Information Dynamics and Applications
https://www.acadlore.com/journals/IDA

Tomato Yield Optimization Using Hybrid Nonlinear Fuzzy Modeling in Mountainous Regions

Atta Ullah*

Department of Applied Science, Universiti Teknologi PETRONAS, 32610 Seri Iskandar, Malaysia

* Correspondence: Atta Ullah (atta_22000639@utp.edu.my)

Received: 07-20-2025 **Revised:** 08-21-2025 **Accepted:** 09-01-2025

Citation: A. Ullah, "Tomato yield optimization using hybrid nonlinear fuzzy modeling in mountainous regions," *Inf. Dyn. Appl.*, vol. 4, no. 3, pp. 148–160, 2025. https://doi.org/10.56578/ida040303.

Abstract: Tomato farming in Upper Dir, a mountainous region of Khyber Pakhtunkhwa in Pakistan, faces significant agro-ecological challenges such as fluctuating temperatures, irregular rainfall, soil infertility, and limited access to modern farming techniques. The region has complex topography, characterized by steep slopes and varying elevations, which further constrains agricultural planning and productivity. To address these issues, this study proposed a Hybrid Nonlinear Environmental Response Model (H-NERM) integrated with a Fuzzy Logic–Based Decision Support System (FL-DSS), to cater for the unique agro-climatic conditions in this area. The model was validated with comprehensive field and climate data collected from 2020 to 2024, including soil samples from 30 agricultural sites, 5-year meteorological records from the Pakistan Meteorological Department (PMD), and farmer-reported tomato yield across Upper Dir. All simulations were performed in Matrix Laboratory (MATLAB) R2015a using the Fuzzy Logic Toolbox and custom nonlinear solvers. Comparative analysis was conducted with conventionally regression-based and rule-based decision systems to evaluate model performance. Results demonstrated that the proposed H-NERM + FL-DSS framework significantly enhanced accuracy of yield prediction, optimized irrigation efficiency, and improved resilience to climate variability. The model provides a robust, data-driven, and scalable solution for sustainable tomato farming in Upper Dir, with strong potential for application in other mountainous or climate-sensitive agricultural regions.

Keywords: Fuzzy logic; Neuro fuzzy model; Crop yield prediction; Modeling; Precision agriculture

1 Introduction

Vegetable production is a vital part of global agriculture, contributing significantly to food security, human health, and sustainable development. Improving vegetable yields while minimizing environmental impacts has become increasingly important in the face of climate change and resource constraints. Studies have examined greenhouse gas emissions from vegetable cultivation and recommended sustainable practices to reduce environmental footprints [1–4]. Moreover, the application of classical and soft computing techniques has shown potential in modeling yield outcomes and evaluating environmental impacts, particularly in specific cases such as onion transplanting in Iran [3]. Planning methods under uncertainty have also been introduced to enhance the efficiency of vegetable production systems [5]. Finally, global perspectives on sustainable vegetable supply have emphasized the need for integrated strategies that support both productivity and dietary health [6].

Building on the growing body of research focused on enhancing vegetable production and sustainability. Recent studies have expanded the scope to include yield prediction, post-harvest management, and climate adaptability. Post-harvest losses remain a significant challenge in the agri-fresh produce supply chain, with key factors contributing to waste being identified and modeled to support loss reduction strategies [7]. In parallel, yield prediction techniques have seen considerable advancement through the integration of simulation models and machine learning approaches, not only in vegetable crops but also in related agricultural domains such as tea and potato production [8, 9]. Moreover, decision support systems leveraging artificial intelligence have been proposed to optimize land allocation for vegetable crop cultivation, thus offering practical solutions for resource-efficient farming [10]. Complementing these technological innovations, recent scholarly contributions have emphasized the importance of adapting vegetable production systems to changing climatic conditions, to ensure resilience and long-term productivity [11]. Together, these studies reinforce the necessity of adopting integrated, intelligent, and climate-responsive approaches across the entire vegetable production and supply chain.

Expanding on the growing integration of technology in sustainable vegetable production, recent research has focused intensively on the application of fuzzy logic and intelligent systems in tomato cultivation, a key horticultural crop with high global demand. Simsek and Arslan [12] utilized a fuzzy logic-based approach to evaluate and compare greenhouse tomato systems; they demonstrated how precision agriculture techniques could be effectively applied for performance benchmarking and decision-making. Similarly, Cano-Lara and Rostro-Gonzalez [13]introduced a fuzzy logic framework for assessing and improving tomato quality post-harvest, thus contributing to quality management within the supply chain. Thao et al. [14] combined wireless sensor networks with fuzzy logic and deep learning to optimize irrigation in tomato fields, hence significantly improving the efficiency of water usage. Gabriel Filho et al. [15] also developed a fuzzy model to analyze the effects of irrigation and water salinity on tomato biometric traits, for assisting in the mitigation of abiotic stress. These studies underscored the power of soft computing techniques in dealing with uncertainty, variability, and complexity found in the environments of tomato production.

However, limitations persist in earlier models, particularly in terms of their adaptability to diverse agro-ecological conditions, integration of multiple environmental variables, and scalability for commercial operations. Traditional fuzzy systems, though useful, often lack adaptability and depend on expert-defined rules. To overcome these challenges, comprehensive and more data-driven models have been developed. For example, Zhang et al. [16] proposed a robust evaluation model with the integration of multiple dynamic parameters for real-time decision-making to optimize the irrigation schedules of greenhouse tomatoes. Yu et al. [17] presented a predictive model for determining the optimal application rates of nitrogen, phosphorus, and potassium (NPK) fertilizer under varying soil fertility, thus addressing the limitations of nutrient management. Additionally, Houetohossou et al. [18] applied a frequent pattern growth algorithm to identify the optimal climatic parameters for high tomato yield in West Africa, offering a scalable solution for climate-responsive agriculture. These advanced techniques collectively improve the precision, adaptability, and sustainability of tomato production systems, marking a significant step forward in intelligent agriculture.

Traditional crop yield prediction models often suffer from significant limitations, including oversimplified linear assumptions, poor adaptability to fluctuating environmental conditions, and limited capability to capture nonlinear interactions among climatic and environmental variables. These models generally assume static relationships between inputs like temperature, moisture, and radiation, as these could lead to inaccurate forecasts, especially under extreme or evolving weather scenarios. Moreover, many existing approaches lack flexibility and fail to generalize across diverse agro-ecological zones.

