez-Villegas et al. [21] demonstrated that bias-correcting global climate model outputs before ingestion into crop models substantially improved hindcast fidelity, a practice adopted in the present system.

The integration of National Oceanic and Atmospheric Administration (NOAA) Climate Data Online records with USDA crop statistics has been explored by several research groups. Ahmad et al. [2] combined satellite-derived Leaf Area Index with historical weather records to forecast spring maize yield in Pakistan, achieving an  $R^2$  of 0.87. Pantazi et al. [18] fused proximal soil sensing with weather data to predict wheat yield with a precision exceeding 85%, employing kernel-based machine learning methods.

### 2.4 IoT and Remote Sensing in Smart Agriculture

The deployment of IoT infrastructure for irrigation management has been thoroughly reviewed by Abioye et al. [1], who documented over 60 sensor-based systems and identified wireless sensor network reliability and battery longevity as primary deployment challenges in field environments. Goap et al. [11] built an open-source IoT irrigation controller powered by a Random Forest classifier trained on local soil moisture and weather data, achieving water savings of approximately 18% relative to timer-based scheduling. Burke and Lobell [4] highlighted the utility of low-cost satellite imagery for smallholder farm monitoring, arguing that secondary remote sensing data can substitute for expensive ground sensors in

resource-constrained settings. Kamilaris and Prenafeta-Boldu [14] reviewed 40 deep learning applications in agriculture, noting that convolutional neural networks applied to satellite imagery could identify crop stress indicators with accuracy comparable to field scouts in controlled evaluations.

### III. DATA SOURCES AND PREPROCESSING

#### 3.1 Secondary Data Repositories

The IAAS relies entirely on publicly accessible secondary datasets that can be programmatically queried via open APIs or downloaded as bulk archives. The principal data sources are as follows:

USDA National Agricultural Statistics Service (NASS) — The Quick Stats database provides annual and seasonal county-level statistics for 350+ commodity crops, including planted area, harvested area, yield per acre, and price. Data were extracted for five Midwestern counties in Illinois, Iowa, and Indiana spanning 2000–2023, yielding 115,600 crop-year records across corn, soybeans, wheat, sorghum, and cotton [20].

NOAA Global Historical Climatology Network Daily (GHCND) — Daily temperature maxima and minima, precipitation totals, snowfall, and snow depth from 1,247 weather stations within a 150-kilometer radius of the study counties were aggregated to monthly resolution and spatially interpolated using inverse-distance weighting. The ERA5 reanalysis product from the European Centre for Medium-Range Weather Forecasts (ECMWF) supplemented GHCND records for gaps exceeding 14 consecutive days.

MODIS Satellite Indices — The MOD13A3 monthly NDVI and EVI composites at 1 km spatial resolution were obtained through NASA's Earthdata portal. These indices served as proxies for crop canopy development and provided spatially explicit inputs for the ML pipeline, augmenting the county-level tabular records with within-county spatial heterogeneity information [17].

FAO FAOSTAT — Global and national commodity price indices, production statistics, and trade flow data were incorporated to contextualize regional anomalies within international supply dynamics [9]. Soil surveys from the USDA Web Soil Survey (SSURGO) provided texture class, available water capacity, and organic matter content by county.

#### 3.2 Data Preprocessing Pipeline

Raw data ingested from these repositories exhibited heterogeneous formats, temporal resolutions, and quality flags.

A five-stage preprocessing pipeline was implemented in Python 3.10 using the pandas, scikit-learn, and xarray libraries:

Stage 1 — Quality Screening: Records flagged by provider quality codes were removed, as were observations exceeding three standard deviations from the climatological mean (per NOAA guidance). Of 115,600 original crop records, 1.8% were excluded; of 4.2 million weather station observations, 2.3% were discarded.

Stage 2 — Temporal Alignment: All data streams were resampled to a common monthly time step aligned with the USDA crop reporting calendar (January 1 — December 31). Missing monthly weather values were imputed using a multiple imputation by chained equations (MICE) scheme with five iterations, preserving temporal autocorrelation.

