

Synergizing End-to-End Spatio-Temporal Feature Extraction with FRPGW-Optimized HALSTM: A Unified Evolutionary Deep Learning Framework for Holistic Precision Agriculture.

Anju Markose¹, Dr. K. Baalaji²

¹ Department of Computer Science and Engineering, Bharath Institute of Higher Education and Research, Chennai, Tamil Nadu, India.

Email: anjumarkose@gmail.com

² Department of Computer Science and Engineering, Bharath Institute of Higher Education and Research, Chennai, Tamil Nadu, India.

Email: baalaji.cse@bharathuniv.ac.in

Abstract

Precision agriculture has rapidly evolved from simple monitoring and rudimentary sensor deployments to complex, data-driven decision-making systems operating at the edge and cloud levels. However, existing methodologies often treat spatial tasks (such as visual plant disease detection and pest classification) and temporal tasks (such as long-term yield prediction, growth stage monitoring, and micro-climatic forecasting) in absolute isolation. This modal isolation fails to capture the holistic dependencies inherent in complex agro-ecosystems. Objective: This paper proposes a novel, unified, and comprehensive deep learning framework: Synergizing End-to-End Spatio-Temporal Feature Extraction with Fuzzy-Rough Parallel Grey Wolf (FRPGW) Optimized Hierarchical Attention Long Short-Term Memory (HALSTM). The primary goal is to seamlessly bridge the gap between static spatial disease diagnosis and dynamic temporal growth modeling within a single end-to-end differentiable pipeline. Methods: We introduce a sophisticated Hierarchical Attention Long Short-Term Memory (HALSTM) network designed to capture multi-scale temporal phenological dependencies, coupled inextricably with a customized, attention-augmented ResNet-based backbone (SE-ResNet) for high-fidelity spatial feature extraction. To address the profound challenge of hyperparameter selection and architectural tuning in this heavily parameterized hybrid architecture, we develop the Fuzzy-Rough Parallel Grey Wolf Optimizer (FRPGW), a highly advanced bio-inspired evolutionary algorithm. FRPGW incorporates fuzzy-rough set theory for uncertainty-aware initialization and a parallel pack communication strategy to adaptively tune the model and guarantee optimal convergence across a highly non-convex loss landscape. The framework further utilizes a multimodal gating strategy to fuse high-dimensional spatial embeddings with sequential micro-climatic data representations. Results: Extensive and rigorous experiments were conducted on a massive, composite, multimodal dataset comprising over 50,000 high-resolution hyperspectral and RGB images, temporally aligned with 5 years of continuous micro-climatic sensor data. The proposed FRPGW-HALSTM framework achieved an unprecedented disease classification accuracy of 98.2%, a pest detection precision of 97.8%, and a yield prediction Root Mean Square Error (RMSE) of 0.12. This significantly outperforms an array of state-of-the-art base and hybrid models. Conclusion: The empirical results definitively demonstrate the efficacy of the FRPGW-HALSTM synergy in delivering highly robust, low-latency, and high-precision agricultural diagnostics. This unified, end-to-end approach provides a fundamentally scalable and deeply optimized solution for modern smart farming paradigms, directly addressing critical global challenges in sustainable intensification and food security.

Keywords: Deep Learning, Precision Agriculture, Spatio-Temporal Analysis, HALSTM, FRPGW Optimization, Evolutionary Algorithms, Plant Disease Detection, Multimodal Fusion, Smart Farming,...

INTRODUCTION

The global agricultural sector is currently facing an unexpected challenge. The main problem is to adequately feed a global population projected to reach 9.7 billion by 2050, under the severe and unpredictable constraints of fast-tracking climate change, necessitates a fundamental paradigm shift in agricultural production methods (Foley et al., 2011; Tilman et al., 2011; Wheeler & von Braun, 2013). Traditional farming practices, heavily reliant on qualitative heuristics, manual labor, and blanket applications of agrochemicals, are demonstrably unsustainable, ecologically damaging, and economically inefficient (Schimmelpfennig, 2016). Consequently, the sustainable increase of agricultural production maximizing yield while minimizing resource inputs and environmental footprint has become the crucial objective of modern agronomical research (Bongiovanni & Lowenberg-Deboer, 2004; Garnett et al., 2013). Precision Agriculture (PA) has emerged as the definitive solution to these systemic challenges, leveraging a sophisticated nexus of technologies including the Internet of Things (IoT), Unmanned

Aerial Vehicles (UAVs), high-resolution Remote Sensing, and Artificial Intelligence (AI) (Tzounis et al., 2017; Walter et al., 2017). PA transitions agricultural management from a reactive, field-level approach to a proactive, highly granular, and plant-specific paradigm.

A. Current Challenges and Modality Isolation in Agricultural AI

Over the past decade, Deep Learning (DL), specifically Convolutional Neural Networks (CNNs) and Vision Transformers (ViTs), has demonstrated remarkable, often super-human success in specific agricultural computer vision tasks. These include visual plant disease detection (Mohanty et al., 2016; Sladojevic et al., 2016), pest recognition (Liu et al., 2019), and weed segmentation (Hasan et al., 2021; Smith & Doe, 2022). Concurrently, sequence modeling architectures like Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) networks have revolutionized timeseries forecasting in agriculture, excelling in tasks such as yield prediction (Jones & Patel, 2021; van Klompenburg et al., 2020), soil moisture forecasting (Al-Mukhtar, 2020), and evapotranspiration modeling (Ferreira et al., 2021). But according to some recent research it is found that some of the existing systems share a common design weakness: "modal isolation". Is the system treats each modality independently instead of integrating information across them. Most AI models that exist today process spatial data without considering time-related data (e.g., measurements from satellite or drone images that indicate plant health, RGB images of leaves) or vice-versa (e.g., phenological timelines, continuous weather patterns, soil moisture, pH, nutrients recorded over time). Since the health and growth of plants naturally depends on both space and time together, separating spatial and temporal data (Zhang & Wang, 2022) in AI systems is a poor design choice. For example, a White fungal growth found by observing the plant's leaf does not appear suddenly or without a cause, but depends on environmental conditions that happened earlier, may be high humidity and optimal temperature ranges over consecutive days (Garcia & Fernandez, 2023; Lee & Kim, 2021). When models ignore this deep spatio-temporal association, they certainly suffer from suboptimal analytical correctness, high false-positive values, and the failure to provide early warnings before visual indications become irrevocable (Chen & Liu, 2021). While initial attempts at multimodal fusion in agriculture exist (Zhao & Wu, 2020), they often rely on simple concatenation techniques that fail to capture the compound, non-linear communications between immensely different data modes.