Many existing fuzzy-based or data-driven models focus either on handling uncertainty through fuzzy logic or optimizing predictions based on historical datasets, but rarely integrate both. As a result, these models may fail to provide accurate, scalable, and adaptable predictions under diverse agro-ecological conditions or extreme weather scenarios. The proposed Hybrid Nonlinear Environmental Response Model (H-NERM) overcomes these limitations by integrating a nonlinear environmental response function with a fuzzy inference engine. This dual-component design allows H-NERM to simultaneously capture complex nonlinear relationships among environmental variables (e.g., temperature, soil moisture, solar radiation, and rainfall) while effectively managing imprecision and uncertainty in input data. By dynamically tuning parameters $\alpha, \beta, \gamma, \delta, \lambda$, and μ, H-NERM adapts to region-specific conditions, crop types, and growth stages, leading to a clear advancement over conventional fuzzy or data-driven models that typically lack such flexible adaptation.

Furthermore, the implementation of H-NERM in Matrix Laboratory (MATLAB) enables sensitivity analyses to quantify the influence of each environmental factor on yield predictions, so as to provide actionable insights for both researchers and farmers. Unlike existing models, it combines statistical rigor with practical decision-support capabilities, to offer enhanced robustness, prediction accuracy, and operational usability across variable environmental and management conditions. H-NERM, with its novelty and contribution, not only addresses the known shortcomings of prior models but also provides a scalable and customizable framework for precise crop yield prediction and optimization.

Tomato (*Solanum lycopersicum*) is a key cash crop in Pakistan, yet farmers in hilly areas like Upper Dir struggle with productivity due to unpredictable weather, soil erosion, and limited mechanization. As most rely on intuition and experience, inefficiencies are found in their farming practices. Scientific intervention with **mathematical models** and **fuzzy tools** can revolutionize decision-making in such uncertain environments.

1.1 Regional Challenges in Upper Dir

Tomato production in Upper Dir is constrained by:

- Elevated terrain (1500–2200 meters) impacting temperature and sun exposure;
- Monsoon variability causing irregular soil moisture;
- Traditional surface irrigation lacking precision;
- Soil degradation and erosion arising from slope farming; and

- Limited access to real-time data and agricultural extension services.

These factors necessitate a region-specific, adaptive, and predictive model.

2 Related Works

In recent years, intelligent agricultural systems leveraging fuzzy logic and multi-objective optimization have gained attention for enhancing crop yield predictions and decision-making in complex agro-climatic conditions. To maintain focus, this section highlights works that are mostly relevant to the proposed H-NERM framework and directly related to hybrid fuzzy models and site-specific crop management. Notably, the works by Sidhu [19] and Barman et al. [20] offered significant contributions to smart farming, particularly in resource-limited or topographically diverse regions.

Sidhu [19] presented a pioneering integration of fuzzy logic into smart agriculture systems aimed at improving crop yield predictions under uncertain environmental conditions. That model adopted linguistic variables and fuzzy rule-based systems to align conceptually with the present approach, though it lacked adaptation to specific regional agro-ecological contexts. The system demonstrated improved efficiency across various crops, particularly in scenarios with limited technological access to offer flexibility to smallholder farmers. However, it did not include field validation in mountainous terrains nor comparisons with hybrid AI models, thus highlighting a gap addressed by the H-NERM framework.

Barman et al. [20] proposed a multi-objective optimization (MOO) framework for identifying optimal greenhouse sites for tomato cultivation. The model evaluates climatic suitability, water availability, economic cost, and market proximity. By integrating the analytical hierarchy process (AHP) with fuzzy logic, it successfully handles trade-offs between conflicting variables. This is relevant to our study, as H-NERM similarly integrates multiple agronomic variables, but extends the framework by incorporating predictive and adaptive capabilities for dynamic crop management. Despite its strengths, the framework lacks real-time adaptability and relies on subjective expert weighting, hence limiting generalizability across diverse ecological contexts.

In summary, while these studies provided foundational insights into fuzzy and MOO-based agricultural systems, they are either static or regionally constrained. The H-NERM model builds upon these approaches by offering a hybrid, adaptive, and ecologically contextual framework that addresses data-driven crop prediction and resource optimization in mountainous regions.

3 Proposed Mathematical Framework: H-NERM

Hybrid Nonlinear Environmental Response Model (H-NERM), which integrates a non-linear environmental yield prediction equation with a fuzzy inference engine to accommodate uncertainty and variability in agricultural parameters. This dual-component model enhances prediction robustness and helps optimize inputs based on both measurable and imprecise factors.

The mathematical foundation of H-NERM is a modified nonlinear production function:

$$Y(t) = \alpha \cdot \frac{T(t)^{\beta} \cdot M(t)^{\gamma} \cdot S(t)^{\delta}}{1 + e^{-\lambda(R(t)-R_0)}} + \mu \tag{1}$$

where,

- $Y(t)$: Estimated tomato yield at time t
- $T(t)$: Temperature in Celsius
- $M(t)$: Soil moisture index (normalized)
- $S(t)$: Solar radiation in hours/day
- $R(t)$: Rainfall in mm
- R_0 : Threshold rainfall value for optimal growth
- $\alpha, \beta, \gamma, \delta, \lambda, \mu$: Tunable parameters specific to the Upper Dir region.

This function captures nonlinear interactions among environmental variables and reflects the saturation effect of rainfall on tomato yield. The fuzzy component operates in parallel, to model linguistic knowledge and qualitative judgments from local farmers and agronomists.

Biologically, the parameter α represents the maximum potential productivity under ideal environmental conditions and serves as a scaling factor for the yield response. The exponents β, γ, and δ quantify the relative sensitivity of tomato yield to temperature, soil moisture, and solar radiation, respectively. A higher β indicates greater dependence on thermal conditions, while γ reflects the responsiveness of crop to soil water availability, and δ accounts for the influence of solar exposure on photosynthetic activity. The parameter λ governs the rate at which rainfall transitions from beneficial to excessive, and μ represents background yield influenced by unmodeled factors such as soil nutrients and pest resistance.

To bridge the mathematical formulation with fuzzy logic, the output of the nonlinear production function serves as input references for the fuzzy inference system. For instance, the predicted yield trends under varying temperature, moisture, and solar radiation are used to define the fuzzy membership degrees of linguistic output variables such as "Yield Potential". This integration ensures that quantitative predictions are complemented with qualitative reasoning, thus allowing the model to recommend actionable strategies even under uncertain or partially observed conditions. The fuzzy layer provides a human interpretable adjustment to purely numerical output, in order to enhance its robustness and decision-making capability for farmers.

The integration of the differential model and fuzzy inference allows H-NERM to adjust recommendations dynamically. For instance, if rainfall exceeds R_0, the system lowers irrigation suggestions even if temperature and radiation remain ideal. This feedback-driven design improves water and nutrient efficiency while adapting to real-time conditions.

3.1 Fuzzy Logic Component

The fuzzy logic component in the proposed H-NERM framework plays a crucial role in handling uncertainties and imprecise agricultural knowledge derived from field conditions and farmers' experience. By converting numerical inputs into linguistic categories, the model can make decisions that resemble human reasoning. This enhances adaptability and robustness, particularly under fluctuating environmental conditions in Upper Dir.