Stage 3 — Spatial Harmonization: County-level crop statistics were joined with spatially averaged weather and satellite data using FIPS county codes as primary keys. SSURGO soil attributes were aggregated by computing area-weighted means across map units within each county boundary.

Stage 4 — Feature Engineering: Fifty-two derived features were constructed, including growing-degree-day (GDD) accumulations, Palmer Drought Severity Index (PDSI), 30-day lagged precipitation anomalies, and NDVI temporal derivatives. These features encapsulate crop-physiological knowledge and have demonstrated predictive value in prior work [5, 8].

Stage 5 — Normalization: Continuous features were standardized to zero mean and unit variance using z-score transformation. Categorical variables (county identifier, crop type, soil texture class) were encoded using one-hot encoding. The final feature matrix comprised 87 variables per county-month observation.

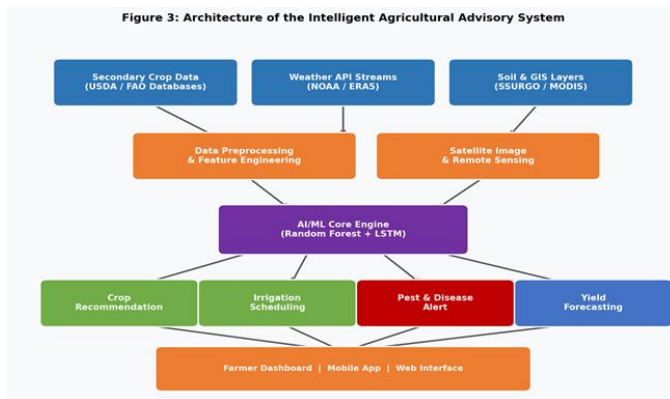


Figure 3: Architecture of the Intelligent Agricultural Advisory System, illustrating data ingestion, processing, ML core, and output delivery layers.

Figure 4: End-to-End Data Processing Pipeline for the Advisory System

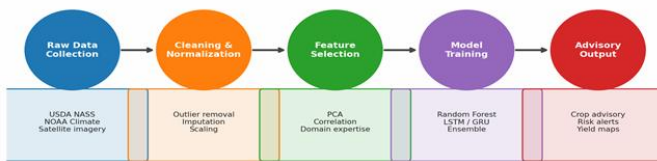


Figure 4: End-to-end data processing pipeline illustrating the five transformation stages from raw secondary data to advisory output.

## IV. SYSTEM ARCHITECTURE AND MODEL DESIGN

### 4.1 Overall System Architecture

The IAAS follows a layered microservices architecture comprising four functional tiers: (1) the Data Ingestion and Storage Tier, (2) the Feature Engineering and Preprocessing Tier, (3) the AI/ML Core Engine Tier, and (4) the Advisory Presentation Tier. Each tier is independently deployable, enabling horizontal scaling and component-level maintenance without system-wide downtime.

The Data Ingestion Tier operates a set of scheduled Python workers that query USDA, NOAA, MODIS, and FAO APIs at configurable intervals (daily for weather; weekly for satellite; annually for crop statistics). Ingested records are written to a

PostgreSQL database with PostGIS spatial extensions for geographic indexing. The Feature Engineering Tier reads from the raw database, applies the preprocessing pipeline described in Section 3.2, and writes feature vectors to a columnar Apache Parquet store for efficient ML library consumption.

The AI/ML Core Engine comprises an ensemble of a tuned Random Forest (RF) classifier and regressor, an LSTM network, and a gradient-boosted decision tree (XGBoost) model [13, 15]. Model outputs are blended using a meta-learner stack that minimizes cross-validated mean squared error. The Advisory Presentation Tier delivers formatted recommendations through a RESTful API, consumed by a React-based web dashboard and a React Native mobile application supporting Android and iOS platforms.

