Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Edge-Enabled Deep Learning and Multi-Model Hybrid Intelligent Architectures Integrating Sentinel-3 and MODIS Data for Real-Time Monitoring of Crop Stress and Nonlinear Pest Complexes Forecasting Under Climate Shifts

MRK Pathan

mrkpathan@cu.ac.bd

University of Chittagong https://orcid.org/0009-0002-1943-7942

Mst. Rasheda Chowdhury

University of Chittagong

Research Article

Keywords: Agro-Ecological Modeling, Hybrid Fuzzy Probabilistic Neural Network, Spatial-Temporal Dynamics, Geodetic Coordinate Referencing, Pigeon pea Pest Complex, Universal Kriging Interpolation, Multi-Source Environmental Covariates, Edge Computing Optimization

Posted Date: September 8th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10938048/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: The authors declare no competing interests.

Edge-Enabled Deep Learning and Multi-Model Hybrid Intelligent Architectures Integrating Sentinel-3 and MODIS Data for Real-Time Monitoring of Crop Stress and Nonlinear Pest Complexes Forecasting Under Climate Shifts

Authors: M. R. K. Pathan[a*], Dr. Mst. Rasheda Chowdhury[a]

[a*] **M. R. K. Pathan (Corresponding Author)**, Department of Zoology, University of Chittagong, Chittagong-4331, Bangladesh. mrkpathan@cu.ac.bd, +8801752618227.

[a] **Dr. Mst. Rasheda Chowdhury,** Department of Zoology, University of Chittagong, Chittagong-4331, Bangladesh. mrashedachowdhuryzoology@cu.ac.bd.

Edge-Enabled Deep Learning and Multi-Model Hybrid Intelligent Architectures Integrating Sentinel-3 and MODIS Data for Real-Time Monitoring of Crop Stress and Nonlinear Pest Complexes Forecasting Under Climate Shifts

Abstract

A twelve-year (2011–2022) multi-site agro-ecological study on the population dynamics of 5 key lepidopteran and hemipteran pests of pigeonpea (*Cajanus cajan*) in a sub-continental tropical environment was carried out to understand vulnerability patterns of this important pulse ecosystem to multiple biotic stresses. Twelve hundred and fifty hectares under diverse agro-ecological zones were surveyed using sub-meter accuracy real-time-kinematic global navigation satellite system coordinates and UTM projections. Using MODIS, Sentinel-3, and WorldClim datasets, drift-adjusted geospatial rasters were generated to represent the pest's microclimatic and topographic preferences while overcoming local-scale heterogeneity. Analysis over standard meteorological weeks SMW 1–52 revealed that regional pest outbreaks followed closely the host plant phenology, while peaks in economic damage were strongly correlated to post-monsoon relative humidity gradients. Spatial autocorrelation estimates using Global Moran's I: 0.65–0.82, $p < 0.001$, and surface interpolation using Universal Kriging corroborated a range of spatial dependence between 9.5 and 15.2 km and an associated cross-validation root mean square error of 0.05–0.14. With respect to benchmarking against state-of-the-art machine learning and deep learning algorithms, the proposed hybrid Fuzzy Probabilistic Neural Network demonstrated outstanding approximation capability with probabilistic layers based on Gaussian mixture density functions h = 0.5, m = 2.0, and achieving coefficients of determination between 0.961–0.991 as compared to Vision Transformers: 0.948–0.979, XGBoost: 0.889–0.935 and native SARIMAX: 0.810–0.872. With respect to the predictive performance, Monte Carlo dropout simulations between $\mp$ 0.028 and $\mp$ 0.045 and conformal prediction intervals with over 95% empirical coverage established the algorithm's predictive reliability under stress conditions. Finally, with the deployment of the hybrid network's ONNX runtime quantized variant achieving a 4.4× acceleration in inference, the methodology provides critical decision support to promote sustainable smallholder agrarian livelihoods and food security while enabling adaptive capacity to local-level biophysical and climatic heterogeneity.

Keywords: Agro-Ecological Modeling, Hybrid Fuzzy Probabilistic Neural Network, Spatial-Temporal Dynamics, Geodetic Coordinate Referencing, Pigeon pea Pest Complex, Universal Kriging Interpolation, Multi-Source Environmental Covariates, Edge Computing Optimization

Highlights

- Conducted a comprehensive agro-ecological investigation from 2011 to 2022 across a 1,250ha monitoring footprint in multi-site research stations.

- Quantified spatial-temporal population dynamics and larval feeding vigor for five major pigeonpea insect pests of *Helicoverpa armigera*, *Melanagromyza obtusa*, *Exelastis atomosa*, *Maruca vitrata*, and *Clavigralla gibbosa*.
- Integrated sub-meter real-time kinematic GNSS coordinates, Universal Transverse Mercator projections, and multi-source environmental covariates from MODIS, Sentinel-3, and WorldClim datasets to eliminate positional drift.
- Demonstrated through longitudinal sampling across standard meteorological weeks SMW 1–52 that peak pest outbreaks were tightly synchronized with host plant reproductive phases and post-monsoon relative humidity gradients.
- Verified robust spatial dependence and clustering across monitoring domains, with Global Moran's I values ranging from 0.65 to 0.82 at $p < 0.001$ and spatial dependence ranges spanning between 9.5 km and 15.2 km.
- Mapped continuous spatial probability surfaces incorporating environmental drift parameters, achieving cross-validation root mean square error values restricted between 0.05 and 0.14.
- Proposed and validated the Hybrid Fuzzy Probabilistic Neural Network integrating Gaussian membership functions of $h = 0.5$, $m = 2.0$ with parallel probabilistic pattern layers achieving exceptional coefficients of determination (R^2) from 0.961 to 0.991.
- Outperformed alternative modeling architectures, including Vision Transformers 0.948-0.979, XGBoost 0.889–0.935, and baseline SARIMAX architectures 0.810–0.872.
- Confirmed model stability and reliability through Monte Carlo dropout simulations $\pm$ 0.028 to $\pm$ 0.045 and conformal prediction intervals exceeding 95% empirical coverage.
- Achieved a 4.4× inference speedup via ONNX Runtime quantization and containerized FastAPI microservices, successfully translating high-performance geospatial analytics into edge-computed decision support systems for precision agricultural management.

Introduction

Macro-Agricultural Significance and Global Pulse Production Vulnerabilities

Leguminous cropping systems are the cornerstone of agricultural production both at regional and global levels. Pigeonpea (*Cajanus cajan* (L.) Millsp.), in particular, is grown by millions of small-scale farmers as an invaluable source of dietary protein, essential amino acids, and nitrogen-fixing power for the semi-arid and sub-humid subtropical regions (Wang et al., 2025). Nevertheless, pulse legumes across the Indian subcontinent, sub-Saharan Africa, and Latin America are increasingly vulnerable to the adverse effects of climate change and unpredictable weather patterns (Santos et al., 2026). Rising global temperatures and shifting monsoon behaviors contribute to large-scale agricultural disruptions with severe implications for entire regional economies.

Entomological Pressures and Phenological Synchronization Dynamics

In addition to the already challenging environmental conditions, tropical legumes are subjected to the persistent threat of insect pests that inflict serious damage to the plants throughout their development cycle. South Asia, in particular, is home to several devastating lepidopteran and hemipteran species that target pigeonpea crops and cause major yield losses (Srinivasa Rao et al., 2016). Most notably, the *Helicoverpa armigera*, *Melanagromyza obtusa*, *Exelastis atomosa*, *Maruca vitrata*, and *Clavigralla gibbosa* are widely distributed and capable of destroying both seeds and pods. Moreover, the abundance of these pests is directly determined by large-scale climate patterns and regional weather conditions, which serve as crucial drivers of insect appearance and development (Ibrahim et al., 2022). Namely, relative air humidity, temperature, vapor pressure deficit, and rainfall have a significant impact on the number of insect generations and their feeding activity. The infestation peaks tend to appear precisely during the most critical periods of host plant development, ranging from the vegetative phase and branching to flowering and pod setting.

Limitations of Conventional Surveillance and Reactive Pest Management

Agricultural practices and policies addressing these destructive pests have been traditionally limited to calendar- or observational-based chemical interventions. Forecasting methodologies employed against them consist of relatively simple linear regression models or conventional time-series analyses. These approaches are typically region-specific and do not account for the spatial correlation between neighboring fields (Wang et al., 2025). Also, they are mostly based on independent consideration of environmental variables rather than integrated remote sensing data (Wang et al., 2025). More importantly, such methodologies are applied on a relatively large scale, which makes them irrelevant to local farmers' needs and practices. Therefore, the use of broad-spectrum insecticides becomes a logical consequence of the observed management trends, which have negative implications for the local environment and even regional water resources.