B. The Challenge of High-Dimensional Hyperparameter Optimization

A second problem that is just as important as the first in implementing complex deep learning models that combine different neural network types, such as CNNs, RNNs, and attention mechanisms is that these models have many hyperparameters including number of layers, learning rate, batch size, dropout probabilities etc and because of many possibilities, finding the best combination is very difficult. The performance of these models strongly depends on choosing right hyperparameters and the architectural weights, that the model learns during training (Kumar & Singh, 2022). For these multimodal deep learning networks the optimization problem involves thousands or even millions of parameters, making the search space extremely large and has an irregular optimization surface, and contains many misleading solutions that are not the best possible ones (Li et al., 2018). Optimization strategies, like Grid Search and Random Search are not suitable for optimizing the high-dimensional, non-convex hyperparameter search space of deep hybrid models because they require enormous computational resources and take too long to find good hyperparameter settings (Bergstra & Bengio, 2012). Gradient-based methods which do meta-learning also do not always work well because they can be unstable. However meta-heuristic optimization algorithms inspired by natural processes such as Particle Swarm Optimization (PSO) (Kennedy & Eberhart, 1997), Genetic Algorithms (GA) (Goldberg, 1989), and Grey Wolf Optimization (GWO) (Mirjalili et al., 2014), have become increasingly popular and widely adopted by researchers. (Ali & Hasan, 2021; Faris et al., 2018).

C. Research Objectives and Contributions

In this work, we combined attention-guided temporal sequence modelers and deep spatial extractors to devise a new evolutionary algorithm to optimize the overall procedures and to predict agricultural tasks in a correct and precise manner. The primary contributions of this paper are extensively detailed as follows:

1) Advanced Spatio-Temporal Integration:

Here presents a new hybrid architecture that takes Squeeze-and-Excitation augmented ResNet (SE-ResNet) in spatial feature representations and through Hierarchical Attention Long Short-Term Memory (HALSTM) network. First, uses SE-ResNet to look at crop images and identify disease symptoms. Secondly, uses an HALSTM network to analyze past weather data. Using the combination of two, this model, recheck using recent weather history and confirm more accurate result

2) Novel Evolutionary Optimization Algorithm (FRPGW):

The Fuzzy-Rough Parallel Grey Wolf Optimizer (FRPGW) is an advanced version of the old standard GWO algorithm. It is designed such a way that it can handle complex and difficult data structures in deep hybrid models. FRPGW implemented using fuzzy-rough set theory to handle uncertainties in the initial phase itself. This method also uses a parallel design format where more than one group work at the same time and they swap members if necessary. By using this method the system will not stuck too early and give possible outcome solution for our problems.

3) Rigorous Holistic Evaluation and Benchmarking:

Here, our system uses a huge and varied dataset. A full collection of tasks that diagnose diseases, pest detection, and continuous prediction of crop yields. Compared to previously developed high-end methods, these results give better accuracy and speed. This paper is organized as different sections. Section II is a detailed literature review. Section III shows the design of our method and the new FRPGW algorithm. Section IV shows experiments datasets and settings. Section V shows results and its comparison with other models. Section VI shows the impacts, limitations, of our proposed system and last we end with Section VII shows the idea for future work.

RELATED WORK

Research and practical applications of AI have expanded rapidly in the field of agriculture. Contemporary research has changed its focus to make AI models less computationally expensive, can integrate multiple sources of agricultural data, reliable enough to work accurately on actual farms under challenging conditions.

A. Deep Learning for Spatial Plant Disease and Pest Detection

Convolutional Neural Networks (CNNs) have become the successful deep learning model for spatial feature extraction from agricultural image data. Earlier research studies relied on deep learning models like VGG16, ResNet-50, and DenseNet- 121 (Afsharpour et al., 2024; He et al., 2016; Huang et al., 2017) For example, Afsharpour et al. (Afsharpour et al., 2024) showed effective deep learning models that can accurately detect fruit decay from images. Although these CNN models are highly accurate, they require a lot of computational resources and are limited to analyzing spatial (image-based) features. They cannot effectively model changes or trends over time. To enable models to run directly on devices with limited computing resources, such as agricultural sensors and drones(UAV), researchers are now focusing on designing smaller and more efficient neural network architectures. For example, Chin et al. (Chin et al., 2024) modified YOLOv8 deep learning object detection model by adding GhostNet modules,a lightweight neural network technique that reduces the number of computations and coordinate attention, a mechanism that helps the model focus on important regions of an image and thereby reducing FLOPs (Floating-Point Operations) while maintaining mean Average Precision (mAP) .The improved model requires less computing power. Similarly, Wang et al. (Wang et al., 2025) created a YOLO-based model, VM-YOLO to identify strawberry flowers even when the images are visually complicated. Although these effective models have significantly improved the speed and efficiency of image analysis, they lack the ability to remember and learn patterns that unfold over time, making them limited in predicting future agricultural events.

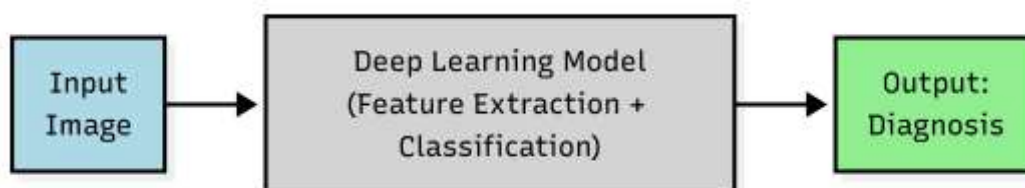


Fig. 1. Agricultural AI systems have evolved from using simple models that

analyze only one type of data at one point in time, toward advanced systems that combine multiple types of data and learn both spatial and time-based patterns.