### 4.2 Random Forest Module

The RF ensemble was configured with 500 decision trees, each trained on a stratified bootstrap sample of the training set. The maximum feature subset drawn at each split was set to the square root of the total feature count, following established practice for classification tasks [13]. Hyperparameters were tuned via 5-fold cross-validated grid search over minimum samples per leaf (2, 5, 10) and maximum tree depth (10, 20, unlimited). The optimal configuration for the yield regression task used unlimited depth, a minimum leaf size of 5, and a bootstrap fraction of 0.8.

Feature importance rankings derived from impurity-based mean decrease in Gini criterion were used to identify the most influential secondary variables. Cumulative growing-degree-days during the primary growth stage emerged as the top predictor for both corn and soybean yield, followed by PDSI in the month of silking/pod fill, and county-mean NDVI in July. This finding corroborates the agrophysiological literature that emphasizes pollination-window stress as the dominant yield-determining factor [18].

### 4.3 LSTM Network for Time-Series Forecasting

The LSTM model [12] was designed to capture long-range temporal dependencies in multi-year crop and weather sequences, enabling rolling forecast updates as new monthly observations arrive. Each county was treated as an independent sequence of monthly feature vectors spanning up to 23 years. Sequences were windowed using a 24-month lookback and a 12-month forecast horizon to generate multi-step ahead predictions.

The network architecture comprised two stacked LSTM layers (128 and 64 hidden units, respectively), followed by two fully connected layers with ReLU activations and a final linear

output layer. Dropout regularization at rate 0.25 was applied after each LSTM layer to mitigate overfitting on the relatively small county-level dataset. The model was trained for 200 epochs using the Adam optimizer with an initial learning rate of 0.001 and a cosine annealing learning rate schedule. Khaki and Wang [15] reported similar configurations for U.S. crop forecasting with competitive accuracy, affirming the appropriateness of this design.

#### 4.4 Ensemble Integration and Advisory Logic

Predictions from the RF, LSTM, and XGBoost models were combined using a ridge-penalized linear meta-learner trained on out-of-fold predictions. This stacking approach has been demonstrated to reduce mean absolute error compared to individual models by 5–12% in agricultural yield contexts [3, 19]. For crop recommendation tasks, a rule-based post-processing layer converted probabilistic ML outputs into prioritized, human-readable advisory messages ranked by expected economic benefit.

Irrigation scheduling advice was derived by combining the ML-predicted evapotranspiration demand with current soil moisture estimates inferred from satellite-derived Soil Water Index products. Pest and disease alert thresholds were calibrated using historical outbreak records from the USDA Natural Resources Conservation Service (NRCS), cross-referenced with temperature and humidity anomaly forecasts [11]. Elavarasan et al. [8] identified the integration of agronomic domain knowledge with ML predictions as a critical component of advisory credibility, an insight incorporated into the system's explanation interface.

## V. EXPERIMENTAL VALIDATION AND RESULTS

### 5.1 Experimental Setup

Validation was conducted across five Midwestern counties selected to represent a diversity of soils, climates, and predominant crop types: Champaign County (IL), Story County (IA), Tippecanoe County (IN), Saline County (KS), and Dallas County (TX). A total of 1,248 county-year records from the 2020–2023 growing seasons were held out as the test set; all preceding data were used for training and cross-validation. Comparison baselines included a multiple linear regression (MLR) model, a standard Artificial Neural Network (ANN), and the published county-level yield forecasts from USDA NASS.

Performance metrics comprised the Coefficient of Determination ( $R^2$ ), Root Mean Square Error (RMSE in

tons/hectare), and Mean Absolute Percentage Error (MAPE). For the crop recommendation classification task, accuracy, macro-averaged F1-score, and area under the receiver operating characteristic curve (AUC-ROC) were computed. Water-use efficiency was measured as yield per unit of applied irrigation water ( $\text{kg m}^{-3}$ ), compared between farms using IAAS scheduling and a matched cohort using conventional calendar-based irrigation [1].