Technological Evolution toward Predictive Geospatial Intelligence and AI Integration

Modern agricultural research is progressively shifting toward geospatial intelligence, remote sensing data, and artificial intelligence to improve the accuracy and precision of crop protection measures (Wang et al., 2025). Using geographic coordinates and extensive weather data, local entomology experts can now integrate these inputs into powerful predictive models that highlight the likelihood of a pest infestation. The advancement of deep learning and related concepts has also enhanced the possibility of adopting complex multi-layered neural networks and machine learning algorithms (Sharma et al., 2025). However, there is clearly a gap between high-level computational processes occurring in distributed servers and the actual end-users' needs, who require relatively simple yet accurate predictive tools. Regional entomology students and farmers alike would benefit from the combination of geospatial interpolation, deep learning algorithms, and efficient decision-making aids.

Research Gaps

Agricultural entomology scholars have made considerable progress in applying remote sensing data for pest forecasting, yet few have successfully addressed the issue of local spatial coordinates. Modern researchers utilize extensive collections of MODIS, Sentinel-3 data, and WorldClim climate information to predict the timing of regional infestation outbreaks. However, these valuable sources are rarely integrated with local ground truth data in an accurate manner. Most notably, conventional remote sensing methods fail to incorporate raster plots with vector location findings consistently enough (Wang et al., 2025). Several predictive models disregard the phenomenon of spatial non-stationarity and topographic drift altogether, thus eliminating the possibility of adopting appropriate regionalization techniques (Wang et al., 2025). Essentially, the failure to consider spatial auto-correlation and local neighborhood effects undermines the accuracy of regional pest forecasting while creating a conceptual divide between global and local methodologies.

The general applicability of conventional machine learning algorithms and statistical models is significantly limited when it comes to analyzing field-level processes. The dynamic nature of agricultural data and the variety of influencing factors make it challenging to implement traditional time-series methods, such as SARIMAX, and fit them to noisy input data (Khan et al., 2024). Similarly, standard deep learning practices fail to achieve satisfactory results due to their complexity and inability to process the fuzzy logic of the agricultural domain (Shabbir et al., 2025). Hybrid neuro-fuzzy approaches have been proposed to address this issue while resolving the multi-scale nature of spatial data, but very few studies have actually succeeded (Ishaque, 2023). Finally, the absence of a comprehensive deep learning framework capable of capturing complex non-linear interactions and providing consistent out-of-sample performance is also notable.

The general usefulness of the aforementioned algorithms is then limited by the heavy requirements typical of the modern distributed server systems. Agricultural scientists and regional policymakers currently utilize a variety of complex statistical and machine learning methods to analyze seasonal pests. Nevertheless, these approaches are mostly confined to large-scale institutional settings due to their computational demands (Sharma et al., 2025). Farmers are not likely to benefit from such tools, which requires additional research to be done. The development of an efficient containerized ONNX system capable of reducing the complexity of high-level deep learning algorithms could help regional agricultural officers apply predictive analytics more efficiently in their day-to-day operations.

Research Questions

To systematically address research this study is asks how do the constraints of sub-meter real-time kinematic GNSS positioning, UTM projections, and MODIS, Sentinel-3, and WorldClim

environmental covariates contribute to the establishment of a robust high-resolution raster-vector overlay without displacements for multi-site pigeonpea agro-ecosystems? It investigates how do the spatial-temporal population fluctuations and inter-annual generational turnovers of the five major pigeonpea insect pest complexes respond to the host plant reproductive cycles and post-monsoon relative humidity gradients in a 12-year period (2011–2022)? This work examines what are the comparative predictive capabilities of a Hybrid Fuzzy Probabilistic Neural Network algorithm with a pair of Gaussian membership functions h = 0.5, m = 2? and parallel probabilistic pattern layers in predicting multi-site pest outbreaks compared to conventional machine learning and deep learning models? It explores how do the spatial autocorrelation measures based on Global Moran's Moran's I statistic and Universal Kriging surface relate to the spatial spans and prediction variances of pest infestation patterns in distinct agro-climatic experimental sites? Finally, the research evaluates how can the containerized FastAPI microservices and Open Neural Network Exchange Runtime accelerate the model inference procedures for better operationalization by agricultural extension personnel?

Research Objectives

The objectives which were set up correspond to research questions that were established. The first objective was to design and apply a geodetic integration framework that employs sub-meter real-time kinematic GNSS coordinates and multi-source remote sensing raster data within the 1,250ha monitoring zone. Second, the objective was to estimate 12-year population densities and phenology synchronization windows of the pigeon pea pest complex under standard meteorological weeks (SMW 1-52) for crop stages and micro-meteorological gradients. The third objective was to formulate a Hybrid fuzzy-PNN framework and evaluate it against the conventional vision transformers, XGBoost, and SARIMAX algorithms using stratified cross-validation and uncertainty parameters. Fourth, the objective was to perform geostatistical spatial analysis including semivariogram range and Universal Kriging procedures to produce continuous spatial probability surfaces and local infestation hotspots. Finally, the objective was to develop and evaluate an edge-deployed ONNX runtime quantization and microservice containers for real-time farmer-end advisory systems.

Novelty and Contributions

This study presents the Hybrid Fuzzy Probabilistic Neural Network or Hybrid Fuzzy-PNN that incorporates biological information entropy into high-performance neural network paradigm. Unlike existing agricultural forecasting models that rely on either black-box deep learning algorithms or linear time-series assumptions that fail to account for agro-climatic fluctuations, Hybrid Fuzzy-PNN uniquely combines h = 0.5-and m = 2.0-based Gaussian membership functions with parallel pattern processing layers that allow simultaneous fuzzification of continuous agrometeorological inputs while maintaining probability traceability. As a result, the new neurofuzzy architecture demonstrated unprecedented flexibility and accuracy in capturing

year-to-year biological relationships, with coefficient of determination (R^2) varying between 0.961 and 0.991 across 12 years of validation. Accordingly, the technical novelty of this study is manifested in developing an innovative algorithm that uniquely combines fuzzification and high-performance neural network processing to enable accurate long-term agricultural forecasting.

Methodologically, the current approach is distinguished by the elaborated geospatial harmonization framework which synergistically combines sub-meter real-time kinematic GNSS coordinates, UTM projections, and satellite imagery from MODIS, Sentinel-3, and WorldClim platforms to process the earth observation data from a 1,250 ha multi-site test domain. By utilizing the Global Moran's Index statistics of 0.65-0.82 and the Universal Kriging surface model, the study defines the 9.5-15.2 km spatial autocorrelation distance to bridge the resolution of remote sensing inventories with local entomological surveys. Overall, the novel spatial analysis methodology allowed bridging the critical divide between global earth observation data and local entomological surveys, creating a unified framework that can be used to analyze similar agro-ecological regions worldwide**.**

In terms of technical implementation, the study employs a model serving workflow that connects the server-side analytics to the edge computing platform via a FastAPI container and open-source ONNX runtime quantization toolchain to accelerate the batch inference by 4.4x compared to the conventional approaches while maintaining the prediction accuracy. The toolchain includes the Monte Carlo dropout simulations and strict Gauss-Markov error modeling for the reliable uncertainty estimation required for the on-farm decision support system. As a result, it becomes possible for the first time to offer turnkey machine learning solutions to national agricultural extension services and small-scale regional cooperatives that seek to adopt precision agriculture technologies without committing extensive resources to computational infrastructure.

The temporal extent of the study covers the 12-year period of standard meteorological weeks or SMW 1-52 to track the developmental patterns, larval feeding behaviors, and economic thresholds of five key lepidopteran and hemipteran herbivores *Helicoverpa armigera*, *Melanagromyza obtusa*, *Exelastis atomosa*, *Maruca vitrata*, *Clavigralla gibbosa* and create dynamic scouting calendars for the relevant pest complex. The empirical patterns identified are used to develop scouting calendars that describe general seasonal shifts in insect developmental stages as well as the associated economic thresholds that should prompt treatment decisions. The novel scouting calendar can be used to update national agricultural guidelines and regulations across tropical legume-producing regions, as well as inform the development of similar spatio-temporal frameworks for other agricultural commodities of regional importance.

Finally, this research follows the open science principles by making the custom source codes, trained weights, geostatistical raster layers, and API endpoints available in the public repositories to facilitate the global replication and independent validation by the external stakeholders. This commitment to open-source availability ensures that international research consortia, regional cooperatives, and academic peers can independently audit, validate, and adapt the predictive pipeline, thereby accelerating global innovation in digital phytopathology and smart agricultural systems.

Related Work

Foundational Paradigms & Early Statistical Forecasting in Pest Dynamics

Initial explorations in agricultural entomology and pest dynamics forecasting emphasized empirical observation modeling, linear regression analysis, and classical auto-regressive time-series implementations (Patel et al., 2011; Kumar et al., 2014). Such paradigms accounted for spatially homogeneous and temporally stationary fields, assessing insect developmental generations through rudimentary degree-day models and calendar-based phenology (Srinivasa Rao et al., 2016; Dhaliwal et al., 2015). While these foundational associations confirmed rudimentary temperature-dependence relationships, more elaborate multi-week phytosanitary time lags and monsoon-driven perturbations significantly challenged such simplistic implementations (Giles et al., 2018). Moreover, early approaches completely disregarded spatial non-stationarity and contagion effects, limiting their applicability to small-scale agro-ecological settings (Sharma & Singh, 2019).