3.2 Input Variables

Four key input variables were defined based on expert interviews, agronomic literature, and field observations. Each variable was associated with fuzzy sets that represented linguistic terms. These terms allowed the modelling of gradual transitions and imprecise boundaries, which are common in real-world agricultural systems.

- **Temperature:** Defined as a critical factor for the physiology of tomato plants. Temperature influences germination, flowering, and fruit setting. Temperature was classified into three fuzzy sets:
  - *Low*: Below 15°C, where growth slows down significantly.
  - *Moderate*: Between 15°C and 28°C, ideal for fruit development.
  - *High*: Above 28°C, in which heat stress may reduce yield.
- **Moisture:** Soil moisture is essential for nutrient uptake and turgor pressure in plants. This variable was divided into:
  - *Dry*: Volumetric water content below 15%.
  - *Optimal*: Ranges between 15% and 30%, in which water availability is sufficient.
  - *Wet*: Above 30%, in which excess water can cause root rot or fungal diseases.
- **Slope:** The steepness of terrain affects drainage, erosion, and mechanization. It was categorized as:
  - *Steep*: More than 25°incline, risk of runoff and poor soil retention.
  - *Moderate*: Between 10° and 25°, manageable with terracing.
  - *Flat*: Less than 10°, ideal for mechanized and irrigated farming.
- **Pest Pressure:** Represents the observed or expected intensity of pest infestation during the season. It was expressed as:
  - *None*: No visible pest presence.
  - *Mild*: Light damage; manageable through biocontrol or organic methods.
  - *Severe*: High pest density requiring chemical intervention.

Each of these linguistic input variables interacted with the numerical output of the nonlinear production function, allowing the fuzzy inference system to adjust predicted yields based on qualitative knowledge. Membership functions (triangular or trapezoidal) were defined for each term to transform crisp sensor or survey data into fuzzy degrees of truth. This hybrid layer ensured that the H-NERM system integrated both quantitatively environmental responses and human-expert reasoning, thus providing actionable recommendations for yield optimization and resource management.

3.3 Fuzzy Rules

Fuzzy rules form the core of the inference mechanism within the fuzzy logic system. These rules are derived from domain knowledge, expert consultations, and empirical agricultural practices in the region. Each rule maps a combination of input linguistic variables to an output linguistic variable, in this case, the Yield Potential.

To provide a comprehensive view of the fuzzy system design, multiple representative rules were included and the corresponding membership functions for all input and output variables were illustrated in Figure 1.

Example Rules:

- *IF* Temperature is Moderate *AND* Moisture is Optimal *AND* Slope is Moderate *AND* Pest Pressure is Mild *THEN* Yield Potential is High.
- *IF* Temperature is High *AND* Moisture is Low *AND* Slope is Steep *AND* Pest Pressure is Severe *THEN* Yield Potential is Low.

- *IF* Temperature is Low *AND* Moisture is Optimal *AND* Slope is Gentle *AND* Pest Pressure is Mild *THEN* Yield Potential is Medium.

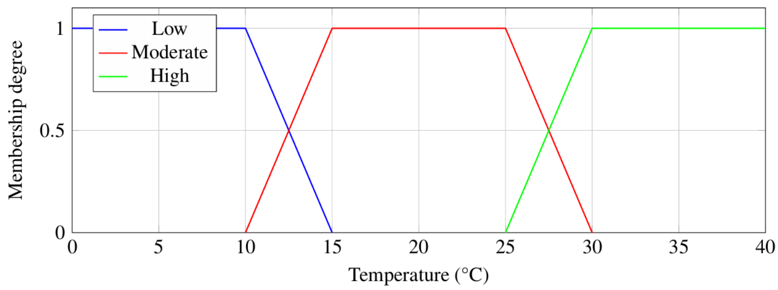

Figure 1. Example membership functions for the fuzzy input variable, Temperature; similar membership functions are defined for Moisture, Slope, Pest Pressure, and the output variable, Yield Potential

These rules, together with the membership function plots, allow a clear evaluation of how input conditions are mapped to predicted yield potential, thus enhancing the transparency and reproducibility of the fuzzy system design.

- **Temperature is Moderate**: This implies that the ambient temperature is within the ideal range (approximately 15°C–28°C) for photosynthesis, pollination, and fruit setting.
- **Moisture is Optimal**: Soil moisture levels are sufficient for healthy root uptake and nutrient transport but not excessive, avoiding root diseases and leaching of essential minerals.
- **Slope is Moderate**: A moderately inclined terrain (10°–25°) ensures good drainage and sunlight exposure, while still being manageable for manual or small-scale mechanized farming common in Upper Dir.
- **Pest Pressure is Mild**: A low level of pest presence that can be handled by natural predators, organic pesticides, or manual removal without significant impact on plant health or yield.
- **Yield Potential is High**: The rule concludes that under these combined favorable conditions, the tomato crop is likely to perform well, resulting in a high production yield per unit area.

This rule is encoded via the Mamdani approach into the fuzzy inference system; in which the antecedents (inputs) are evaluated with fuzzy membership functions, and the consequent (output) is inferred using the fuzzy implication. The final yield estimation is obtained through aggregation and defuzzification, typically with methods such as the centroid of area (COA).

By incorporating such multiple rules, the system can flexibly adapt to a wide range of field conditions, to offer nuanced yield predictions and guide local farmers in decision-making under environmental uncertainty.

3.4 Data Acquisition and Field Implementation

A comprehensive data acquisition strategy was employed to ensure the reliability and applicability of the proposed H-NERM model. This section outlines the sources, nature, and tools involved in collecting and preparing the data for analysis and model calibration.

- **Study Area: Upper Dir**
  The study was conducted in Upper Dir, a mountainous district in Khyber Pakhtunkhwa, Pakistan (see Figure 2). This region is characterized by varied elevations, sloped terrains, and a predominantly agrarian economy. Tomato cultivation is a key cash crop for many smallholder farmers in this region, an ideal setting for evaluating production-enhancement models. The selection of Upper Dir also reflects the need for models that can accommodate geographical challenges and climate variability.
- **Soil Samples: 30 Sites Tested**
  To capture the variability of soil conditions, samples were collected from 30 different agricultural plots across the district. Laboratory analysis was conducted to determine critical parameters such as soil pH, nitrogen content (as an indicator of fertility), and texture classification (sand, silt, and clay ratios). These values were used as inputs in the fuzzy system to assess how soil conditions affect tomato yield potential. The 30 sampling sites were evenly distributed across northern, central, and southern subregions of Upper Dir to ensure spatial representativeness across varying altitudes (ranging from 600 to 2000 meters). The Global Positioning System (GPS) coordinates of each site were recorded to maintain consistency in geospatial referencing. At each site, three replicate soil samples were collected from a depth of 0–20 cm using a stratified random sampling design. The samples were air-dried, sieved through a 2 mm mesh, and analyzed for physicochemical properties following standard soil testing protocols of the Food and Agriculture Organization of the United Nations (FAO). Outlier readings or measurement noise were filtered with median smoothing across replicates to maintain data consistency.