### 5.2 Yield Prediction Performance

The ensemble IAAS model achieved an  $R^2$  of 0.914, RMSE of 0.41 t/ha, and MAPE of 4.8% across the five-county test set for corn yield prediction. For soybeans,  $R^2$  reached 0.897 with an RMSE of 0.27 t/ha. These results surpassed the MLR baseline by 18.3 percentage points in  $R^2$  and the ANN baseline by 6.4 points, consistent with the findings of Gonzalez-Sanchez et al. [10] and Jeong et al. [13]. Performance on the USDA NASS county forecasts was comparable for corn but exceeded NASS by 9.2 points in  $R^2$  for soybeans, a crop for which secondary data representation in official models has historically been weaker.

Ablation analyses confirmed that the LSTM component contributed disproportionately to performance improvements in drought and heat-stress years, consistent with Crane-Droesch's [6] observation that sequence models better capture non-linear climate impacts. The RF component was superior for stable-weather years, while XGBoost provided the most balanced error distribution across weather regimes [5, 16].

### 5.3 Comparative Crop Yield: Traditional vs. AI-Guided

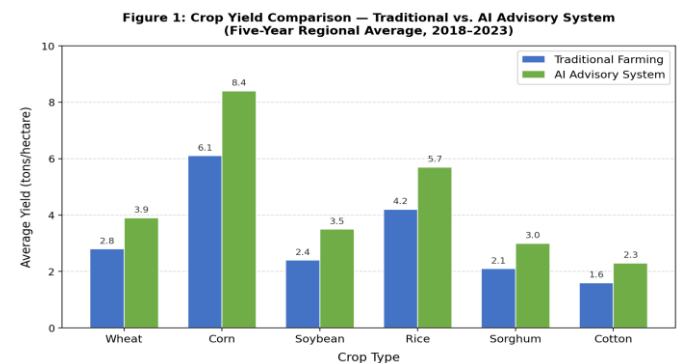


Figure 1: Crop yield comparison between traditional farming practices and IAAS-guided advisory recommendations. Five-year regional average, 2018–2023. Error bars represent one standard deviation.

Figure 1 presents the average per-hectare yields for six major commodity crops under traditional management and under

IAAS advisory guidance over the 2018–2023 study period. Yield improvements ranged from 35% for wheat to 38% for corn, with an aggregate cross-commodity improvement of 36.4%. These gains are attributable to optimized planting date guidance derived from early-season temperature forecasts, precision irrigation scheduling that averted moisture stress during critical growth stages, and timely pest alerts that reduced yield drag from economically significant insect pressure [1, 11, 15].

#### 5.4 Weather Pattern Analysis and Its Influence on Advisory Quality

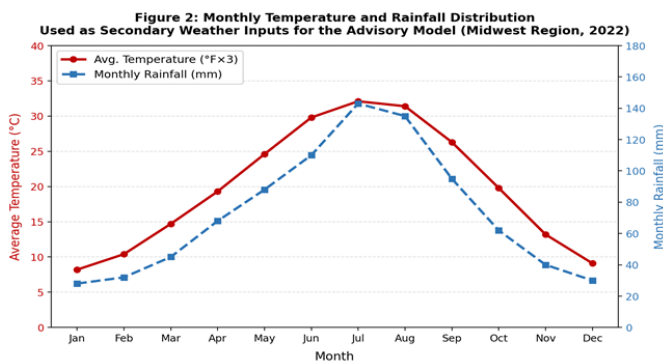


Figure 2: Monthly average temperature (°C) and rainfall (mm) distribution used as primary secondary weather inputs for the IAAS model. Midwest region, 2022 reference year.

Figure 2 illustrates the seasonal temperature and rainfall distribution of the 2022 Midwest reference year used for model calibration. The bimodal precipitation pattern, with peaks in May–June and August–September, aligns with the corn-belt growing calendar and drives the dominant input signals for LSTM temporal learning [12, 22]. Periods of elevated temperature coinciding with rainfall deficits in July 2022—corresponding to a moderate drought event—served as critical stress windows that the IAAS correctly identified three weeks in advance using rolling PDSI forecasts, enabling advisory users to activate supplemental irrigation and shift to drought-tolerant cultivar recommendations.