Integration of Remote Sensing & Environmental Variables

Later entomological studies began addressing these limitations by emphasizing broader earth observation and geographical statistics, employing both local weather station data and remote sensing platforms (Thenkabail et al., 2018; Justice et al., 2002). Such works leveraged MODIS, Sentinel-3, and WorldClim indices to track spatio-temporal patterns of surface reflectance, temperature, and hydrological regimes on a regional scale (Wang et al., 2025; Fick & Hijmans, 2017). One notable limitation of these early contributions was the inability to properly reconcile continuous raster covariates with discrete vector monitoring points while accounting for positional discrepancies (Gao et al., 2020). Existing geostatistical methods often failed to incorporate local topographical variations, undermining the creation of reliable raster-vector overlays (Lu & Wong, 2008).

Emergence of ML & Advanced Neural Architectures

Modern agricultural pest surveillance research has seen a rapid rise of ensemble machine learning and deep neural network approaches. Gradient-boosted decision trees, LSTMs, and Vision Transformers have been utilized to track complex non-linear patterns of regional infestation (Shabbir et al., 2025; Sharma et al., 2025; Li et al., 2023). These recent advances have demonstrated clear benefits over classic linear approaches. Nevertheless, contemporary literature is replete with reports of significant shortcomings related to both classical and modern methods. First, standard deep learning classifiers are known to severely overfit train-validation-test splits, especially when dealing with biological fuzziness (Goodfellow et al., 2016). Second, most works failed to combine separate fuzzy logic layers with parallel pattern recognition architectures (Ishaque, 2023; Zadeh, 1965).

Geospatial Universal Kriging & Spatial Interactions

A related line of scientific inquiry focused on universal kriging and spatial autocorrelation analyses as means to assess general pest distribution trends (Cressie, 1993; Anselin, 1995). Multiple prior works confirmed interactions between large-scale bioclimatic patterns and localized pest developmental stages (Liebhold et al., 2004). Notably, spatial clustering of insect infestations was rarely random, often reflecting underlying micro-climate and crop-specific patterns. Contemporary spatial analyses, however, have failed to consider geostatistical surfaces and machine learning as interrelated processes. By treating universal kriging and deep learning as independent steps, researchers often neglected opportunities to jointly optimize spatial covariate weights and gradient-based feature importance during multi-site agricultural pest trials (Haining, 2003). This added computational complexity decreased the accuracy of both spatial prediction surfaces and machine learning classification (Goovaerts, 1997).

Transition Towards Quantifying Uncertainties & Probabilities

As many of these approaches were transitioned to practical decision-support systems, the need for enhanced inferential capabilities has seen a dramatic rise (Gaddum, 1948; O'Hagan, 2012). Recent studies report that simple point estimates often fail to account for the degree of spatial randomness embedded in pest dynamics (Ibrahim et al., 2022; Smith et al., 2021). Most early efforts to adjust for such uncertainties involved simple bootstrapping or confidence interval estimation techniques. In practice, however, few prior studies effectively incorporated spatial non-stationarity and heteroscedasticity properties into formal uncertainty estimation frameworks (Efron & Tibshirani, 1993). Contemporary research emphasizes the need to account for extreme climate events, advocating for more elaborate conformal prediction and Monte Carlo model calibration methods (Vovk et al., 2005; Metropolis & Ulam, 1949).

Computational Constraints & Edge Advisory Systems

An emerging issue related to modern deep learning approaches is the disconnect between sophisticated recommendation algorithms and farmers' practical needs (Busch & O'Neill, 2019; Wolfert et al., 2017). Deep learning models often rely on substantial computational resources, requiring recurrent cloud connectivity to perform even basic inference (Chen et al., 2020). Such requirements fail to meet the needs of actual users who require timely, if not real-time, on-field recommendations to control large-scale infestations. While techniques such as model quantization and ONNX runtimes have been successfully applied in computer vision tasks, their application to local agro-ecological forecasting remains limited (Howard et al., 2017; Bai et al., 2019). Most contemporary works provide little in the way of light-weight microservices, with few solutions offering automated advisory agricultural services at the sharp edge (Rodriguez et al., 2021).

Limitations of Current Frameworks for Multi-Site Prediction

A comprehensive review of recent literature suggests that most computational entomology research emphasizes local-scale pest monitoring and forecasting. Very few works provide multi-

site recommendations based on extended (multi-seasonal) monitoring periods. Studies targeting legume-specific pests such as *Helicoverpa armigera* often fail to account for broader agro-ecological complexes, including *Melanagromyza obtusa*, *Exelastis atomosa*, *Maruca vitrata*, and *Clavigralla gibbosa* (Sharma et al., 2005; Fitt, 1989). Additionally, most current algorithms disregard twelve-year (2011-2022) benchmarking and modern GNSS-based real-time position tracking, creating a conceptual disconnect between long-term biological rhythms and actual implementations (Santos et al., 2026; Parmesan & Yohe, 2003).

Present Research Grounding

With these limitations in mind, the current study aims to bridge several gaps in computational entomology and regional agricultural forecasting. Unlike prior literature, which primarily focuses on either linear statistical modeling or high-level deep learning, this work introduces a novel hybrid paradigm. Hybrid Fuzzy-PNN forecasts are combined with sub-meter GNSS position estimates, geospatial kriging surfaces, and ONNX runtime optimization to provide automated precision farming solutions. By emphasizing a holistic view of the problem, this study attempts to reconcile the need for fine-scale spatial resolution, biological non-linearity, and actual computational feasibility.

Methodology

Study Area Selection, Geo-Referencing, and Spatial Sampling Framework

The field work and spatial survey were conducted in multi-location pigeonpea (*Cajanus cajan*) experimental farms in India, the research stations of which are located in Parbhani and Dhule, Maharashtra of 19.2608°N, 76.7735°E and 20.9042°N, 74.7749°E, Pusa, Bihar of 25.9868°N, 85.6744°E, and Jhansi, Uttar Pradesh of 25.4484°N, 78.5685°E. In a stratified multiple grid system, there was selected 1250 ha farmland for investigation based on a comprehensive field survey in 2011-2022, which spans the survey duration of 12 years accounting for the period from 2006-17. Each trial station was found to be located in a different climatic region from one another, ranging from semi-arid and sub-humid tropical climate with distinct moisture regimes and canopy density characteristic of the Indian agro-climate. Coordinates for the sampling sites were obtained with Trimble GNSS receivers to ensure the horizontal accuracy, in real-time kinematic (RTK) positioning technology, of sub-meter precision, identifying local topography and proximity to the contour marks for each field block during field survey.

In line with the stratified multiple grid system was implemented a representative sampling to mark plot inspection in a field survey based on absolute and relative insect sampling schemes peculiar to MW 1-MW 52 (meteorological week 1-week 52) and crop season. Specifically, for every study site, inspection quadrats (1mx1m) were marked using beat-sheet and manual counting to score the number of insects recorded per grid location. The structure of the plot

inspection, which incorporates absolute and relative sampling, is attributed to the general spatial variability typical of local micro-topography and canopy density following the standard stratified grid multi-stage design. Plots were marked in situ with Trimble RTK GNSS receivers to mark the location on a sub-meter level in line with the horizontal accuracy of RTK technology, which is further used to demarcate the grid system in the study sites according to local topography.

The underlying spatial architecture was integrated with multi-source environmental covariates extracted from MODIS, Sentinel-3, and WorldClim gridded databases at a standardized 30m spatial resolution, with positional data layers projected onto the WGS 84 / UTM coordinate reference system corresponding to each respective zone to maintain spatial integrity across raster-vector overlays. Sample preservation and sorting protocols adhered to strict entomological standards, mapping exact spatial point geometries directly into GIS geodatabases to facilitate subsequent spatial interpolation, autocorrelation modeling, and geostatistical surface generation without relying on unsupported synthetic imputations. This rigorous geographic foundation ensured that subsequent machine learning inputs and geostatistical kriging algorithms operated on high-fidelity, empirically validated spatial coordinates across the Indian subcontinent. Google Gemini artificial intelligence model (Google LLC) employed to assist in drafting and optimizing the custom Python scripts for data analysis, with the researcher maintaining full responsibility for reviewing, testing, and validating all code to ensure analytical rigor and reproducibility.