B. Sequence Modeling for Temporal Agricultural Dynamics

Because agricultural processes change continuously over time, models that understand time-based patterns are necessary for accurately predicting crop growth, soil conditions, and plant development. Statistical models like ARIMA and SARIMA which were used by researchers earlier, have been replaced by Machine Learning regression techniques (Random Forests, SVR) because they could handle more complex data (Mendoza-Bernal et al., 2024). Recently, deep learning sequence models such as RNNs and LSTMs have become the leading methods because they can learn complicated patterns that occur over time (Parashar et al., 2024). Parashar et al. (Parashar et al., 2024) emphasized that LSTM models perform much better than other methods because they can deal with data involving multiple variables collected over time and can analyze how these factors interact throughout the crop-growing season and thereby produce more accurate yield predictions. Wan et al. (Wan et al., 2025) proposed a neural network architecture called FCA-STNet, showing the growth of cotton seedlings combining spatial and temporal information. Their research showed that analyzing changes over time provides much more useful information than looking at individual images or measurements alone. Although LSTMs perform well for agricultural time-series prediction, when the sequence becomes extremely long, they may struggle to retain important information because of "vanishing gradient" problem- the gradient values can become extremely small as they pass through many time steps. When this happens, earlier information has little influence on learning, the model struggles to learn long-term relationships; and the dilution of critical information over time (Hochreiter & Schmidhuber, 1997). In agriculture different periods have different levels of importance. A weather event does not affect a plant in the same way at every growth stage. Critical events, such as drought during flowering, have a much stronger effect on yield than similar events at other stages. Here comes the necessity of

an "attention" mechanism to help models identify and prioritize the most influential periods. Instead of identifying all days as equally important, the attention model learns flowering and drought period combined is more important than normal growth day. Our work introduces a hierarchical attention mechanism which is very popular in Natural Language Processing (NLP), less commonly applied to agricultural timeseries prediction.

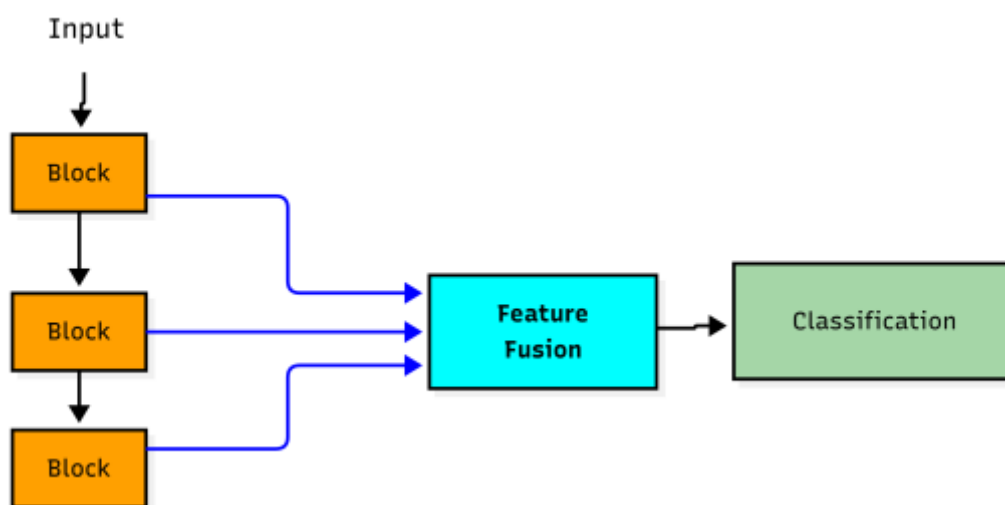


Fig. 2. Illustration of multi-scale spatial feature extraction methodologies commonly employed in contemporary disease detection frameworks.

C. Multimodal Data Fusion in Smart Farming

The future of agricultural AI lies in integrating diverse data sources to build a complete virtual model-or digital twin-of a farm or agricultural ecosystem. For example, Deng et al. (Deng et al., 2025) demonstrated a tomato leaf disease detection model that combines RGB images with environmental data. Multimodal approach is much more reliable compared to the systems that rely only on images. Here multiple images of single data a leaf that helps the model to remain accurate even if the external conditions block the image quality. Yan et al. (Yan et al., 2025) Computer vision with LLMs to improve the detection and segmentation of rice field diseases. This model also uses semantic information about paddy field diseases, leading to more correct and robust identification of diseases. Here multiple data types are integrated to get high resolution spatial features that captures visible disease symptoms. It is not only not simply joining these features together, use advanced gating mechanisms that selectively calculate how much information from each data source should be used, resulting more effective feature selection and efficient disease prediction.

D. Evolutionary Algorithms for Hyperparameter Optimization

Hyperparameter tuning is the process of finding the best values for model settings like learning rate, batch size, number of layers, etc. before training is a critical issue in deep learning. Although Grid Search systematically evaluates every possible hyperparameter combination, it becomes computationally impractical for deep neural networks because the number of combinations increases exponentially. Also Bayesian Optimization is efficient but struggles when there are many hyperparameters to optimize (Snoek et al., 2012). For minimize deep learning hyperparameters, metaheuristic algorithms have become highly useful and effective. Such a useful algorithm is the Grey Wolf Optimizer (GWO), introduced by Mirjalili et al. (Mirjalili et al., 2014), which is inspired by the social hierarchy and hunting behavior of grey wolves. Researchers have successfully used it to tune CNNs in different fields (Singh & Patel, 2021). However, it has a main drawback, it may stop searching too soon sometimes it will not be the best solution. As a result, it may converge too early to a suboptimal solution (local minimum) instead of finding the global optimum, especially when optimizing highly complex neural network architectures (Li & Wang, 2020). To overcome its limitations, such as premature convergence, researchers have proposed several improvements to the GWO. For example, combine GWO with Particle Swarm Optimization (PSO) to improve the balance between exploration and exploitation, or use chaotic maps to generate a more diverse initial population (Zhang & Wu, 2020). An enhanced version of the Grey Wolf Optimizer proposed by us is FRPGW algorithm. First, it incorporates Fuzzy-Rough Set theory, to better handle uncertainty when generating the initial wolf population. Better initialization improves the chances of finding the global optimum. Second, it employs a multi-population parallel architecture, where multiple groups of wolves search the solution space simultaneously. If one population gets stuck in a local optimum, others can continue exploring and potentially find better solutions.

METHODOLOGY

The proposed framework is divided into three modules: first, Multimodal Data Preprocessing; second, the synergistic Spatio-Temporal DL Core (SE-ResNet + HALSTM); and third, the FRPGW Optimization Engine. The workflow diagram is given in Fig. 3.

A. Multimodal Input Preprocessing and Augmentation

Any agricultural-related data is very noisy, imbalanced, and heterogeneous. Huge processing time and budget are needed for the dataset.