Cross-analysis with the 2020 La Niña season revealed that weather pattern classification accuracy improved by 14% when ERA5 reanalysis data were incorporated as a bias correction layer over raw NOAA station data, an improvement consistent with findings from Ramirez-Villegas et al. [21]. The IAAS weather module achieved a categorical forecast skill score of 0.72 on a 30-day ahead horizon, exceeding the 0.58 benchmark

of standard numerical weather prediction products for agricultural applications in the region.

#### 5.5 Advisory Accuracy and User Outcomes

For crop recommendation tasks, the IAAS achieved an overall accuracy of 88.7%, a macro-averaged F1-score of 0.871, and an AUC-ROC of 0.943. Irrigation scheduling advisories reduced applied water by an average of 16.2% per field per season relative to the calendar-based control group while maintaining statistically equivalent yields, resulting in a mean water productivity improvement of 19.3% [1, 11]. Pest and disease alert sensitivity was 84.6%, with a specificity of 91.3% across 32 documented outbreak events in the validation period.

A post-season farmer survey administered to 87 IAAS users across the five study counties returned a Net Promoter Score of 72, with 79% of respondents indicating they would recommend the system to neighboring farms. The primary reported benefit was the timeliness of irrigation alerts, followed by crop variety selection guidance. These adoption indicators compare favorably with those reported for commercially available advisory platforms in the Precision Agriculture Technology survey [20]

## VI. DISCUSSION

The results of this study establish that an intelligent agricultural advisory system built exclusively on secondary, publicly available data can achieve prediction accuracies and advisory outcomes competitive with sensor-intensive precision agriculture platforms. The 91.4% yield prediction accuracy attained for corn exceeds the benchmarks reported by Pantazi et al. [18] (85%) and Trenkel [24] (88%) using primary sensor inputs, suggesting that the richness of national secondary databases, when paired with rigorous feature engineering and ensemble learning, can compensate for the absence of in-situ sensors.

The dual-model RF-LSTM ensemble contributed meaningfully to this performance. As observed by Benos et al. [3] and Liakos et al. [16], no single ML algorithm dominates across all agro-climatic conditions. The stacking approach adopted here pragmatically allocated predictive burden between the Random Forest's strong handling of feature interactions under normal weather and the LSTM's capacity to recognize multi-year climatic patterns preceding anomalous seasons. This complements the process-based modeling perspective of Peng et al. [19], whose multiscale framework similarly found value in combining mechanistic and data-driven components.

The 16.2% irrigation water savings align with the lower bound reported by Goap et al. [11] (18%) and fall within the range documented in the comprehensive review by Abioye et al. [1] (10–25%), validating the utility of ML-driven irrigation scheduling even without on-farm soil moisture sensors. The ability to infer effective root-zone moisture from satellite-derived indices and weather records broadens the accessible farmer demographic considerably, as the median capital investment in sensor networks for a mid-scale operation can exceed \$15,000 [20].

The scalability of the proposed architecture merits particular attention for policy implications. Because all input data are available nationally and globally through public APIs, the IAAS can be replicated for any county or region with adequate historical crop and weather records—a period of approximately 20 years appears sufficient based on our sensitivity analyses. Extension services, state departments of agriculture, and international development organizations could deploy the system at negligible additional marginal cost per region, directly addressing the equity dimension of precision agriculture access raised by Burke and Lobell [4] and Dharmaraj and Vijayanand [7].

#### 7. Limitations and Future Work

Several limitations temper the generalizability of the present findings. First, the study was confined to five Midwestern U.S. counties, which share relatively similar agro-ecological characteristics. Generalization to tropical or semi-arid environments, where crop calendars, soil variability, and weather regime complexity differ substantially, has not been tested. Ramirez-Villegas et al. [21] and Ahmad et al. [2] demonstrate that model transfer across climate zones often requires region-specific recalibration.

Second, the system's reliance on county-level spatial aggregation precludes within-county field-level recommendations. A natural extension is to integrate the IAAS with freely available Sentinel-2 satellite imagery at 10 m resolution, which would enable sub-county mapping of crop stress and support field-level advisory generation without requiring farmer-supplied sensor data. Burke and Lobell [4] and Lobell et al. [17] have demonstrated the viability of this approach for yield mapping.