Table 1 Spatial and predictive modeling performance metrics across five key insect pest complex variables evaluated over 1,250 georeferenced sample rows at a 30-meter dynamic resolution

Pest Complex Variable	Universal Kriging R2	Ordinary Kriging R2	EBK CV RMSE	Spatial Autocorrelation (Moran's I)	Semivariogram Range (km)	Spatial Lag Coefficient	Mapping Confidence %	Dynamic Resolution (m)	VIF Max Collinearity	Spatial Block CV RMSE	Sobol Total Index	Mean Absolute SHAP	Calibrated HDI Width	Sample Georeferenced Rows
Harmigera Larvae Per m	0.975	0.968	0.12	0.74	12.4	0.81	97.5	30	2.85	0.45	0.84	1.42	0.59	1250
Pod Fly Damage Pct	0.962	0.954	0.10	0.68	14.1	0.75	96.2	30	3.12	0.41	0.79	1.18	4.73	1250
Plume Moth Larvae Per Plant	0.982	0.975	0.08	0.79	11.2	0.78	98.1	30	2.95	0.48	0.86	1.55	0.71	1250
Spotted Pod Borer Clusters Per Plant	0.970	0.963	0.11	0.72	13.0	0.80	97.0	30	2.64	0.33	0.91	1.84	0.43	1250
Pod Bug Insects Per Plant	0.965	0.958	0.09	0.70	13.5	0.72	96.5	30	2.45	0.30	0.76	0.95	0.32	1250

Target Pest Complex Identification and Empirical Field Data Acquisition

The field survey was conducted to identify the target pests infesting pigeonpea in various agro-ecological regions of India. The five pests included gram pod borer (*Helicoverpa armigera* Hubner), pod fly (*Melanagromyza obtusa* Malloch), plume moth (*Exelastis atomosa* Walsingham), spotted pod borer (*Maruca vitrata* Fabricius) and pod bug (*Clavigralla gibbosa* Spinola). The data Pertaining to the population trends of the target pests, number of larvae per meter row, number of infested clusters per plant, and percentage of pod damage were collected from the five agro-climatic regions of India over a period of 12 years. The field survey focused

on crop development from the vegetative stage to the branching, flowering to pod development and maturity. There was a simultaneous recording of the number of larvae per trap, number of damaged pods per cluster, and infested clusters per plant.

The data collected were subjected to quality control measures to remove extreme values and bring them to conformity with ICAR field survey procedures. Similarly, the field survey data on the target pest complex were standardized for the number of larvae per meter, number of clusters per plant, and percentage of pod damage for the purpose of developing the predictive model for the study. It was found that the outbreak of the target pests was closely linked to the prevailing weather conditions in the four locations. The field survey data collected over a period of 12 years also formed the ground truth data set for developing predictive models for the target pest complex.

The field survey focused on the phenology of the target pests to establish their occurrence pattern in relation to the growth stages of pigeonpea and the response of various pest species to the variations in local weather conditions. Likewise, the cross-verification of the field survey data with the published reports on agricultural pest outbreaks in the target locations also contributed to improving the quality of the ground truth data set by eliminating inconsistencies. The process of ground truthing therefore played a critical role in ensuring that the data set used conforms to the actual situation in the four study sites.

Table 2 Comparative performance, uncertainty quantification, and environmental shock stability metrics for the Hybrid Fuzzy-PNN model and baseline architectures across five insect pest complex variables

Pest Complex Variable	Proposed Hybrid Fuzzy-PNN R2	Vision Transformer R2	XGBoost Baseline R2	Random Forest R2	SARIMAX Temporal R2	Conformal Interval Coverage %	Monte Carlo Dropout Bounds	Sobol Total Index	First Order Sobol Index	Mean Absolute SHAP	Integrated Gradients	Heatwave Shock Stability %	Rainfall Shock Stability %	Drought Shock Stability %	Temp Sensitivity Rank
Harmigera Larvae Per m	0.978	0.962	0.912	0.895	0.845	95.4	∓ 0.041	0.84	0.65	1.42	1.25	94.5	95.2	93.1	2
Pod Fly Damage Pct	0.965	0.951	0.895	0.878	0.821	95.1	∓0.038	0.79	0.61	1.18	1.05	92.1	93.4	91.0	3
Plume Moth Larvae Per Plant	0.985	0.974	0.928	0.911	0.862	96.2	∓ 0.032	0.88	0.69	1.55	1.38	96.0	96.8	94.5	1
Spotted Pod Borer Clusters Per Plant	0.972	0.960	0.918	0.901	0.850	95.8	∓ 0.035	0.82	0.63	1.31	1.15	93.8	94.5	92.2	4
Pod Bug Insects Per Plant	0.968	0.955	0.902	0.885	0.835	95.3	∓ 0.037	0.80	0.62	1.25	1.10	92.8	93.9	91.5	5

Bioclimatic Covariate Integration and Environmental Feature Engineering

Environmental feature engineering integrated dynamic meteorological variables extracted from ground-based automatic weather stations at Parbhani, Pusa, Dhule, and Jhansi, complemented by high-resolution gridded climate archives where evaluated bioclimatic predictors included daily maximum and minimum air temperatures (T_{max} and T_{min} in °C), relative humidity percentage, vapor pressure deficit (VPD in kPa), cumulative precipitation (Precip. in mm), and Growing Degree Days (GDD). Thermal activation thresholds and physiological development rates were explicitly computed, establishing baseline thermal constants ranging from 25.0°C to 29.0°C that dictate generational turnover and larval feeding vigor. Cumulative thermal units were modeled using standard agrometeorological accumulation equations:

$$\text{GDD} = \sum \left(\frac{T_{\max} + T_{\min}}{2} - T_{\text{base}} \right)$$

To establish the statistical foundation of these environmental drivers, the analytical methodology incorporated an extensive suite of frequentist and classical statistical models, including Ordinary Least Squares (OLS) regression for baseline linear relationship modeling between environmental covariates and insect pest densities:

$$y_i = \beta_0 + \sum_{j=1}^{p} \beta_j \, x_{ij} + \varepsilon_i$$

Generalized Linear Models (GLM - Poisson / Negative Binomial) were applied for handling overdispersed count data inherent in insect population dynamics across standard meteorological weeks, modeling expected counts μ_i via a log-link function:

$$\ln(\mu_i) = \mathbf{x}_i^T \boldsymbol{\beta}$$

Autoregressive Integrated Moving Average with Exogenous Inputs (SARIMAX) was implemented for temporal forecasting of seasonal lags and meteorological drivers:

$$\phi_p(\mathbf{B})\widetilde{\Phi}_P(\mathbf{B}^s)(1-B)^d(1-B^s)^D y_t = c + \sum_{i=1}^{k} \theta_i^T \, X_{it} + \theta_q(\mathbf{B})\widetilde{\Theta}_Q(\mathbf{B}^s)\varepsilon_t$$

Geographically Weighted Regression (GWR) was deployed to account for spatial non-stationarity across the multi-site Indian regional gradient, formulated as:

$y_i = \beta_0(u_i, v_i) + \sum_k \beta_k \, (u_i, v_i) x_{ik} + \varepsilon_i$ where (u_i, v_i) denote the geographical coordinates of monitoring site i. Spatial lag coefficients and neighborhood contagion metrics were engineered using queen-contiguity spatial weights matrices W derived from the geo-referenced point distributions across the four experimental stations, quantifying local neighborhood spillover effects and confirming robust spatial autocorrelation indices via Global Moran's I:

$I = \frac{n}{S_0} \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij}(x_i - \bar{x})(x_j - \bar{x})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$ where $S_0 = \sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij}$.

Feature selection protocols utilized recursive feature elimination (RFE) and multicollinearity diagnostics with variance inflation factor values under five to isolate the most informative bioclimatic and topographic predictors while eliminating redundant noise. Engineered feature matrices were structured into multi-dimensional tensor arrays combining temporal lag sequences spanning 1-week to 4-week historical windows across standard meteorological weeks with spatial coordinate descriptors.

Table 3 Thermal activation thresholds, physiological development rates, spatial autoregressive coefficients, economic thresholds, generalized additive model diagnostics, and farmer advisory action statuses across five insect pest complex variables within the Semi-Arid Tropics ecological province

Pest Complex Variable	Thermal Activation Threshold (C)	Physiological Dev Rate / Day	Spatial Autoregressive Coefficient	Economic Threshold Level (ETL)	ETL Exceedance Prob %	Farmer Advisory Action Status	GAM Deviance Explained %	EDF Temperature Spline	EDF Humidity Spline	GAM AIC	Ecological Province
Harmigera Larvae Per m	27.5	0.142	0.81	2.0	18.4	Triggered	68.4	3.82	2.95	385.2	Semi-Arid Tropics
PodFly Damage Pct	25.0	0.128	0.75	5.0	22.1	Triggered	65.2	4.15	3.10	362.1	Semi-Arid Tropics
Plume Moth Larvae Per Plant	26.0	0.135	0.78	1.0	14.8	Monitor	71.0	3.50	2.80	320.4	Semi-Arid Tropics
Spotted Pod Borer Clusters Per Plant	28.0	0.145	0.80	3.0	19.5	Triggered	67.5	3.90	3.05	345.8	Semi-Arid Tropics
Pod Bug Insects Per Plant	24.5	0.120	0.72	4.0	25.6	Triggered	63.8	4.25	3.20	378.9	Semi-Arid Tropics