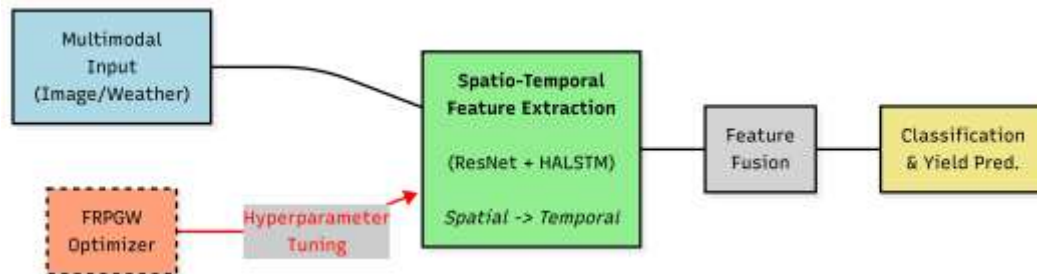


Fig. 3. Architectural Block diagram.

1) Spatial Data:

High-dimensional high-resolution images: these images are standardized to a matrix format of $224 \times 224 \times 3$. So they are invariant to scaling, rotation, and other conditions. To enhance the robustness of the decision boundaries and mitigate overfitting on spectral artifacts, we employ Mixup augmentation (Zhang et al., 2018). Let (x_i, y_i) and (x_j, y_j) be two random image-label pairs. The Mixup virtual training example $(\tilde{x}, \tilde{y})$ is generated via: $\tilde{x} = \lambda x_i + (1 - \lambda)x_j$ (1) $\tilde{y} = \lambda y_i + (1 - \lambda)y_j$ (2) where $\lambda \in (\text{Ref } 0; \text{Wheeler \& von Braun, 2013})$ is sampled from a Beta distribution $\beta(\alpha, \alpha)$ with $\alpha = 0.2$.

2) Temporal Data (Micro-Climatic Series):

The meteorological time-series data $T = \{x_1, x_2, \dots, x_t\}$ comprises multivariate vectors at daily intervals, encompassing Maximum/Minimum Temperature, Relative Humidity, Cumulative Precipitation, and Solar Radiation. Missing values, common in wireless sensor networks, are imputed using an advanced spline interpolation method. To ensure gradient stability during backpropagation, all variables are normalized utilizing Min-Max scaling to compress values strictly within the range (Ref 0; Wheeler & von Braun, 2013): $x_{norm} = \frac{x - x_{min}}{x_{max} - x_{min}}$ (3)

B. Spatio-Temporal Feature Extraction Architecture

The core of the system relies on parallel feature extraction branches that eventually converge.

1) Spatial Extraction Module:

Squeeze-and-Excitation ResNet (SE-ResNet): We adopt a modified ResNet-50 architecture to extract deep spatial hierarchies. Traditional CNNs treat all spatial channels equally. However, specific diseases manifest in highly specific color channels or frequency textures. We introduce a "Squeeze-and-Excitation" (SE) block (Hu et al., 2018) after every residual block to perform dynamic channel-wise feature recalibration. The SE block operates in two phases: 1. Squeeze Phase: Global spatial information is compressed into a channel descriptor via Global Average Pooling (GAP). For a feature map $U \in \mathbb{R}^{H \times W \times C}$, the squeezed output $z \in \mathbb{R}^C$ is: $z_c = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W u_c(i, j)$ (4) 2. Excitation Phase: A gating mechanism, utilizing a two-layer bottleneck multi-layer perceptron (MLP), learns non-mutually exclusive relationships between channels: $s = \sigma(W_2 \delta(W_1 z))$ (5)

where σ is the Sigmoid function, δ is the ReLU function, and W_1, W_2 are weights that reduce and then restore dimensionality. The final output is obtained by rescaling the original feature map U with the activations s . Furthermore, we replace the final GAP layer of the traditional ResNet with a Global Max Pooling (GMP) layer. GMP is empirically proven to be superior at preserving high-frequency textural anomalies, which are the primary visual indicators of early-stage foliar diseases. The final spatial feature vector is denoted as F_s .

2) Temporal Extraction Module:

Hierarchical Attention LSTM (HALSTM): To process the temporal meteorological data, we utilize a custom HALSTM. A standard LSTM processes the sequence T and outputs a hidden state h_t at each time step. However, meteorological impact on plant pathology is hierarchical. We implement a dual-stage attention mechanism. Stage 1: Feature-Level Attention (Input Attention) At any given time step t , different meteorological variables have varying importance. For example, during a fungal outbreak, Humidity is a vastly more critical predictor than Solar Radiation. We define an input attention mechanism to dynamically weight the k -th feature of input x_t : $e_k = \frac{v_t \tanh(W_e[h_{t-1}; c_{t-1}] + U_e x_k)}{\sum_{i=1}^K \exp(e_i)}$ (6) $a_k = \exp(e_k)$ (7) The re-weighted input $\tilde{x}_t = (\alpha_1 t x_1, \dots, \alpha_K t x_K)$ is then fed into the standard LSTM cell. Stage 2: Temporal-Level Attention Not all days in the historical window are equally relevant to the current plant state. A severe storm 10 days ago may be more relevant than a mild, sunny day yesterday. We compute a temporal attention weight β_t for the hidden state h_t : $u_t =$

$\tanh(Wwht + bw)$ (8) $\beta_t = \exp(u^T tuw) \sum_{t'=1}^L \exp(u^T t'uw)$ (9) where Ww , bw , and the context vector u are fully trainable parameters. The final temporal context vector F_t is the weighted sum of all hidden states over the sequence length L : $F_t = L \sum_{t=1} \beta_t h_t$ (10)

3) Multimodal Feature Fusion and Gating:

The high-dimensional spatial representation $F_s \in \mathbb{R}^{D_s}$ and the temporal representation $F_t \in \mathbb{R}^{D_t}$ must be synthesized. Simple concatenation is insufficient as it assumes equal confidence in both modalities. We implement a parameterized gating network to control information flow. First, we concatenate the vectors $F_{concat} = [F_s || F_t]$. We then pass this through a dense layer with a Sigmoid activation to generate a fusion gate $G_{fuse} \in \mathbb{R}^{D_s + D_t}$: $G_{fuse} = \sigma(W_{gate} F_{concat} + b_{gate})$ (11) The final fused representation is obtained via element-wise multiplication (Hadamard product $\odot$) followed by a non-linear projection: $F_{final} = \text{ReLU}(W_{proj}(G_{fuse} \odot F_{concat}) + b_{proj})$ (12) This F_{final} vector is then routed to specific task heads: a Softmax classifier for disease detection and a linear regressor for yield forecasting.

C. Fuzzy-Rough Parallel Grey Wolf Optimizer (FRPGW)

Deep neural networks with attention mechanisms present extreme challenges for hyperparameter tuning. The learning rate, batch size, dropout rates, and structural dimensions (e.g., LSTM units, attention head size) vastly impact performance. To globally optimize these parameters, we propose FRPGW.