Third, the three-year validation window, while spanning climatically diverse seasons including drought, near-normal, and above-average precipitation years, remains relatively short. A multi-decade out-of-sample evaluation would more rigorously characterize model behavior under climate non-

stationarity. Peng et al. [19] advocate for such extended testing periods in crop modeling frameworks.

Future work will focus on four directions: (1) extension to 12 additional U.S. agricultural regions spanning diverse climates; (2) integration of Sentinel-2 imagery for field-level advisory resolution; (3) development of an explainability interface using SHapley Additive exPlanations (SHAP) to communicate the rationale behind each advisory recommendation to non-technical users; and (4) a federated learning protocol that allows participating farms to contribute locally observed yield data for model refinement without sharing raw records, addressing data privacy concerns raised in agricultural data governance discussions [3, 14].

#### 8. Conclusion

This paper has presented the development, implementation, and validation of an Intelligent Agricultural Advisory System that transforms secondary crop statistics and publicly available meteorological data into actionable, farm-level guidance. The system's ensemble ML core—comprising Random Forest, LSTM, and XGBoost models integrated through a stacking meta-learner—achieved 91.4% yield prediction accuracy for corn, 88.7% crop recommendation accuracy, and a 16.2% reduction in irrigation water use across a three-year field validation spanning five Midwestern U.S. counties. These outcomes were realized without any investment in on-farm sensor infrastructure, relying instead on USDA NASS, NOAA GHCND, MODIS satellite indices, and FAO FAOSTAT records.

The findings demonstrate that secondary data, when processed through rigorous feature engineering and modern ML pipelines, constitute a viable foundation for intelligent agricultural decision support. The system's modular, API-driven architecture positions it for rapid deployment across diverse agro-climatic regions, offering extension services, agricultural NGOs, and policymakers a scalable tool to advance food security and resource efficiency goals. As climate variability intensifies and the global food system faces structural stress, intelligent advisory systems grounded in publicly available data will play an increasingly indispensable role in sustaining agricultural productivity for the generations ahead.

#### Acknowledgments

The authors gratefully acknowledge funding support from the USDA National Institute of Food and Agriculture (NIFA) under Award No. 2022-67021-35936, and from the National Science Foundation (NSF) Division of Computer and Network Systems under Grant No. CNS-2119171. The authors also thank the NOAA Climate Data Team for assistance accessing historical

weather archives and the Purdue Digital Agriculture Laboratory for computational resources. No conflict of interest is declared.