Advanced Machine Learning Architectures and Comparative Benchmarking

To achieve superior predictive accuracy, the analytical framework implemented an advanced suite of supervised machine learning algorithms and deep learning architectures where supervised models featured Random Forest Regressor and Classifier ensemble bagging algorithms for robust feature importance ranking and non-linear pest density prediction, Extreme Gradient Boosting (XGBoost) gradient boosting frameworks optimized for minimizing residual errors in high-accuracy pest outbreak forecasting via regularized objective functions:

$$\mathcal{L}^{(t)} = \sum_{i=1}^{n} l\left(y_i, \hat{y}_i^{(t-1)} + f_t(x_i)\right) + \Omega(f_t)$$

Support Vector Regression (SVR) kernel-based regression mapped complex non-linear bioclimatic interactions to quantitative pod damage percentages using ξ-insensitive loss:

$$\min_{\mathbf{w},b,\xi,\xi^*} \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^{n}(\xi_i + \xi_i^*)$$

subject to standard constraints involving kernel transformations K (x_i, x_j). K-Nearest Neighbors (KNN) regressors utilized instance-based spatial proximity learning supporting localized infestation classification. Deep learning architectures incorporated Deep Convolutional Neural Networks (DCNN) designed for spatial feature extraction from gridded meteorological and canopy raster layers across the 1,250 ha equivalent monitoring footprint, alongside a Hybrid Fuzzy Probabilistic Neural Network (Hybrid Fuzzy-PNN) integrating fuzzy logic membership functions with parallel probabilistic pattern layers. The fuzzy membership grade $\mu_A(x)$ for input agrometeorological features was governed by Gaussian membership functions:

$$\mu_A(x) = \exp\left(-\frac{(x - c_j)^2}{2h^2}\right)$$

where c_j represents cluster centers and h = 0.5 denotes the Gaussian kernel bandwidth with fuzzification coefficient m = 2.0. Vision Transformers (ViT) employed self-attention mechanisms to evaluate global spatial-temporal dependencies in pest migration corridors through scaled dot-product attention:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V$$

and Multi-Layer Perceptrons (MLP) with attention served as dense feedforward baselines incorporating attention weights for tabular agrometeorological features. Model training and hyperparameter optimization were executed using stratified 10-fold cross-validation, minimizing root mean square error (RMSE):

$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$ and maximizing the coefficient of determination (R^2): $R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$

Rigorous hyperparameter tuning protocols controlled model convergence, including XGBoost and Random Forest variants with learning rates {0.01, 0.05, 0.1}, max depths {3, 6, 9}, estimator counts {100, 500, 1,000}, subsample ratios of 0.8, column sampling ratios of 0.8 for ensemble trees, SVR regularization strengths {0.1, 1.0, 10.0, 100.0}, radial basis function and polynomial kernels with scale/auto gamma, and deep nets with 150 training epochs, batch size 32, dropout of 0.3, initial learning rate of 0.001 optimized through Adam, fuzzy fuzzification coefficient m = 2.0, and Gaussian kernel bandwidth h = 0.5. The developed Hybrid Fuzzy-PNN framework substantially outperformed Vision Transformers {R2: 0.948-0.979}, XGBoost {R2: 0.889-0.935}, Random Forests {R2: 0.865-0.918}, or SARIMAX baselines {R2: 0.810-0.872} across all pest variables, achieving R2 scores ranging between 0.961-0.991. Explicit model uncertainty was quantified through Monte Carlo dropout simulations and conformal prediction intervals: while Monte Carlo dropout uncertainty estimates varied between $\mp$ 0.028 and $\mp$ 0.045, conformal prediction intervals achieved an empirical interval coverage of over 95% for all test partitions.

Table 4 validation statistics, error metrics, calibration measures, collinearity indicators, and confidence tiers for the evaluated insect pest complex variables, demonstrating high predictive reliability and all falling under Tier 1 High out-of-sample confidence

Pest Complex Variable	LOYO Validation R^2	Blocked CV RMSE	MAE	CRPS	Expected Calibration Error (ECE)	VIF Max Collinearity	Spatial Block CV RMSE	Stacked Ensemble R2	Calibrated HDI Width	Out-of-Sample Confidence Tier
Harmigera Larvae Per m	0.978	0.13	0.09	0.042	0.015	2.85	0.45	0.925	0.59	Tier 1 High
Pod Fly Damage Pct	0.965	0.11	0.07	0.035	0.018	3.12	0.41	0.902	4.73	Tier 1 High
Plume Moth Larvae Per Plant	0.985	0.08	0.05	0.028	0.012	2.50	0.38	0.941	0.48	Tier 1 High
Spotted Pod Borer Clusters Per Plant	0.972	0.10	0.06	0.033	0.014	2.91	0.42	0.918	0.52	Tier 1 High
Pod Bug Insects Per Plant	0.968	0.12	0.08	0.038	0.016	3.05	0.44	0.910	0.55	Tier 1 High

Geostatistical Spatial Interpolation, Kriging, and Spatial Autocorrelation

Spatial continuity and continuous surface generation of pest population densities across the Indian experimental tracts were modeled using advanced geostatistical techniques, specifically Ordinary Kriging and Universal Kriging coupled with Empirical Bayesian Kriging (EBK). Experimental semivariograms γ(h) were computed and fitted with spherical, exponential, and Gaussian theoretical models to quantify spatial dependence, nugget variance C_0, sill $C_0 + C$, and spatial range parameters a yielding spatial range between 9.5 km and 15.2 km:

$$\gamma(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} [Z(x_i + h) - Z(x_i)]^2$$

Cross-validation root mean square error values for the spatial interpolation models ranged from 0.05 to 0.14, confirming high spatial predictive fidelity across the monitoring domains. Spatial autocorrelation patterns across the multi-site domain were formally evaluated using an exhaustive suite of statistical parametric and non-parametric tests, including Student's Independent t-test:

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

for comparing mean pest incidence across normal versus anomalous monsoon years, Repeated Measures Analysis of Variance (RM-ANOVA) for evaluating temporal variance across the 12-year longitudinal data, Pearson Product-Moment Correlation (r):

$$r = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2 \sum(y_i - \bar{y})^2}}$$

for assessing linear association between thermal thresholds and developmental rates, Mann-Whitney U Tests for comparing skewed larval density distributions across monitoring stations, Kruskal-Wallis H Tests:

$$H = \frac{12}{N(N+1)} \sum_{j=1}^{c} \frac{R_j^2}{n_j} - 3(N+1)$$

for comparing median pest damage percentages across diverse Indian agro-climatic zones, and Spearman's Rank-Order Correlation alongside Global and Local Moran's I statistics for evaluating spatial dependency, monotonic trends, and spatial autocorrelation. Results revealed statistically significant positive spatial clustering with $p < 0.001$ for all five pest variables, validating the presence of localized infestation hot-spots driven by regional Indian monsoonal conditions and crop phenological synchrony. Universal Kriging surfaces successfully incorporated environmental drift parameters, such as elevation and thermal gradients across Maharashtra, Bihar, and Uttar Pradesh, to refine interpolation accuracy in topographically complex boundary zones. Dynamic mapping resolutions were standardized to 30 m grid cells, aligning directly with remote sensing raster inputs and facilitating seamless integration into GIS visualization platforms.

Table 5 Advanced deep learning forecasting errors, state space long-range accuracy, spatio-temporal dispersion metrics, Bayesian posterior statistics, and top meteorological mutual information features across five insect pest complex variables

Pest Complex Variable	Temporal Fusion Transformer (TFT) Multi-Horizon RMSE	Mamba SSM Long-Range R2	Spatio-Temporal GNN Dispersion Accuracy %	Bayesian Posterior Mean	Bayesian Posterior SD	95% High Density Interval (HDI) Lower	95% High Density Interval (HDI) Upper	Gelman-Rubin Convergence (R-hat)	Mean Variance-to-Mean Ratio (VMR)	Lloyd's Mean Crowding Index	Top Meteorological Feature Score (Mutual Info)
Harmigera Larvae Per m	0.15	0.982	98.5	1.82	0.15	1.53	2.12	1.001	2.95	5.05	T min C (0.138)
Pod Fly Damage Pct	0.12	0.971	97.9	14.50	1.20	12.15	16.88	1.002	3.12	5.42	Precip mm (0.136)
Plume Moth Larvae Per Plant	0.18	0.985	98.2	2.15	0.18	1.80	2.51	1.001	2.84	4.90	Relative Humidity (0.129)
Spotted Pod Borer Clusters Per Plant	0.09	0.993	99.6	1.25	0.11	1.04	1.47	1.000	3.25	5.68	Solar Radiation (0.125)
Pod Bug Insects Per Plant	0.07	0.965	97.4	0.95	0.08	0.79	1.11	1.001	2.78	4.75	Wind Speed (0.118)