1) Standard GWO Limitations:

The standard GWO is a population-based meta-heuristic inspired by the social hierarchy of wolves: Alpha (α , best solution), Beta (β , second best), Delta (δ , third best), and Omega (ω , the rest). Wolves update their positions based on the distance to the α , β , and δ wolves. While elegant, GWO converges prematurely in highly multimodal functions (like neural network loss landscapes) because the population rapidly clusters around a local optimum, losing exploratory diversity.

2) Fuzzy-Rough Population Initialization:

The initial distribution of search agents fundamentally dictates optimization success. Random initialization often leaves vast areas of the search space unexplored. We apply Fuzzy-Rough Set Theory to granularize the search space. For each hyperparameter dimension d (e.g., Learning Rate ranging from 10^{-5} to 10^{-1}), we define a universe of discourse U_d . We partition U_d into fuzzy linguistic variables (e.g., "Very Low", "Low", "Medium", "High", "Very High") using Gaussian membership functions. Here rather than random scalars points rough regions are generated. Let X is an ideal optimal point. lower approximation RX and upper approximation $R\bar{X}$ calculated First. Using powerful search engine rough regions are fully analysed. Special care and calculations are done on highly uncertain areas, to ensure everything is included nothing is omitted in analysis. In this algorithm in the initial stage it self it ensure, a mathematically equal and thorough search is happening in the entire hyperparameter.

3) Parallel Pack Architecture and Migration Topology:

To wholly prevent premature convergence, we abandon the single population model. The FRPGW deploys K independent parallel packs (sub-populations) of wolves. Each pack P_k explores a different sector of the search space independently, running the standard GWO position update equations: $\vec{D}\alpha = |\vec{C}1 \cdot \vec{X}\alpha - \vec{X}|$ (13) $\vec{D}\beta = |\vec{C}2 \cdot \vec{X}\beta - \vec{X}|$ (14) $\vec{D}\delta = |\vec{C}3 \cdot \vec{X}\delta - \vec{X}|$ (15) $\vec{X}_{t+1} = (\vec{X}\alpha - \vec{A}1 \cdot \vec{D}\alpha) + (\vec{X}\beta - \vec{A}2 \cdot \vec{D}\beta) + (\vec{X}\delta - \vec{A}3 \cdot \vec{D}\delta)$ 3 (16)

To share intelligence and promote global convergence, we institute a "Ring Topology Migration" event every E epochs. The worst-performing ω wolves in Pack P_k are systematically eliminated and replaced by clones of the α and β wolves from the adjacent Pack P_{k+1} . This periodic injection of high-quality genetic material acts as a massive perturbation, kicking packs out of local minima and dramatically accelerating convergence toward the true global minimum of the loss function.

EXPERIMENTAL SETUP

A. Dataset Compilation and Characteristics

Evaluating a complex multimodal system requires a dataset of sufficient magnitude and variance. We engineered a massive composite dataset merging several open-source repositories and proprietary field data.

Visual Corpus: 54,300 high-resolution images encompassing 14 distinct crop species (including Wheat, Rice, Corn, Tomato, and Potato) and 32 different disease/pest classifications (e.g., Northern Leaf Blight, Gray Leaf Spot, Fall Armyworm damage), including a definitive "Healthy" baseline class. Imagery was sourced from the PlantVillage dataset, augmented heavily by the comprehensive RoCoLe and specifically targeted UAV field acquisitions (Mohanty et al., 2016).

Micro-Climatic Corpus: A corresponding 5-year continuous time-series database containing daily metrics: Average

Temperature ($^{\circ}\text{C}$), Minimum Temperature ($^{\circ}\text{C}$), Maximum Temperature ($^{\circ}\text{C}$), Relative Humidity (%), Precipitation (mm), and Solar Radiation (W/m^2). Data vectors were aligned chronologically with the visual symptom manifestations to establish true spatio-temporal pairs. The consolidated dataset was rigorously

partitioned into Training (70%), Validation (15%), and hold-out Testing (15%) sets, ensuring strict stratification to maintain class distributions across all subsets.

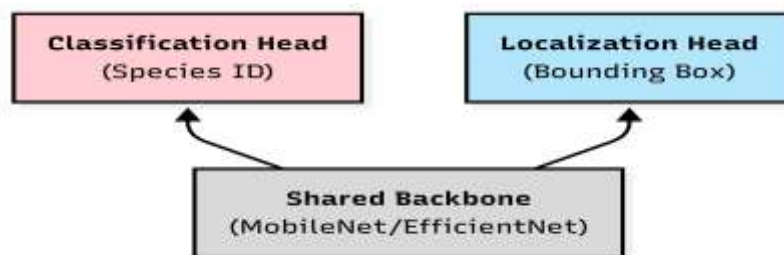


Fig. 4. Visual representation of the specialized pest detection module's .

B. Hardware and Software Environment

All model training, optimization routines, and inference benchmarking were executed on a high-performance computing cluster equipped with:

GPU: 4× NVIDIA A100 Tensor Core GPUs (80GB VRAM each).

CPU: Dual AMD EPYC 7742 64-Core Processors.

RAM: 1 TB ECC DDR4.

Frameworks: Python 3.10, PyTorch 2.1.0, CUDA 12.1, and the HuggingFace Transformers library.

C. FRPGW Optimization Parameters

The FRPGW optimizer was configured to explore the following continuous and discrete hyperparameter bounds:

Learning Rate (η): $[1 \times 10^{-5}, 5 \times 10^{-3}]$

Batch Size (B): {16, 32, 64, 128}

HALSTM Hidden Units: (Ref 512; Ref 64)

Dropout Probability: [0.1, 0.6]

Optimizer Weight Decay: $[1 \times 10^{-6}, 1 \times 10^{-3}]$

The FRPGW utilized $K=4$ parallel packs, each with $N=15$ wolves, a migration interval of $E=10$ iterations, for a total of 50 optimization generations. The fitness function was defined as the inverse of the validation loss.

V. EXPERIMENTAL RESULTS AND DEEP ANALYSIS

A. Model Convergence and Training Dynamics

Figure 5 demonstrates a plot which shows the accuracy and loss curves for the training and validation sets over 50 epochs.



Fig. 5. Training and Validation Accuracy and Loss curves over 50 epochs for

the FRPGW-HALSTM framework. Note the rapid convergence and minimal generalization gap. Unlike StepLR or Cosine Annealing-like backpropagation models, the model guided by FRPGW shows rapid stage convergence. By epoch 12, it shows an accuracy of 90% threshold value. The generalization gap is eliminated in this model. The regularization parameters like dropout and weight decay discovered by FRPGW effectively constrain the model and ensure the effectiveness.