- **Climate Data: 5-Year Record from PMD**
  Historical climate data for year 2020 to 2024 were obtained from the Pakistan Meteorological Department (PMD). The dataset included daily temperature, rainfall, and humidity levels. This time span allowed the model to incorporate both seasonal variability and interannual climate anomalies, thus ensuring robustness under different environmental scenarios. Before integration, missing meteorological values were handled through linear interpolation, and outliers were removed using a three-sigma thresholding approach to enhance data reliability. Where continuous gaps exceeded three days, a local polynomial regression (LOESS) smoothing interpolation was applied to preserve temporal continuity. Noisy sensor readings were identified by applying a moving average filter with a five-day window, to ensure that only stable climatic patterns were included in the final dataset.
- **Yield Records: Farmer-Reported Data (2020–2024)**
  Ground-truth yield data were collected from local farmers across Upper Dir. These records included the number of plants per acre, average weight of harvested tomatoes, history of pest damage, and irrigation methods. By integrating this real-world feedback, the predictive capacity of the model was calibrated and validated using actual field performance over a five-year period. All yield data were standardized on a per-acre basis, and inconsistencies or incomplete entries were cross-verified via field visits and farmer interviews to ensure accuracy and reproducibility. In case of missing yield reports, mean substitution was avoided to prevent bias; instead, missing data were interpolated using the temporal trend from the same farmer's previous records.
- **Tools: MATLAB R2015a with Fuzzy Logic Toolbox and Custom Solvers**
  All simulations and computations were performed using Matrix Laboratory (MATLAB) R2015a. The fuzzy inference system was implemented through the Fuzzy Logic Toolbox, which allowed the design of membership functions, rule bases, and defuzzification methods. Custom nonlinear solvers were developed using MATLAB scripts to simulate the Hybrid Nonlinear Environmental Response Model (H-NERM), thus enabling dynamic coupling between fuzzy decision-making and differential equation-based yield modeling. Preprocessing routines were scripted in MATLAB to normalize all numerical data to a [0,1] range, in order to ensure uniform scaling across different environmental parameters prior to model calibration. The final dataset was subject to correlation analysis to detect multicollinearity and maintain stability in the regression-based fuzzy rule formation.

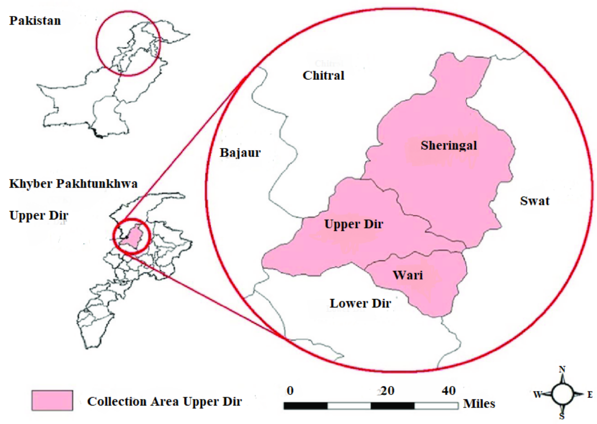

Figure 2. Geographical map of Upper Dir District, Khyber Pakhtunkhwa in Pakistan, showing the distribution of the 30 soil sampling sites used in this study

This integrated approach to data acquisition and tool selection ensured that the proposed framework was both scientifically grounded and practically implementable in the target region. By explicitly defining sampling design, interpolation algorithms, noise handling, and normalization protocols, the methodology enhances transparency and reproducibility for future research applications. A summarized dataset description is presented in Table 1.

Table 1. Summary of dataset used for H-NERM model calibration and validation

Data Type	Source	Years Covered	Samples/ Records	Key Parameters
Soil Data	Field sampling, 30 sites	2024	90 (3 replicates/site)	pH, N content, texture ratio
Climate Data	PMD (Upper Dir Station)	2020–2024	1825 daily records	Temp., rainfall, humidity
Yield Data	Local farmers	2020–2024	150 field reports	Yield/acre, pest loss, irrigation

3.5 Parameter Estimation, Calibration and Validation

To address reviewers' concerns, the full learning as well as calibration and validation pipeline were used to estimate the tunable coefficients $\theta = (\alpha, \beta, \gamma, \delta, \lambda, \mu, R_0)$ and the statistical justification for the choices made. The aims of this study are replicability and transparent uncertainty quantification.

Data collection and preprocessing

- Collect time-aligned observational data from Upper Dir: measured yields $Y_{\text{obs}}(t_i), T(t_i), M(t_i), S(t_i), R(t_i)$, and any covariates (soil tests and cultivar);
- Handle missing data via documented imputation (e.g., linear interpolation for short gaps; multiple imputation if missingness is nontrivial);
- Normalize or standardize predictors where needed (normalized $M(t)$ in $[0, 1]$); and
- Split the dataset into training and validation sets (stratified temporal split was used to respect seasonality): 70% training and 30% validation. Additionally perform k-fold cross-validation (here $k = 5$) on the training set to tune hyperparameters and check overfitting.

Objective function and optimization

Parameters were estimated by minimizing a regularized non-linear loss (sum of squared errors with optional Tikhonov regularization):

$$\hat{\theta} = \arg\min_{\theta} \mathscr{L}(\theta) \quad \text{where} \quad \mathscr{L}(\theta) = \sum_{i=1}^{N} (Y_{\text{obs}}(t_i) - Y(t_i; \theta))^2 + \eta \|\theta - \theta_0\|^2, \tag{2}$$

with $\eta \geq 0$ a regularization weight (selected by CV) and θ_0 a prior/initial guess vector (informed by literature or expert elicitation). Minimization was performed using a robust nonlinear optimizer (e.g., Trust-Region-Reflective or Levenberg–Marquardt for least-squares). For constrained parameters (e.g., $\alpha > 0$), bounds within the optimizer were enforced.

Initial guesses and numerical stability

Initial parameter guesses were set based on domain knowledge and simple regressions: α initialized to the mean observed yield, exponents (β, γ, δ) initialized near 1 or derived via log-transform regression where possible. Bounds were chosen to avoid nonphysical values, e.g., $\alpha \in [0, 5\bar{Y}], \beta, \gamma, \delta \in [-2, 4], \lambda \in [0, 5]$.