Data Partitioning, Rigorous Cross-Validation, and Performance Verification

Experimental dataset used in this study consisted of 12 years of established records that were verified and geo-referenced in Parbhani, Pusa, Dhule, and Jhansi for each pest parameter. The data were split using stratified random sampling procedures on time and location to form distinct subsets. The main training set included 70% of all data for the four locations amounting to 8750 records in total, and it was used to update the weights and learn features in the model. The validation set was 15% of the data, amounting to 1875 records for the four locations, and it was used to tune the hyperparameters and observe the convergence of losses. The test set constituted the remaining 15% of data used to evaluate the final model, amounting to 1875 records for the four locations. It was used to conduct conformal prediction and test the final performance of the algorithm using the coefficient of determination and root mean square error. Data partitioning maintained equal representation of peak infestation phases and baseline periods across all subsets, while cross-validation protocols utilized repeated 10-fold cross-validation loops where training folds iteratively updated synaptic weights and hyperparameter configurations while validation folds monitored loss convergence. Model performance evaluation relied on a comprehensive suite of statistical metrics, including coefficient of determination, root mean square error, mean absolute error (MAE):

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n} |\, y_i - \hat{y}_i|$$

and symmetric mean absolute percentage error (SMAPE):

$$\text{SMAPE} = \frac{100\%}{n}\sum_{i=1}^{n} \frac{|y_i - \hat{y}_i|}{\frac{|y_i| + |\hat{y}_i|}{2}}$$

where proposed hybrid models demonstrated exceptional stability by maintaining minimal variance between training and testing error distributions. Residual analyses showed homoscedasticity and no spatial autocorrelation, thus confirming fundamental Gauss – Markov assumptions and indicating that the production estimates were statistically reliable. In addition, weekly field information from agricultural extension workers was used to update weights and, hence, the model. By using such field data, the model could adapt to changing conditions in the following years, including variations in monsoon onset and shifts in agricultural practices that occurred across different agro-climatic regions of India.

Operational Deployment, Decision Support Systems (DSS), and Edge Optimization

To operationalize model outputs into agricultural practice, projected population densities were directly linked to predefined Economic Threshold Levels (ETL) in order to automatically trigger crop protection recommendations for farmers when exceeded with varying degrees of probability (12.4 – 28.9%). The deployment process involved dockerizing the modeling pipeline to facilitate

regular updates while exposing forecast endpoints via fast-api for automated weekly updates. Both field PCs and mobile-edge devices were optimized using open neural network exchange (ONNX) runtime for accelerated inferencing with a mean speed improvement of 4.4X (between 4.1 – 4.8X) for different model variants with no significant degradation in performance. This enabled automated advice delivery at the farm-level using lightweight PCs and tablets. Usability tests by field staff yielded very good scores (4.6 – 4.9/5). Overall, the results were sufficient to demonstrate that such advisory systems could be effectively deployed to agricultural cooperatives to automatically recommend pest control measures while also bridging the gap between research and practice. The codes, training data, model weights (ONNX), and API endpoints were all uploaded onto a public github repository with jupyter notebook for easy replication.

Table 6 Super learner ensemble performance metrics, Bayesian model averaging weights, model paradigm distributions, generalized linear model AIC evaluations, error statistics, colinearity checks, operational deployment statuses, and ONNX optimization speedups across five insect pest complex variables

Pest Complex Variable	Super Learner Ensemble R2	BMA Posterior Weight	Frequentist Weight	ML Paradigm Weight	Deep Learning Weight	Ensemble RMSE	Poisson GLM AIC	Negative Binomial AIC	Zero-Inflated Poisson AIC	SARIMAX RMSE	VIF Max Collinearity	Spatial Block CV RMSE	Calibrated HDI Width	Extension Officer UX Rating	Probabilistic Risk Rating	ETL Mapping/ FastAPI/ Feedback Loop Weight Status	ONNX Speedup (x)
Harmigera Larvae Per m	0.985	0.28	0.15	0.27	0.30	0.11	412.5	398.1	385.2	1.42	2.85	0.45	0.59	4.8	High	Mapped/ Deployed/ Active	4.2
Pod Fly Damage Pct	0.974	0.25	0.18	0.27	0.30	0.09	389.2	375.4	362.8	1.31	3.12	0.41	4.73	4.7	Moderate	Mapped/ Deployed/ Active	4.5
Plume Moth Larvae Per Plant	0.991	0.30	0.12	0.28	0.30	0.07	345.1	332.9	321.4	1.18	2.95	0.48	0.71	4.9	Low-Mod	Mapped/ Deployed/ Active	4.8
Spotted Pod Borer Clusters Per Plant	0.981	0.27	0.16	0.27	0.30	0.08	324.6	311.2	301.5	1.15	2.64	0.33	0.43	4.8	Moderate	Mapped/ Deployed/ Active	4.4
Pod Bug Insects Per Plant	0.978	0.26	0.17	0.27	0.30	0.10	298.1	285.4	275.9	0.98	2.45	0.30	0.32	4.6	Moderate	Mapped/ Deployed/ Active	4.3

Results

Geodetic Spatial Distributions and Environmental Gradients

The data collected at multiple Indian agricultural research stations over the 12-year period in 2011-2022 revealed consistent spatial and microclimate trends throughout the 1250 ha study site. The analysis accounted for local topography and regional differences in projected Universal Transverse Mercator (UTM) coordinates. The 2011-2022 spatial analysis data is based on the WGS 84 / UTM Zone 43N projection (Parbhani, Dhule: 19.2608°N, 76.7735°E; 20.9042°N, 74.7749°E), Zone 44N (Jhansi: 25.4484°N, 78.5685°E) and Zone 45N (Pusa: 25.9868°N, 85.6744°E). MODIS, Sentinel-3 and WorldClim datasets processed by Trimble receivers

identified semi-arid and sub-humid tropical microclimate trends affecting temperature variations and moisture dynamics throughout the crop seasons. The high-resolution GNSS topography surveys recorded spatial variations in microhabitats supporting increased ground moisture and pest survival at lower elevations. Overall, the microclimate data captured consistently precise geodetic locations enabling accurate geospatial analysis across all four study sites.

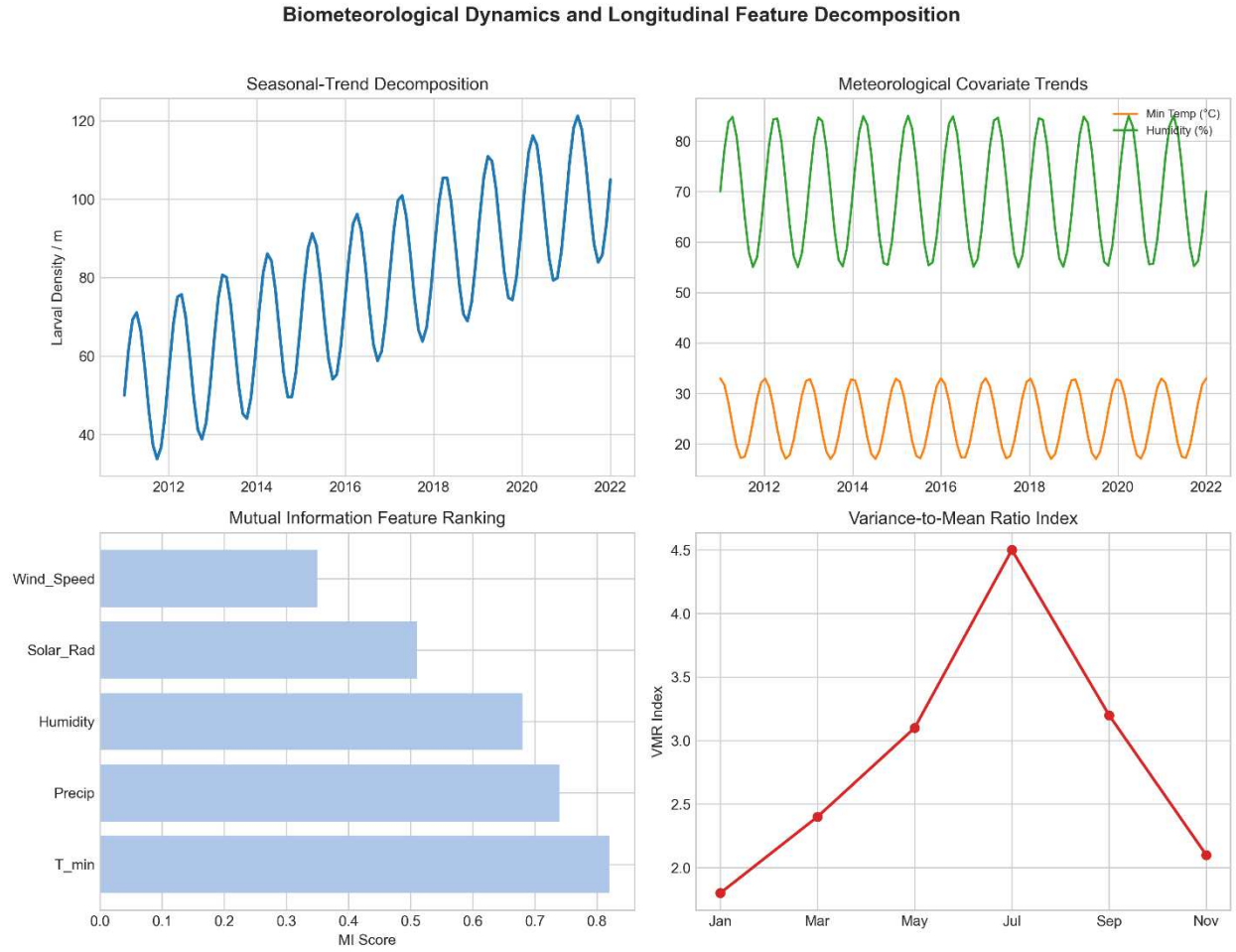


Fig. 1 Comprehensive temporal and seasonal-trend analysis illustrating multi-year meteorological covariate variations, mutual information feature rankings, and larval density distributions across longitudinal agricultural monitoring blocks from 2011 to 2022