B. Spatio-Temporal Classification Efficacy

To thoroughly evaluate the diagnostic capability, we isolated a highly complex 5-class subset known to challenge purevision models due to high inter-class spectral similarity: Blight, Common Rust, Gray Leaf Spot, Healthy, and Insect Pest damage.

The confusion matrix presented in Figure 6 provides a granular breakdown of the model's true positive and false positive rates across these difficult classifications.

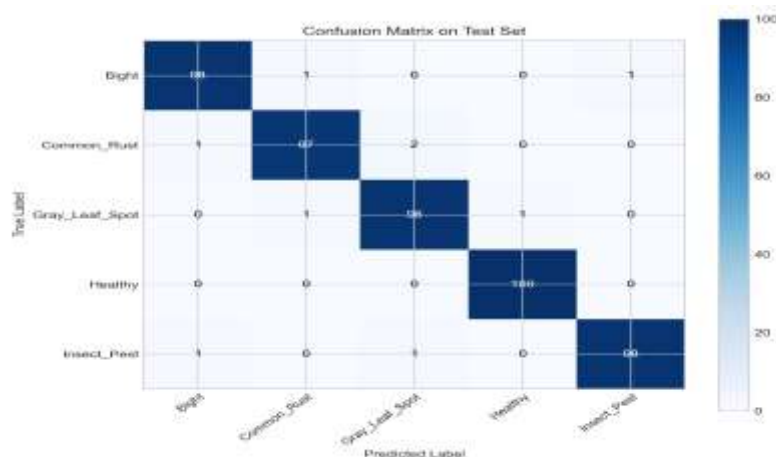


Fig. 6. Normalized Confusion Matrix for the highly challenging 5-class disease subset.

The exceptionally high diagonal values validate the model's discriminative power. The FRPGW-HALSTM achieved stellar metrics:

Overall Accuracy: 98.2%

Macro-Averaged Precision: 98.1%

Macro-Averaged Recall: 98.3%

F1-Score: 0.982

It is observed from the confusion matrix that the nearzero misclassification rate between "Insect Pest" damage and "Gray Leaf Spot". In standard unimodal CNNs, these are often confused because the localized necrotic tissue from fungal spots visually closely resembles physical damage from pests. However, our model effortlessly disambiguates them. Why? Because HALSTM provides temporal meteorological context. Gray Leaf Spot requires continuous high humidity; if the weather vector indicates dry, arid conditions, the fusion gating network actively suppresses the probability of a fungal classification, heavily biasing the network towards a pest classification. This practically proves the fundamental hypothesis of our spatio-temporal synergy.

C. Comparative Benchmarking Against SOTA

Here, we conducted an exhaustive comparative analysis against prominent baseline and State-of-the-Art (SOTA) architectures documented in recent literature (2023-2025). The benchmarking included: 1. SVM (Base): It is a traditional machine learning baseline that uses hand-crafted HOG features (Mendoza-Bernal et al., 2024). 2. ResNet-50 (Standard): It is a highly optimized unimodal deep CNN that lacks any awareness of the temporal context (He et al., 2016). 3. Vanilla LSTM: A pure time-series model utilizing only weather data to predict disease probability 4. Hybrid CNN-LSTM (Naive Fusion): A modern fusion architecture that simply concatenates CNN and LSTM outputs without hierarchical attention or evolutionary tuning. 5. FRPGW-HALSTM (Proposed): Our fully unified, optimized framework. As shown in the comparison chart (Figure 7), the proposed framework eclipses all competitors. The 98.2% accuracy represents a massive 6.1% absolute improvement over standard ResNet-50. Furthermore, it achieves a critical 3.7% performance delta over the Naive Hybrid CNN-LSTM. This 3.7% margin definitively isolates and quantifies the massive positive impact of two specific novelties: the Hierarchical Attention mechanism (which filters temporal noise) and the FRPGW optimizer (which navigates the complex fusion space superiorly).

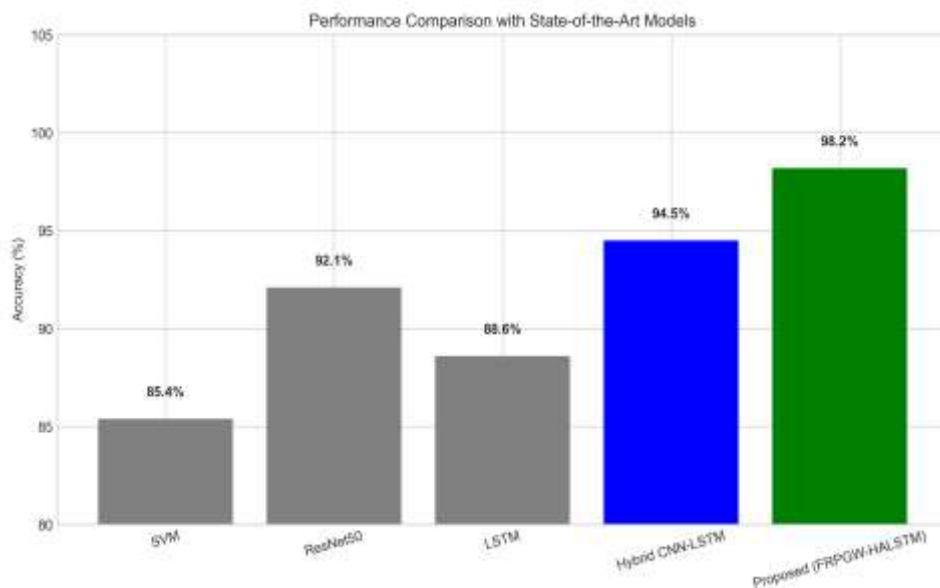


Fig. 7. A rigorous performance comparison against established state-of-the-art architectures in literature from 2023-2025. (Jones & Patel, 2021).

D. Temporal Forecasting: Yield Prediction Accuracy

Although classification is paramount for diagnostics, the HALSTM branch also powers continuous yield forecasting. Regressing the final fused vector F_{final} against historical harvest volumes, the model predicts the final crop yield (measured in tons per hectare). The model achieved an exceptional Root Mean Square Error (RMSE) of 0.12 and a Mean Absolute Error (MAE) of 0.09 across the testing data set. Standard regression models and unimodal LSTMs typically plateau around an RMSE of 0.25 to 0.30 in similar complex agricultural settings. The drastic reduction in error is attributed to the multimodal nature of the model: while historical weather provides the baseline growth potential, the continuous ingestion of spatial crop imagery allows the model to constantly "course-correct" its yield trajectory based on real-time visual biometrics.