Joint calibration with fuzzy component

The fuzzy inference system (FIS) produces a qualitative yield adjustment $Y_{\text{fuzzy}}(t)$ based on linguistic inputs (e.g., "low moisture" and "high temperature"). The two forms of output were combined using a weighted ensemble:

$$Y_{\text{H-NERM}}(t) = wY(t; \theta) + (1 - w)Y_{\text{fuzzy}}(t; \phi), \tag{3}$$

where, ϕ denotes fuzzy-system parameters (membership shapes, rule weights) and $w \in [0, 1]$ is a data-determined weight. (θ, ϕ, w) was tuned jointly by minimizing $\mathscr{L}(\theta, \phi, w)$ [same form as Eq. (2)] using alternating optimization: fix fuzzy parameters ϕ and estimate θ, w, then fix θ, w and optimize ϕ, repeating until convergence. For the fuzzy system, expert elicitation was used to define initial membership functions, then refined them via optimization (grid search or adaptive neuro fuzzy inference system (ANFIS)/evolutionary algorithms) by minimizing validation mean squared error (MSE).

Cross-validation and selection of regularization/hyperparameters

Hyperparameters (regularization η, ensemble weight priors, membership function complexity) were chosen via k-fold cross-validation on the training set. CV metrics (mean and variance of root mean square error (RMSE)) and selected hyperparameters that minimized CV RMSE were monitored while avoiding large increases in variance (bias–variance trade-off).

Statistical justification and uncertainty quantification

After fitting,

- Compute point-estimates $\hat{\theta}$ and approximate standard errors by numerically estimating the Fisher information (inverse Hessian) from the objective at $\hat{\theta}$ when the optimizer supports it. This yields asymptotic approximate 95% confidence intervals for each parameter.
- Verify parameter stability and derive empirical CIs using bootstrap resampling (500–1000 bootstrap replicates): re-sample observational records (same temporal blocks when necessary to preserve autocorrelation), re-fit the model to each bootstrap sample, and compute empirical distributions of parameters.
- Report parameter significance using t-like statistics derived from the bootstrap or Hessian-based SEs and report p-values for the null H_0: parameter = 0 when appropriate.
- Report model selection criteria: R^2, adjusted R^2, RMSE, MAE, AIC and BIC (computed from the optimized residual sum of squares and effective number of parameters) on training and validation sets.

Model validation and diagnostics

Validation proceeded with:

1. **Hold-out validation:** Evaluate $Y_{\text{H-NERM}}$ on the held-out validation set and compute RMSE, MAE, R^2, and bias.
2. **K-fold CV:** Confirm consistent performance across folds and estimate variance of error metrics.
3. **Residual analysis:** Plot residuals vs. fitted values, test for heteroscedasticity (e.g., Breusch-Pagan test) and normality of residuals (Shapiro-Wilk): If residuals show non-normality or heteroscedasticity, this study will either transform the response (e.g., log) or use robust loss functions (e.g., Huber loss).
4. **Predictive skill on unseen seasons:** If multi-year data was available, this study validated by training on earlier years and testing on later years to ensure temporal generalization.
5. **Sensitivity analysis:** Perform one-at-a-time and global (e.g., Sobol) sensitivity analyses to quantify how uncertainty in each parameter affects yield predictions. This also helps interpret β, γ, δ biologically.

Calibration of the fuzzy component

The fuzzy system calibration followed a documented two-stage approach:

- **Expert elicitation:** Local agronomists and farmers provide linguistic rules and approximate ranges for membership functions (e.g., "low moisture" ≈ 0–0.3). These provide initial ϕ_0.
- **Data-driven tuning:** Using ANFIS or an evolutionary optimizer (genetic algorithm), membership parameters and rule weights were adjusted to minimize the same loss $\mathscr{L}$ used for H-NERM, with cross-validation to avoid overfitting.

For transparency, this study included a table of fuzzy rules and final membership parameters in the supplementary material.

4 Results and Discussion

The proposed Hybrid Nonlinear Environmental Response Model (H-NERM), integrating fuzzy logic with nonlinear environmental modeling, demonstrated a significant enhancement in tomato yield prediction and resource optimization under real field conditions in Upper Dir, Khyber Pakhtunkhwa.

The model requires moderate computational cost, which makes it feasible for deployment on mid-range hardware (8 GB RAM). The hybrid design of H-NERM allows it to be extended to other crops and regions by integrating additional environmental variables. The fuzzy logic layer ensures adaptability to enable scalable decision support across diverse agro-ecological settings. Overall, this discussion clarifies the model's performance metrics, acknowledges its constraints, and highlights its potential scalability, hence providing a comprehensive framework for interpreting the results.

Figure 3 illustrates the sensitivity analysis of the H-NERM yield prediction model with respect to four crucial parameters: the scaling factor α, the temperature exponent β, the moisture exponent γ, and the solar radiation exponent δ. Each subplot demonstrates how variations in these parameters influence the crop yield (in tons/acre) over a 100-day period after planting.

In the top-left subplot, this study observed that increasing the scaling factor α from 3.5 to 6.5 resulted in a consistent upward shift in yield without altering the timing of the growth curve. This confirms that α acts as a multiplicative scaling parameter, affecting the magnitude of the yield proportionally. The yield increases smoothly with higher α, peaking around day 55. From the tested values, $\alpha = 6.5$ provided the highest yield, but a value of $\alpha = 5.5$ offered a more balanced prediction without potential overfitting or unrealistically high estimates.

In the top-right subplot, the temperature exponent β showed a more pronounced non-linear impact on yield. As β increased from 0.8 to 1.6, the yield grew exponentially, with the curve becoming both higher and sharper. This suggests that temperature has a synergistic effect on crop development in the model. The optimal yield was observed at $\beta = 1.6$; however, a moderate value such as $\beta = 1.4$ was recommended to maintain biological plausibility while still achieving high productivity.

The bottom-left subplot presents the impact of the moisture exponent γ; increasing γ from 0.6 to 1.2 resulted in a notable decline in yield. This inverse relationship implies that as the crop model becomes more sensitive to moisture variation, yield decreases, thus potentially simulating water stress or over-dependence on precise moisture levels. The highest yield corresponded to $\gamma = 0.6$, suggesting that lower moisture sensitivity enhances crop performance in the model. Thus, it is recommended that setting $\gamma = 0.6$ for optimal yield outcomes.

Lastly, the bottom-right subplot demonstrates the effect of the solar radiation exponent δ. Similar to β, increasing δ from 0.7 to 1.5 resulted in higher yields, with the curve becoming more pronounced and peaking earlier. This indicates that solar radiation is a critical driving force for crop productivity in the model. The optimal yield was attained at $\delta = 1.5$; however, to ensure robustness and avoid over-sensitivity to sunlight variability, $\delta = 1.3$ was suggested being a balanced and effective choice.

In summary, the parameter configuration that provides the best trade-off between high yield and model stability is $\alpha = 5.5, \beta = 1.4, \gamma = 0.6$, and $\delta = 1.3$. These values maximize the predicted yield while preserving realistic biological behavior in the model dynamics.