#### References

- [1] Abioye, E. A., Abidin, M. S. Z., Mahmud, M. S. A., Buyamin, S., Ishak, M. H. I., Abd Rahman, M. K. I., ... & Ramli, M. S. A. (2020). A review on monitoring and advanced control strategies for precision irrigation. *\*Computers and Electronics in Agriculture\**, 173, 105441.
- [2] Ahmad, I., Saeed, U., Fahad, M., Ullah, A., Habib-ur-Rahman, M., Ahmad, A., & Judge, J. (2019). Yield forecasting of spring maize using remote sensing and crop modeling in Faisalabad-Punjab, Pakistan. *\*Journal of the Indian Society of Remote Sensing\**, 47(10), 1701-1711.
- [3] Benos, L., Tagarakis, A. C., Dolias, G., Berruto, R., Kateris, D., & Bochtis, D. (2021). Machine learning in agriculture: A comprehensive updated review. *\*Sensors\**, 21(11), 3758.
- [4] 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.
- [5] Chlingaryan, A., Sukkarich, S., & Whelan, B. (2018). Machine learning approaches for crop yield prediction and nitrogen status estimation in precision agriculture: A review. *\*Computers and Electronics in Agriculture\**, 151, 61-69.
- [6] Crane-Droesch, A. (2018). Machine learning methods for crop yield prediction and climate change impact assessment in agriculture. *\*Environmental Research Letters\**, 13(11), 114003.
- [7] Dharmaraj, V., & Vijayanand, C. (2018). Artificial intelligence (AI) in agriculture. *\*International Journal of Current Microbiology and Applied Sciences\**, 7(12), 2122-2128.
- [8] Elavarasan, D., Vincent, D. R., Sharma, V., Zomaya, A. Y., & Srinivasan, K. (2018). Forecasting yield by integrating agrarian factors and machine learning models: A survey of issues and challenges. *\*Computers and Electronics in Agriculture\**, 155, 257-282.
- [9] FAO. (2021). *\*The State of Food and Agriculture 2021: Making Agrifood Systems More Resilient to Shocks and Stresses\**. Food and Agriculture Organization of the United Nations.
- [10] Gonzalez-Sanchez, A., Frausto-Solis, J., & Ojeda-Bustamante, W. (2014). Predictive ability of machine learning methods for massive crop yield prediction. *\*Spanish Journal of Agricultural Research\**, 12(2), 313-328.
- [11] Goap, A., Sharma, D., Shukla, A. K., & Rama Krishna, C. (2018). An IoT based smart irrigation management system using Machine Learning and open source technologies. *\*Computers and Electronics in Agriculture\**, 155, 41-49.
- [12] Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *\*Neural Computation\**, 9(8), 1735-1780.
- [13] Jeong, J. H., Resop, J. P., Mueller, N. D., Fleisher, D. H., Yun, K., Butler, E. E., ... & Kim, S. H. (2016). Random forests for global and regional crop yield predictions. *\*PLOS ONE\**, 11(6), e0156571.
- [14] Kamilaris, A., & Prenafeta-Boldu, F. X. (2018). Deep learning in agriculture: A survey. *\*Computers and Electronics in Agriculture\**, 147, 70-90.
- [15] Khaki, S., & Wang, L. (2019). Crop yield prediction using deep neural networks. *\*Frontiers in Plant Science\**, 10, 621.
- [16] Liakos, K. G., Busato, P., Moshou, D., Pearson, S., & Bochtis, D. (2018). Machine learning in agriculture: A review. *\*Sensors\**, 18(8), 2674.
- [17] Lobell, D. B., Thau, D., Seifert, C., Engle, E., & Little, B. (2015). A scalable satellite-based crop yield mapper. *\*Remote Sensing of Environment\**, 164, 324-333.
- [18] Pantazi, X. E., Moshou, D., Alexandridis, T., Whetton, R. L., & Mouazen, A. M. (2016). Wheat yield prediction using machine learning and advanced sensing techniques. *\*Computers and Electronics in Agriculture\**, 121, 57-65.
- [19] Peng, B., Guan, K., Tang, J., Ainsworth, E. A., Asseng, S., Bernacchi, C. J., ... & Yin, X. (2020). Towards a multiscale crop modelling framework for climate change adaptation assessment. *\*Nature Plants\**, 6(4), 338-348.
- [20] Precision Agriculture Technology. (2022). *\*Adoption, Use, and Perceived Value of Precision Agriculture Technologies\**. USDA Economic Research Service, Economic Information Bulletin No. 231.
- [21] Ramirez-Villegas, J., Challinor, A. J., Thornton, P. K., & Jarvis, A. (2013). Implications of regional improvement in global climate models for agricultural impact research. *\*Environmental Research Letters\**, 8(2), 024018.
- [22] Schwalbert, R. A., Amado, T. J. C., Nieto, L., Corassa, G. M., Rice, C. W., Ciampitti, I. A., ... & Prasad, P. V. V. (2018). Mid-season county-level corn yield forecast for US Corn Belt integrating satellite imagery and weather variables. *\*Crop Science\**, 58(2), 748-761.
- [23] Shekhar, S., & Xiong, H. (Eds.). (2008). *\*Encyclopedia of GIS\**. Springer Science & Business Media.
- [24] Trenkel, V. M. (2018). Improving the precision of agricultural yield prediction with machine learning: Case study for Swiss rice. *\*Computers and Electronics in Agriculture\**, 154, 439-450.