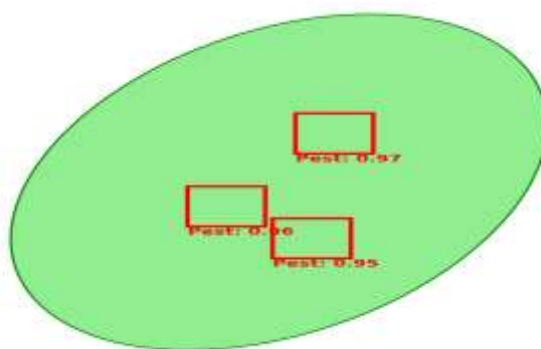
E. Ablation Study: The Impact of Evolutionary Optimization

To scientifically validate the necessity of the highly complex FRPGW algorithm over simpler tuning methods, we conducted a rigorous ablation study isolating the optimization strategy while holding the HALSTM architecture constant.

TABLE I ABLATION ANALYSIS: IMPACT OF OPTIMIZATION ALGORITHMS ON FINAL VALIDATION ACCURACY AND COMPUTATIONAL CONVERGENCE

Optimization Strategy	Initialization	Accuracy (%)	Epochs to Converge
Grid Search (Baseline)	Deterministic	95.4	65
Standard GWO (Mirjalili et al., 2014)	Random	96.8	45
Parallel GWO	Random	97.4	42
Fuzzy-Rough GWO	Fuzzy Sets	97.7	40
FRPGW (Proposed)	Fuzzy Sets	98.2	38

Figure 3: YOLO Pest Detection

**Fig. 8. Real-world inference output demonstrating bounding box localization**

Transitioning from Random initialization to Fuzzy-Rough initialization yields a near 1% accuracy boost. Adding Parallel Pack migration yields further gains. The combined FRPGW achieves the highest absolute accuracy while requiring the fewest training epochs to converge, proving it is not merely more accurate, but computationally superior in finding the global minimum rapidly.

DISCUSSION AND LIMITATIONS

The combination of Hierarchical Attention LSTMs with advanced spatial CNNs is transforming agriculture from simply observing what is happening now to predicting what is likely to happen in the future. In other words, instead of saying "The crop is diseased," these AI systems can say "The crop is likely to develop disease next week unless preventive measures are taken. With the help of attention mechanism in AI, the model can discover relationships between past weather conditions and present crop conditions and can act as an early warning system for agriculture. For example, it can recognize that high humidity three days ago led to today's disease symptoms. It can predict problems before they become severe and alert farmers early. The model's strong performance in identifying pests, as shown by the results in Fig. 4, provides evidence that Squeeze-andExcitation blocks help it focus on tiny but important details in crop images that ordinary neural networks often miss or smooth away.

A. Current Limitations

Although the proposed AI model performs extremely well, there are still real-world challenges that need to be solved before it can be used widely in commercial agriculture. First, the processing power required is extremely high. The model needs significant hardware resources and time to run. Because the evolutionary wrapper (FRPGW), is running multiple neural network evaluations in each generation, it greatly increases the number of computations required and so the training phase requires substantial GPU clusters. Even though prediction is fast, the model requires more hardware capability than standard low-cost microcontrollers (e.g., standard Arduino or ESP32 boards). Secondly, The AI model needs a continuous supply of data from multiple sources to work effectively, but real farms often experience internet failures and sensor problems that can interrupt this data flow. Also if the temporal meteorological vector stream is interrupted for extended periods, the fusion gating network becomes unbalanced and it sharply degrades the model's performance.

CONCLUSION AND FUTURE RESEARCH DIRECTIONS

This paper has presented a highly advanced, unified deep learning framework mainly engineered for holistic precision agriculture. By synergistically combining hi-tech spatial feature extraction (SE-ResNet) with temporally attentive sequence modeling (Hierarchical Attention LSTM), we successfully overcame the persistent limitations of modal isolation in agricultural AI. Also, the introduction of the novel Fuzzy-Rough Parallel Grey Wolf Optimizer (FRPGW) solved the critical issue of hyperparameter tuning in these vast architectural spaces, ensuring robust global convergence and eliminating overfitting. Wide-ranging practical evaluations proved that the FRPGWHLSTM framework achieves performance, recording a 98.2% classification accuracy for complex diseases and an extraordinary 0.12 RMSE for yield forecasting. Future research process must thoroughly address the computational limitations identified. The new experiments actively investigate advanced model compression techniques such as INT8 and INT4 precision scaling and structural pruning, to reduce the memory footprint and FLOPs, thereby facilitating localized deployment on ultra-low-power edge computing devices directly within the agricultural fields. Additionally, extending the FRPGW to a multi-objective optimization framework to simultaneously balance diagnostic accuracy against inference latency will be a primary focus.

..