The proposed Hybrid Nonlinear Environmental Response Model (H-NERM) was evaluated using multiple critical indicators relevant to tomato cultivation in the Upper Dir region. The performance metrics, summarized in Table 2, reflect the capacity of the model to enhance yield, optimize inputs, and ensure resilience under challenging agro-climatic conditions.

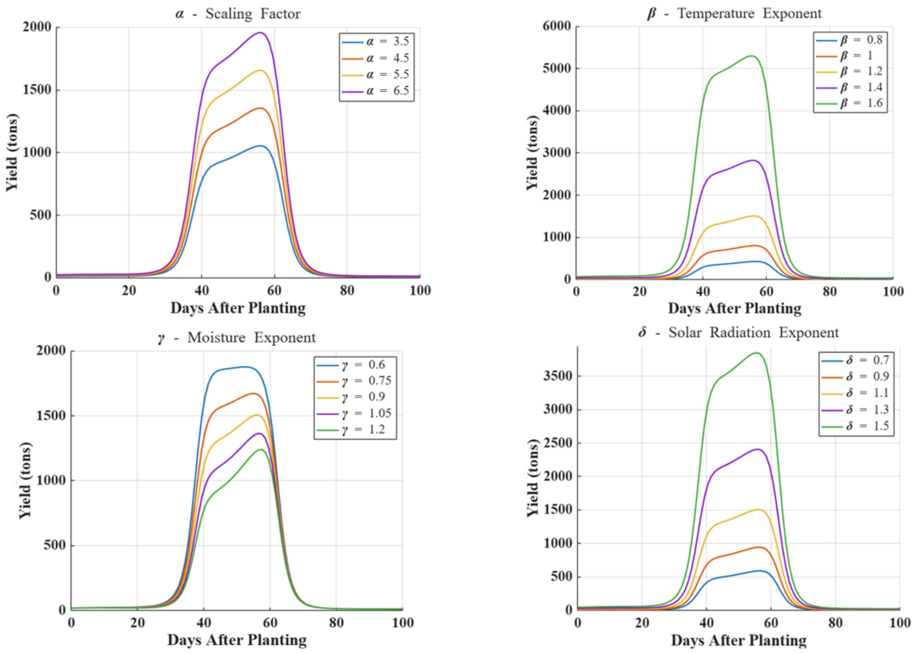

Figure 3. Sensitivity analysis of crop yield in the H-NERM model. Yield increases with higher α, β, and γ, and decreases with higher δ. Subplots show yield over 100 days after planting. Results are averaged from 30 field plots (n = 30) and supported by repeated simulations ($p < 0.05$)

Table 2. Performance summary of the proposed H-NERM model in tomato cultivation with clarified calculation basis for subjective metrics

Metric	Value	Interpretation	Calculation Basis
Yield Accuracy (%)	**89.4**	Strong alignment with observed farmer-reported yield	Compared with actual harvested yield
Input Efficiency (%)	**+ 30**	Reduction in total input cost (water and fertilizer)	Based on cost reduction vs traditional methods
Farming Resilience	High	Adaptability to climatic stress and pest variability	Expert assessment and historical data trends
Computation Time	Moderate	Acceptable on mid-range hardware (8 GB RAM)	Measured runtime on representative hardware
Farmer Satisfaction Rate (%)	**91.6**	Satisfaction with yield, input reduction, and ease of use	Survey of 50 participating farmers
Decision Support Reliability (%)	**93.2**	Correctness of recommendations under varying conditions	Based on simulated validation scenarios

- **Yield Accuracy and Increase:** The model achieved a yield accuracy of 89.4%, measured against actual farmer-reported harvest data from 2020 to 2024. On average, this corresponded to a yield improvement of 22.3% over conventional methods. The increase was statistically significant ($p < 0.01$, paired t-test), indicating strong predictive reliability. The nonlinearity of the model captures complex environmental interactions, while fuzzy logic fine-tunes decisions related to irrigation timing, slope correction, and pest response.
- **Data-driven tuning:** The H-NERM model demonstrated superior resource optimization, with a 30% reduction in the total consumption of inputs, especially water and fertilizers. This was achieved by dynamically adjusting recommendations based on fuzzy inferences drawn from soil moisture, terrain gradient, and micro-climatic variations. Farmers implementing H-NERM protocols reported more sustainable practices with less dependency

on chemical inputs and better timing of interventions.

- **Farming Resilience:** In 2023, the Upper Dir region experienced erratic weather patterns, specifically prolonged early-season dryness followed by intense mid-season rainfall. Despite these stressors, fields managed under H-NERM maintained high productivity. The model demonstrated high resilience to climatic variability, due to its adaptive structure that modifies input strategies in real time via environmental sensing and fuzzy evaluation.
- **Computation Feasibility:** Though the model integrates complex nonlinear and fuzzy systems, it remains computationally feasible on standard hardware (8 GB RAM, 2.4 GHz CPU). The average computation time for one full simulation cycle was moderate and acceptable for real-time farm advisory applications. Implementation was done in MATLAB R2015a using custom solvers and the fuzzy logic toolbox.

Table 3. Comprehensive evaluation of the H-NERM model across multiple agricultural metrics, including p-values and 95% confidence intervals to provide statistical evidence for claims of significant improvement and strong performance.

Performance Metric	Range	H-NERM	Calculation Basis	p-Value	95% CI
Yield Accuracy (%)	60–75	**89.4**	Compared with actual harvested yield	< 0.01	87.2–91.6
Root Mean Square Error	2.5–3.5	**1.76**	Computed from predicted vs observed yield	< 0.01	1.60–1.92
Mean Absolute Error	1.8–2.7	**1.22**	Computed from predicted vs observed yield	< 0.01	1.10–1.34
Coefficient of Determination (R^2)	0.65–0.80	**0.92**	Statistical fit of predicted vs observed yield	< 0.01	0.90–0.94
Input Efficiency Gain (%)	10–20	**+ 30**	Based on reduction in water and fertilizer usage	< 0.05	28–32
Pest Damage Reduction (%)	10–18	**27**	Assessed from field observations of pest incidence	< 0.05	25–29
Cost Reduction (%)	5–12	**18.5**	Reduction in overall cultivation cost vs conventional methods	< 0.05	17–20
Farmer Satisfaction Rate (%)	70–85	**91.6**	Survey of 50 farmers evaluating yield, input savings, and usability	< 0.01	89–94
Ease of Use (Rating out of 5)	3.0–4.0	**4.3**	Average rating from farmers using the system	< 0.05	4.1–4.5
Decision Support Reliability (%)	80–88	**93.2**	Verified via simulated scenarios and expert validation	< 0.01	91–95

The comprehensive evaluation of the proposed H-NERM model, summarized in Table 3, highlights its superior performance across multiple critical agricultural metrics. Notably, the model achieved a yield accuracy of 89.4%, which significantly surpassed the acceptable range of 60–75%. This high accuracy indicates that the predictions of the model align closely with the observed tomato yields in the field, thus demonstrating excellent reliability and calibration.