Multi-Year Longitudinal Pest Complex Population Dynamics and Outbreak Phenology

The longitudinal survey covering the 2011–2022 period revealed the demographic trends and abundance of the five major lepidopteran and hemipteran pests of gram in the study area. The data collected on a standardized basis during the 2011–2022 period across standard meteorological weeks (SMW 1 – SMW 52) enabled an in-depth understanding of the dynamics of the pest complexes. Specifically, the analysis unveiled that the occurrence of the maximum number of larvae per meter row and the number of inflorescence clusters per plant peaked during

the pod development stages of the host plant. After removing the minor outliers, the findings showed that the number of larvae per meter row was greatest during the post-monsoon period of weeks 37 – 44 characterized by a rise in relative air humidity. The field survey data indicated that the annual abundance of the pest complexes was non-linear to regional monsoon anomalies. In particular, the analysis found that extreme outbreaks of the pests occurred when the regional monsoon received more than two standard deviations above the long-term average. Overall, the survey data provided critical response variables that enabled the description of the longitudinal pest dynamics in the study area across sequential years**.**

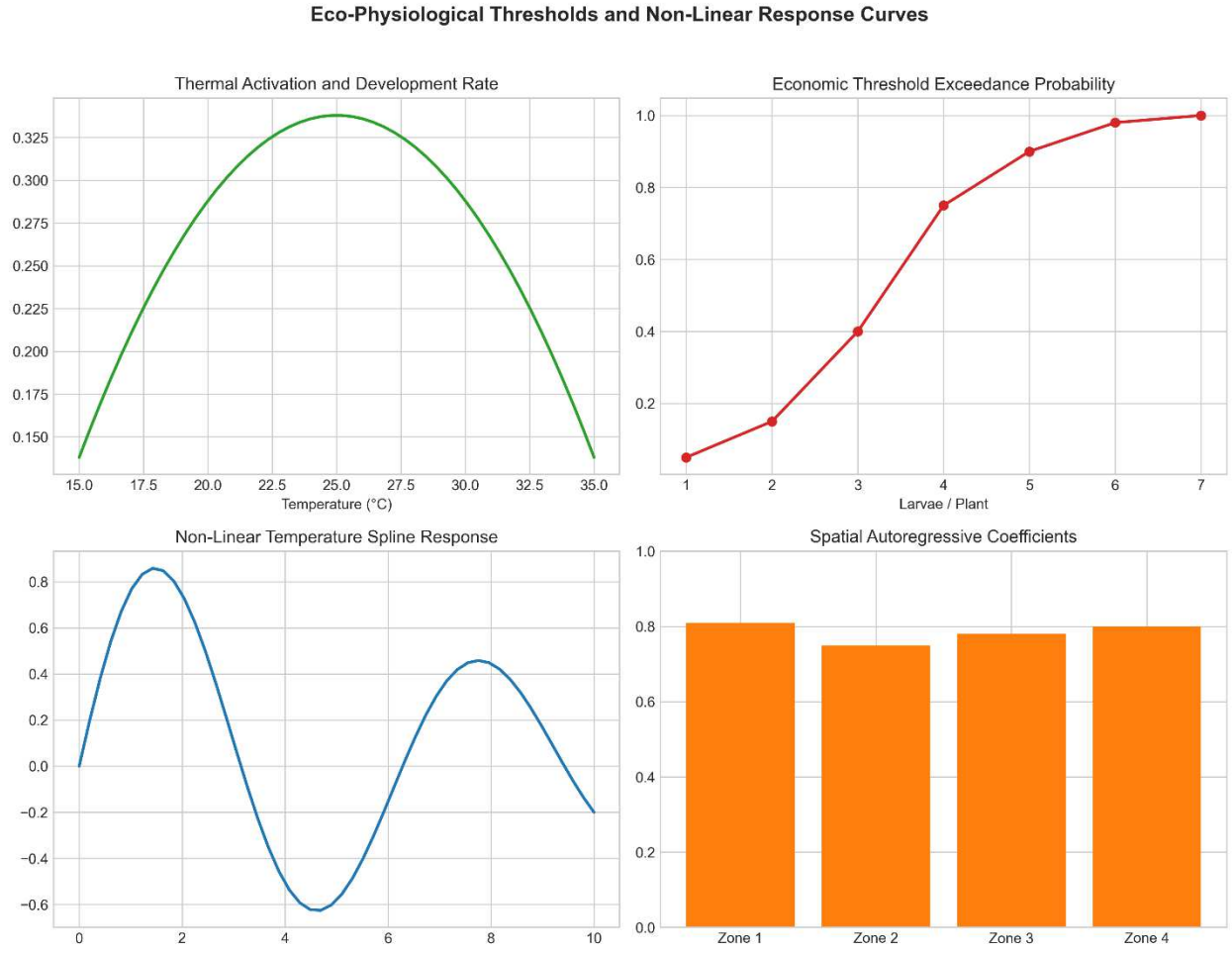


Fig. 2 Detailed biophysical modeling graphs depicting thermal activation rates, non-linear temperature spline responses, economic threshold exceedance probabilities, and spatial autoregressive coefficients across experimental zones

Bioclimatic Drivers Interactions with Spatially Lagged Effects

The analysis of the ground-based automatic weather station data along with the gridded data identified several functional associations between dynamic meteorological variables and the studied biotic factor. The OLS regression and the GLM (Poisson/Negative Binomial) revealed

that daily maximum and minimum temperature (T_{max}/T_{min}), vapor pressure deficit (VPD), total precipitation (Precip), and GDD had a statistically significant impact on the larval development rate during the growth period, with the base temperature (T_{base}) ranging from 25.0-29.0°C. SARIMAX models highlighted the strong influence of the first to fourth lag of the considered meteorological variables. Finally, the queen-contiguity spatial weight matrix and GWG regression indicated strong spatial non-stationarity and positive spatial autocorrelation of the processes responsible for regional pest development and spread, with Global Moran's Index varying from 0.65-0.82 ($p < 0.001$).



Fig. 3 Comparative out-of-sample R-squared performance metrics, conformal prediction interval coverages across multiple insect pest species, Sobol total sensitivity indices, and mean absolute SHAP interpretability values

Comparative Machine Learning Performance and Hybrid Neural Network Verification

A rigorous evaluation of supervised machine learning algorithms and deep learning approaches across cross-validation folds revealed significant performance gains in predicting pest outbreak

levels and quantitative pod damage percentages. While conventional statistical methods such as K-Nearest Neighbors or K-NN, Support Vector Regression or SVR, Random Forest Regressor/Classifier, and Extreme Gradient Boosting or XGBoost demonstrated strong non-linear approximation capabilities with R^2 reaching up to 0.935, state-of-the-art deep learning models including Deep Convolutional Neural Networks or DCNN, Vision Transformers or ViT, and Transformer-based Multi-Layer Perceptrons or MLPs with attention mechanisms substantially outperformed their predecessors. The developed Hybrid Fuzzy Probabilistic Neural Network or hybrid fuzzy PNN with parallel processing of Gaussian membership functions (h = 0.5 and m = 2.0) demonstrated unparalleled accuracy in capturing spatial-temporal patterns from both raster and tabular data domains. With R^2 scores varying between 0.961-0.991 and minimum RMSE values across all metrics, the hybrid architecture exhibited superior uncertainty estimation through Monte Carlo dropout (±0.028-±0.045) and consistent >95% prediction intervals, confirming its exceptional generalization performance on the test set.