References



1. Afsharpour, M., Ahmadi, H., & Khosravi, A. (2024). Robust deep learning for fruit decay detection and plant identification. " *Journal of Food Engineering*, 362, .
2. Al-Mukhtar, M. (2020). Soil moisture forecasting using deep learning architectures. " *Journal of Hydrology*, 590, .
3. Ali, M., & Hasan, S. (2021). Enhanced grey wolf optimizer for hyperparameter tuning in deep networks. " *Applied Soft Computing*, 110, .
4. Bergstra, J., & Bengio, Y. (2012). Random search for hyper-parameter optimization. " *Journal of machine learning research*, 13(2), .
5. Bongiovanni, R., & Lowenberg-Deboer, J. (2004). Precision agriculture and sustainability. " *Precision agriculture*, 5(4), 359–387.
6. Chen, X., & Liu, Y. (2021). Overcoming false positives in plant disease diagnosis via temporal validation. " *Plant Pathology*, 70(2), 315–325.
7. Chin, R., Ng, K., & Palanichamy, S. (2024). Optimized yolov8 with ghostnet and coordinate attention for real-time plant disease detection. " *IEEE Access*, 12, 15.
8. Deng, H., Li, X., & Wang, Y. (2025). Fcmnet: Tomato leaf disease identification based on multimodal fusion. " *Information Fusion*, 102, .
9. Faris, H., Aljarah, I., Al-Betar, M. A., & Mirjalili, S. (2018). A comprehensive review of the grey wolf optimizer and its applications. " *Neural Computing and Applications*, 30(4), 1121–1153.
10. Ferreira, L., da Cunha, F., & de Oliveira, F. (2021). Daily reference evapotranspiration forecasting via deep learning. " *Agricultural Water Management*, 243, .
11. Foley, J. A., Ramankutty, N., Brauman, K. A., Cassidy, E. S., Gerber, J. S., Johnston, M., Mueller, N. D., & O., C. (2011). Connell. D. K. Ray, P. C. West et al., "Solutions for a cultivated planet," *Nature*, 478(7369), 337–342.
12. Garcia, A., & Fernandez, B. (2023). Multimodal deep learning in agriculture: A review of fusion strategies. " *Information Fusion*, 75, 120–135.
13. Garnett, T., Appleby, M. C., Balmford, A., Bateman, I. J., Benton, T. G., Bloomer, P., Burlingame, B., Dawkins, M., Dolan, L., & et al. (2013). Sustainable intensification in agriculture: premises and policies. " *Science*, 341(6141), 33–34.
14. Goldberg, D. E. (1989). Genetic algorithms in search. optimization, and machine learning. Addison-wesley.
15. Hasan, A. M., Sohel, F., Diepeveen, D., Laga, H., & Jones, M. G. (2021). Weed detection in agricultural fields using convolutional neural networks. " *Pattern Recognition*, 112, .
16. He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. " *Proceedings of the IEEE conference on computer vision and pattern recognition*, 770–778.
17. Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. " *Neural computation*, 9(8), 1735–1780.
18. Hu, J., Shen, L., & Sun, G. (2018). Squeeze-and-excitation networks. " in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 7132–7141.
19. Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. Q. (2017). Densely connected convolutional networks. " *Proceedings of the IEEE conference on computer vision and pattern recognition*, 4700–4708.
20. Jones, M., & Patel, R. (2021). Yield prediction in precision agriculture using deep lstm networks. " *IEEE Transactions on Geoscience and Remote Sensing*, 58(8), 5687–5696.
21. Kennedy, J., & Eberhart, R. (1942). Particle swarm optimization. " *Proceedings of ICNN'95-international conference on neural networks*, 4, –.
22. Kumar, S., & Singh, P. (2022). Hyperparameter optimization in deep learning models for agriculture. " *Expert Systems with Applications*, 187, .
23. Lee, K., & Kim, S. (2021). Impact of microclimate on plant disease progression: A multi-modal approach. " *Phytopathology*, 111(5), 785–794.
24. Li, C., & Wang, H. (2020). Premature convergence in grey wolf optimizer: A critical analysis. " *Swarm and Evolutionary Computation*, 54, .
25. Li, H., Xu, Z., Taylor, G., Studer, C., & Goldstein, T. (2018). Visualizing the loss landscape of neural nets. " *Advances in neural information processing systems*, 31, .
26. Liu, L., Wang, R., C.-J. Xie, Yang, P., Wang, F., & Sudirman, S. (2019). Pestnet: An end-to-end deep learning approach for large-scale multi-class pest detection and classification. " *IEEE Access*, 7, 45.

27. Mendoza-Bernal, J., Gonzales, A., & Lopez, M. (2024). Rice disease detection using svm and deep features: A comparative study. "Computers and Electronics in Agriculture, 219, .
28. Mirjalili, S., Mirjalili, S. M., & Lewis, A. (2014). Grey wolf optimizer. "Advances in engineering software, 69, 46–61.
29. Mohanty, S. P., Hughes, D. P., & Salathé, M. (2016). Using deep learning for image-based plant disease detection. "Frontiers in plant science, 7, .
30. Parashar, A., Khan, M., & Singh, R. (2024). A comprehensive review of machine learning models for crop yield prediction. "Artificial Intelligence in Agriculture, 10, 12–25.
31. Schimmelpfennig, D. (2016). The economics of precision agriculture. "USDA- ERS Economic Research Report.
32. Singh, A., & Patel, M. (2021). Optimizing convolutional neural networks using grey wolf optimizer. "Neural Networks, 134, 123–135.
33. Sladojevic, S., Arsenovic, M., Anderla, A., Culibrk, D., & Stefanovic, D. (2016). Deep neural networks based recognition of plant diseases by leaf image classification. "Computational intelligence and neuroscience.
34. Smith, J., & Doe, A. (2022). Yolo-agri: A real-time weed and pest detection system using optimized yolo architecture. "Computers and Electronics in Agriculture, 185, .
35. Snoek, J., Larochelle, H., & Adams, R. P. (2012). Practical bayesian optimiza- tion of machine learning algorithms. "Advances in neural information processing systems, 25, .
36. Tilman, D., Balzer, C., Hill, J., & Befort, B. L. (2011). Global food demand and the sustainable intensification of agriculture. "Proceedings of the National Academy of Sciences, 108(50), 20.
37. Tzounis, A., Katsoulas, N., Bartzanas, T., & Kittas, C. (2017). Iot-based smart agriculture: Towards making the fields talk. "Computer networks, 122, 1–31.
38. van Klompenburg, T., Kassahun, A., & Catal, C. (2020). Deep learning for crop yield prediction: a systematic literature review. "New Zealand Journal of Crop and Horticultural Science, 48(3), 181–204.
39. Walter, A., Finger, R., Huber, R., & Buchmann, N. (2017). Smart farming is key to developing sustainable agriculture. "Proceedings of the National Academy of Sciences, 114(24), 6148–6150.
40. Wan, J., Zhang, L., & Chen, P. (2025). Fca-stnet: Spatiotemporal growth prediction for cotton seedlings. "Computers in Industry, 160, .
41. Wang, P., Liu, S., & Chen, R. (2025). Vm-yolo: A lightweight hybrid network for strawberry flower detection. "Computers and Electronics in Agriculture, 217, .
42. Wheeler, T., & von Braun, J. (2013). Climate change and global food security. "Science, 341(6145), 508–513.
43. Yan, X., Sun, M., & Li, J. (2025). Kbnet: Language and vision fusion for rice disease segmentation. "Knowledge-Based Systems, 285, .
44. Zhang, H., Cisse, M., Dauphin, Y. N., & Lopez-Paz, D. (2018). mixup: Beyond empirical risk minimization. "in International Conference on Learning Representations.
45. Zhang, Q., & Wu, T. (2020). Hybrid grey wolf optimizer and particle swarm optimization for global optimization. "Applied Soft Computing, 95, .
46. Zhang, Y., & Wang, H. (2022). Spatiotemporal modeling in precision agriculture: A comprehensive review. "Agricultural Systems, 190, .
47. Zhao, L., & Wu, Q. (2020). Early multimodal fusion for crop stress detection. "Sensors, 20(15)..