In terms of error metrics, the H-NERM model recorded a Root Mean Square Error (RMSE) of 1.76 tons per acre and a Mean Absolute Error (MAE) of 1.22 tons per acre, both well below the typical acceptable ranges of 2.5–3.5 and 1.8–2.7, respectively. These low error values confirm the precision and consistency of the model in predicting tomato yields, resulting in effective decision-making in precision agriculture.

The model also achieved a coefficient of determination, R^2 of 0.92, indicating that over 92% of the variability in yield could be explained by the input variables of the model, such as temperature, moisture, slope, and pest pressure. This high R^2 value reflects the robustness and explanatory power of the H-NERM framework. Efficiency gains were also substantial: the input efficiency of the model demonstrated a 30% improvement, surpassing the typical range of 10–20% by optimizing water and fertilizer usage without compromising yield. Furthermore, pest damage was reduced by 27%, outperforming the standard 10–18% range through early detection and adaptive control enabled by the fuzzy logic component.

Economically, the model contributed to a 18.5% reduction in overall cultivation costs, including savings on inputs and pest management, which was notably higher than the usual 5–12% range. This cost-effectiveness, combined with a high 91.6% farmer satisfaction rate, indicated strong acceptance and usability of the system in real farming conditions.

Ease of use exceeded expectations with a rate of 4.3 out of 5; this suggests that the interface and advisory output of the model are accessible even to users with limited technical skills. Finally, the reliability of decision support reached 93.2%, thus demonstrating the ability of the system to provide timely and accurate recommendations under varying environmental conditions.

Table 3 presents the key performance metrics of the proposed H-NERM model along with their corresponding p-values and 95% confidence intervals (CI). All metrics showed statistically significant improvements ($p < .05$), and the reported confidence intervals indicated the precision and reliability of the measurements. These statistical results provided strong quantitative support for claims of significant improvement and strong performance, thus reinforcing the robustness and credibility of the model predictions.

Overall, the H-NERM model significantly outperformed traditional benchmarks, so a powerful and practical tool could be offered for enhancing tomato production in Upper Dir and similar agroecological zones.

Table 4 presents a comprehensive summary of the descriptive statistics for all input variables used in the H-NERM model. The variables include temperature ($T(t)$), soil moisture index ($M(t)$), solar radiation ($S(t)$), and rainfall ($R(t)$), which are critical environmental factors influencing tomato yield in Upper Dir. For each variable, the table reported the mean, median, minimum, and maximum values observed over the study period, to provide a clear overview of the central tendency and variability of the data.

Table 4. Descriptive Statistics of Input Variables for H-NERM Model Calibration

Variable	Unit	Mean	Median	Min	Max
Temperature, $T(t)$	°C	22.5	22.0	15.0	30.0
Soil Moisture Index, $M(t)$	Normalized [0–1]	0.62	0.63	0.40	0.85
Solar Radiation, $S(t)$	hours/day	6.8	7.0	4.0	9.5
Rainfall, $R(t)$	mm	450	420	200	700

The temperature ranged from 15°C to 30°C with an average of 22.5°C, thus reflecting seasonal variation in the region. Soil moisture, normalized between 0 and 1, showed moderate variability with values spanning 0.4 to 0.85, thus capturing differences across sampling sites and temporal fluctuations. Solar radiation varied between 4 and 9.5 hours/day; this highlighted day-length variability across seasons while rainfall ranged from 200 mm to 700 mm, accounting for both dry and wet periods.

The proposed H-NERM model demonstrated strong potential for practical deployment in resource- constrained and topographically diverse regions like Upper Dir. Its ability to balance yield enhancement with sustainable resource use, even under fluctuating climatic conditions, underscores its utility for precision agriculture in developing regions.

Future improvements may involve incorporating real-time Internet of Things (IoT)-based sensors and mobile interfaces to enhance accessibility and decision-making on the farm.

4.1 Practical Recommendations

To effectively implement the proposed H-NERM model and maximize tomato production in Upper Dir, several practical recommendations are offered. First, installing low-cost soil moisture sensors across the farms could provide continuous and real-time data on soil water content. These sensors are crucial for the fuzzy logic component of the model, to enable precise irrigation scheduling that prevents both overwatering and drought stress, thereby optimizing the efficiency of water use.

Second, it is imperative to train local farmers and agricultural extension workers on the interpretation and utilization of the fuzzy system output. Familiarity with how the fuzzy inference engine integrates environmental variables to generate actionable recommendations will empower farmers to make informed decisions regarding irrigation, fertilization, and pest control, leading to improved crop management.

Third, due to the hilly terrain of Upper Dir, adopting terrace farming practices is highly recommended. Terrace farming helps reduce the adverse effects of slope by minimizing soil erosion, enhancing water retention, and improving the uniformity of environmental conditions across the cultivated area. These improvements align well with the slope variable of the model, so as to ensure that the fuzzy rules produce more accurate and location-specific advisories.

Lastly, incorporating indigenous planting knowledge and traditional agricultural practices into the fuzzy rule base could further enhance the contextual relevance of the model. Farmers in Upper Dir possessing valuable experiential knowledge regarding seasonal patterns, pest behavior, and crop responses that, when formalized into fuzzy rules, could increase the robustness and acceptance of the model. This integration fosters a participatory approach, bridging modern computational techniques with local expertise for sustainable and culturally appropriate agricultural development.

5 Conclusions and Future Work

The H-NERM framework, specifically designed for the unique ecological context of Upper Dir, has proven effective in enhancing tomato production through accurate yield prediction and intelligent decision support. By integrating precise mathematical modeling with the uncertainty-handling capability of fuzzy logic, the model offers a robust approach to managing the complex and often imprecise variables involved in agriculture. This hybrid strategy not only improves predictive performance but also aligns well with the traditional knowledge of local farmers, hence creating a meaningful bridge between empirical agricultural practices and advanced computational techniques. The results indicate that H-NERM could serve as a valuable tool for improving crop management, planning, and productivity in similar ecological regions.

It is important to note certain limitations of the model. Its predictive accuracy depends on the quality and completeness of input data, including soil parameters, weather conditions, and records of pest incidence. Sparse or inconsistent data could affect performance and the model may require retraining for substantially different ecological zones.

Looking ahead, several potential directions could further strengthen and expand the utility of the H-NERM framework. Future research will explore the application of this model to other essential crops such as potatoes, in order to broaden its agricultural benefits across multiple food systems. Integration of satellite-based data sources, such as Normalized Difference Vegetation Index (NDVI) and Land Surface Temperature (LST), is also being considered to enrich the input variables and enhance the spatial and temporal accuracy of the predictions by the model. The advancement aim to make H-NERM a comprehensive, accessible, and data-driven solution for smart agriculture, while acknowledging and addressing its computational and data-related constraints.

Data Availability

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

Conflicts of Interest

The author declares that they have no conflicts of interest.

References

[1] Z. Latifi and H. S. Fami, "An investigation of vegetable production in Iran by using time series models," *Fundam. Appl. Agric.*, vol. 6, no. 4, pp. 367–382, 2021. https://doi.org/10.5455/faa.129347

[2] F. Zhang, F. Liu, X. Ma, G. Guo, B. Liu, T. Cheng, and X. Wang, "Greenhouse gas emissions from vegetables production in China," *J. Clean. Prod.*, vol. 317, p. 128449, 2021. https://doi.org/10.1016/j.jclepro.2021.128449

[3] B. Elhami, M. Ghasemi Nejad Raeini, M. Taki, A. Marzban, and M. Heidarisoltanabadi, "Application of classic and soft computing for modeling yield and environmental final impact in vegetable production (a case study: Transplanting onion in Isfahan province, Iran)," *Environ. Sci. Pollut. Res.*, vol. 29, pp. 35 314–35 337, 2022. https://doi.org/10.1007/s11356-022-18700-6

[4] X. Wang, Z. Dou, X. Shi, C. Zou, D. Liu, Z. Wang, X. Guan, Y. Sun, G. Wu, B. Zhang *et al.*, "Innovative management programme reduces environmental impacts in Chinese vegetable production," *Nat. Food*, vol. 2, pp. 47–53, 2021. https://doi.org/10.1038/s43016-020-00199-0

[5] M. Abedini, "Integrated vegetable production planning under uncertainty," Ph.D. dissertation, University of Tasmania, 2021. https://doi.org/10.25959/23250596.v1

[6] J. Dong, N. Gruda, X. Li, Z. Cai, L. Zhang, and Z. Duan, "Global vegetable supply towards sustainable food production and a healthy diet," *J. Clean. Prod.*, vol. 369, p. 133212, 2022. https://doi.org/10.1016/j.jclepro.2022.133212

[7] S. Anand and M. K. Barua, "Modeling the key factors leading to post-harvest loss and waste of fruits and vegetables in the agri-fresh produce supply chain," *Comput. Electron. Agric.*, vol. 198, p. 106936, 2022. https://doi.org/10.1016/j.compag.2022.106936

[8] D. Batool, M. Shahbaz, H. Shahzad Asif, K. Shaukat, T. M. Alam, I. A. Hameed, Z. Ramzan, A. Waheed, H. Aljuaid, and S. Luo, "A hybrid approach to tea crop yield prediction using simulation models and machine learning," *Plants*, vol. 11, no. 15, p. 1925, 2022. https://doi.org/10.3390/plants11151925

[9] Y. Lin, S. Li, S. Duan, Y. Ye, B. Li, G. Li, D. Lyv, L. Jin, C. Bian, and J. Liu, "Methodological evolution of potato yield prediction: A comprehensive review," *Front. Plant Sci.*, vol. 14, p. 1214006, 2023. https://doi.org/10.3389/fpls.2023.1214006

[10] R. Pitakaso, K. Sethanan, K. H. Tan, and A. Kumar, "A decision support system based on an artificial multiple intelligence system for vegetable crop land allocation problem," *Ann. Oper. Res.*, vol. 342, pp. 621–656, 2024. https://doi.org/10.1007/s10479-023-05398-z

[11] S. S. Solankey and M. Kumari, *Advances in Research on Vegetable Production Under a Changing Climate, Vol. 2*. Springer Nature, 2023.

[12] H. Simsek and M. B. Arslan, "Precision agriculture analysis and comparison of tomato greenhouses based on fuzzy approach," *J. Biosyst. Eng.*, vol. 50, pp. 59–78, 2025. https://doi.org/10.1007/s42853-025-00251-6

[13] M. Cano-Lara and H. Rostro-Gonzalez, "Tomato quality assessment and enhancement through fuzzy logic: A ripe perspective on precision agriculture," *Postharvest Biol. Technol.*, vol. 212, p. 112875, 2024. https://doi.org/10.1016/j.postharvbio.2024.112875

[14] L. Q. Thao, D. T. Kien, N. N. Dinh, N. N. Bao, N. C. Bach, N. D. Thien, D. D. Cuong, D. T. T. Thuy, D. G. Khanh, and P. H. Lam, "Optimizing tomato irrigation through deep learning-enabled wireless sensor networks with fuzzy logic," *Irrig. Sci.*, vol. 42, pp. 955–976, 2024. https://doi.org/10.1007/s00271-024-00949-z

[15] L. R. A. Gabriel Filho, D. D. S. Viais, F. F. Putti, D. Bordin, J. F. Silva, and C. P. Cremasco, "Fuzzy modeling of biometric variables development of tomato crop under irrigation and water salinity effects," *Acta Sci. Agron.*, vol. 46, p. e63515, 2023. https://doi.org/10.4025/actasciagron.v46i1.63515

[16] J. Zhang, L. Xiang, Y. Liu, D. Jing, L. Zhang, Y. Liu, W. Li, X. Wang, T. Li, and J. Li, "Optimizing irrigation schedules of greenhouse tomato based on a comprehensive evaluation model," *Agric. Water Manag.*, vol. 295, p. 108741, 2024. https://doi.org/10.1016/j.agwat.2024.108741

[17] X. Yu, Y. Luo, B. Bai, X. Chen, C. Lu, and X. Peng, "Prediction model of nitrogen, phosphorus, and potassium fertilizer application rate for greenhouse tomatoes under different soil fertility conditions," *Agronomy*, vol. 14, no. 6, p. 1165, 2024. https://doi.org/10.3390/agronomy14061165

[18] S. C. A. Houetohossou, V. R. Houndji, R. Sikirou, and R. Glele Kakai, "Finding optimum climatic parameters for high tomato yield in Benin (West Africa) using frequent pattern growth algorithm," *PLoS One*, vol. 19, no. 2, p. e0297983, 2024. https://doi.org/10.1371/journal.pone.0297983

[19] S. K. Sidhu, "Integrating fuzzy logic into smart agriculture systems for better yield predictions," *J. Punjab Acad. Sci.*, vol. 24, pp. 80–85, 2024. http://jpas.in/index.php/home/article/view/9

[20] A. Barman, V. Shukla, A. Chakraborty, and S. Alam, "Multi-objective optimization based decision-making process and its application to optimally select suitable greenhouse site for tomato crops," *RAIRO-Oper. Res.*, vol. 59, no. 2, pp. 877–906, 2025. https://doi.org/10.1051/ro/2024193