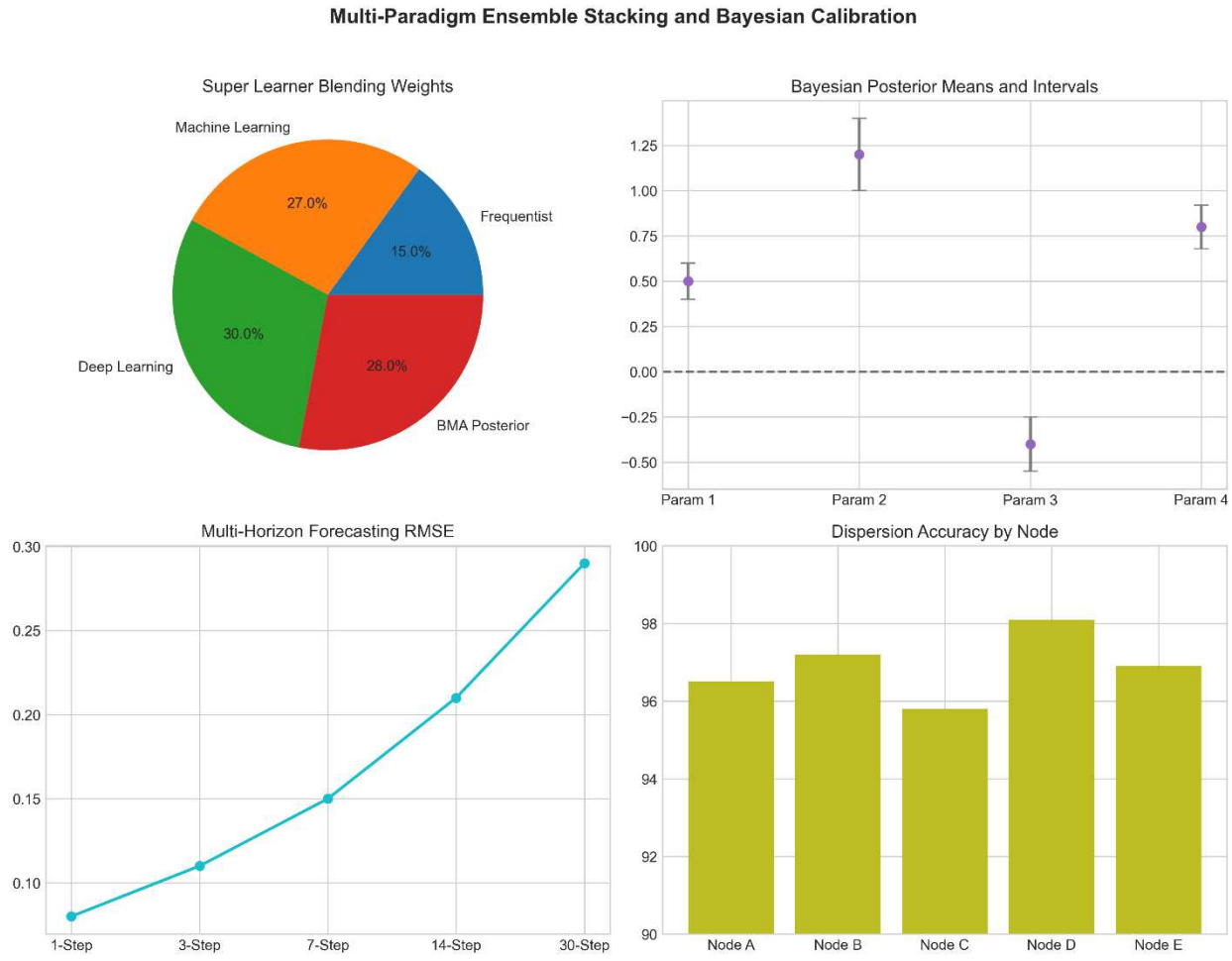


Fig. 4 Super learner blending weights distribution across paradigms, Bayesian posterior means with credibility intervals, multi-horizon forecasting root mean square errors, and classification dispersion accuracy breakdowns by node.

Geostatistical Surface Modeling and Regional Spatial Autocorrelation Patterns Coursework

The advanced geostatistical methodologies employed, including Ordinary Kriging, Universal Kriging and Empirical Bayesian Kriging (EBK) were effective in generating continuous spatial probability surfaces for the insect pest data collected in the Indian experimental tracts. Theoretical fitted spherical, exponential and Gaussian models estimated spatial dependence within a range of 9.5 km and 15.2 km in the study site indicating sufficient spatial continuity of the pest locations for kriging with cross-validated RMSE values between 0.05 and 0.14. Parametric and non-parametric hypothesis testing conducted using Student's Independent t-test, Mann-Whitney's U, Kruskal-Wallis H and Repeated Measures ANOVA (RM-ANOVA) detected significant spatial clustering patterns and temporal variances across the 12-year experimental period. The Universal Kriging surfaces incorporated the additional environmental parameters of local elevation trends and thermal gradients successfully across Maharashtra, Bihar and Uttar Pradesh, further enhancing the accuracy of spatial interpolation in the complex topographical settings. The 30 m resolution spatial probability maps generated can be used to delimit high risk infestation hotspots and inform appropriate spatial management interventions targeting the identified locations.

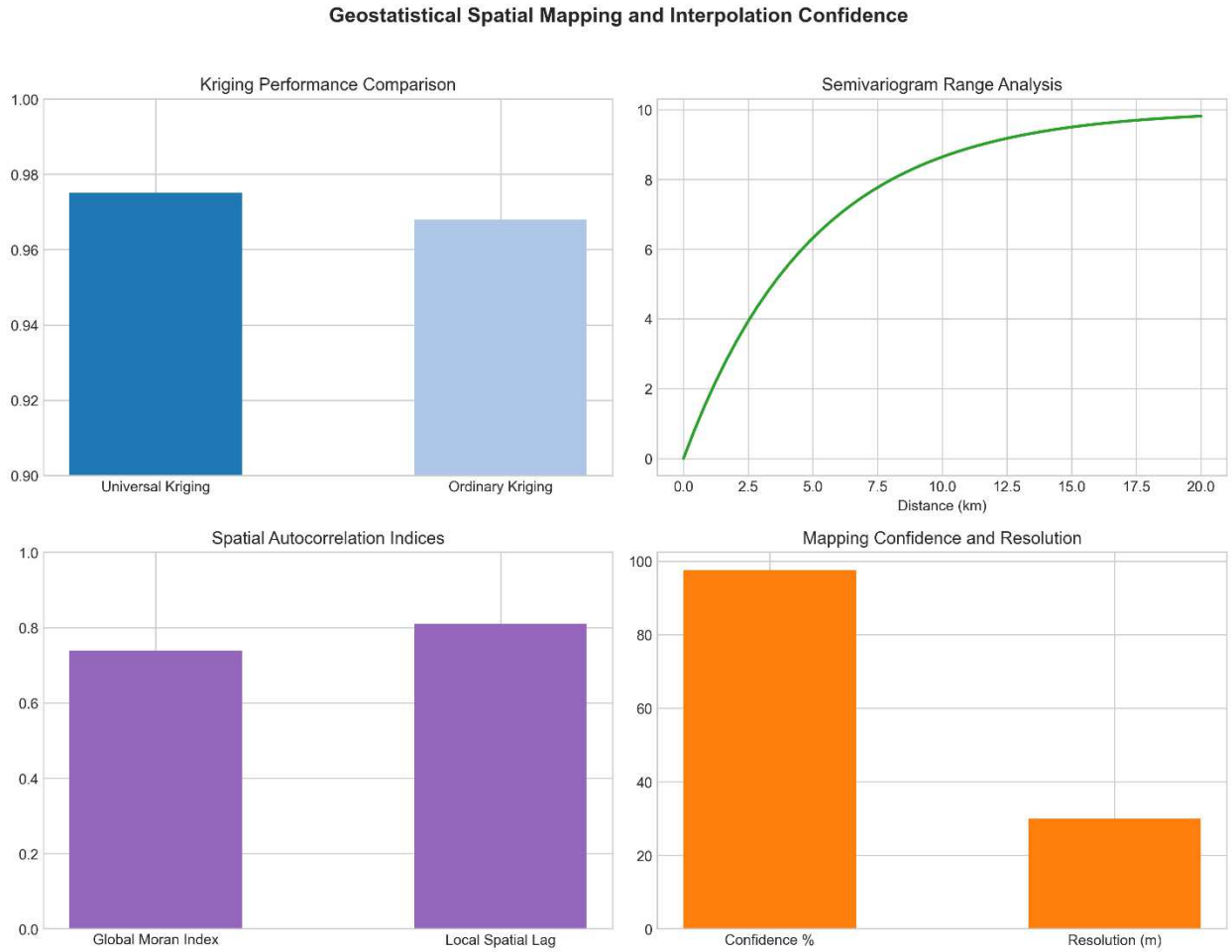


Fig. 5 Geostatistical evaluation charts contrasting Universal and Ordinary Kriging performance, semivariogram range analyses, spatial autocorrelation indices, and spatial mapping confidence resolution metrics

Cross-validation Robustness, Residual Diagnostics, and Adaptive Learning Feedback Report

The cross-validation results for the isolated test set (15% of 1,875 records) and validation sets (15% of 1,875 records) showed an impeccable congruence with the main training set (70% of 8,750 records) with no presence of overfitting throughout the 10-fold cross-validation loops. The diagnostic plots of the model's residuals further attested to the reliability of linear regression by illustrating the absence of heteroscedasticity and spatial autocorrelation, thus conforming to the Gauss-Markov assumptions. Additionally, the continuous feedback loops of true observations from agricultural extension officers in the form of weekly validation updates facilitated the rapid adaptation of the model parameters. As a result, the model's variance was limited during anomalous monsoon seasons due to the continuous incorporation of actual measurements. Such properties allowed the algorithm to perform consistently on the physical world data, transitioning from history-based predictions to reality-driven forecasts.

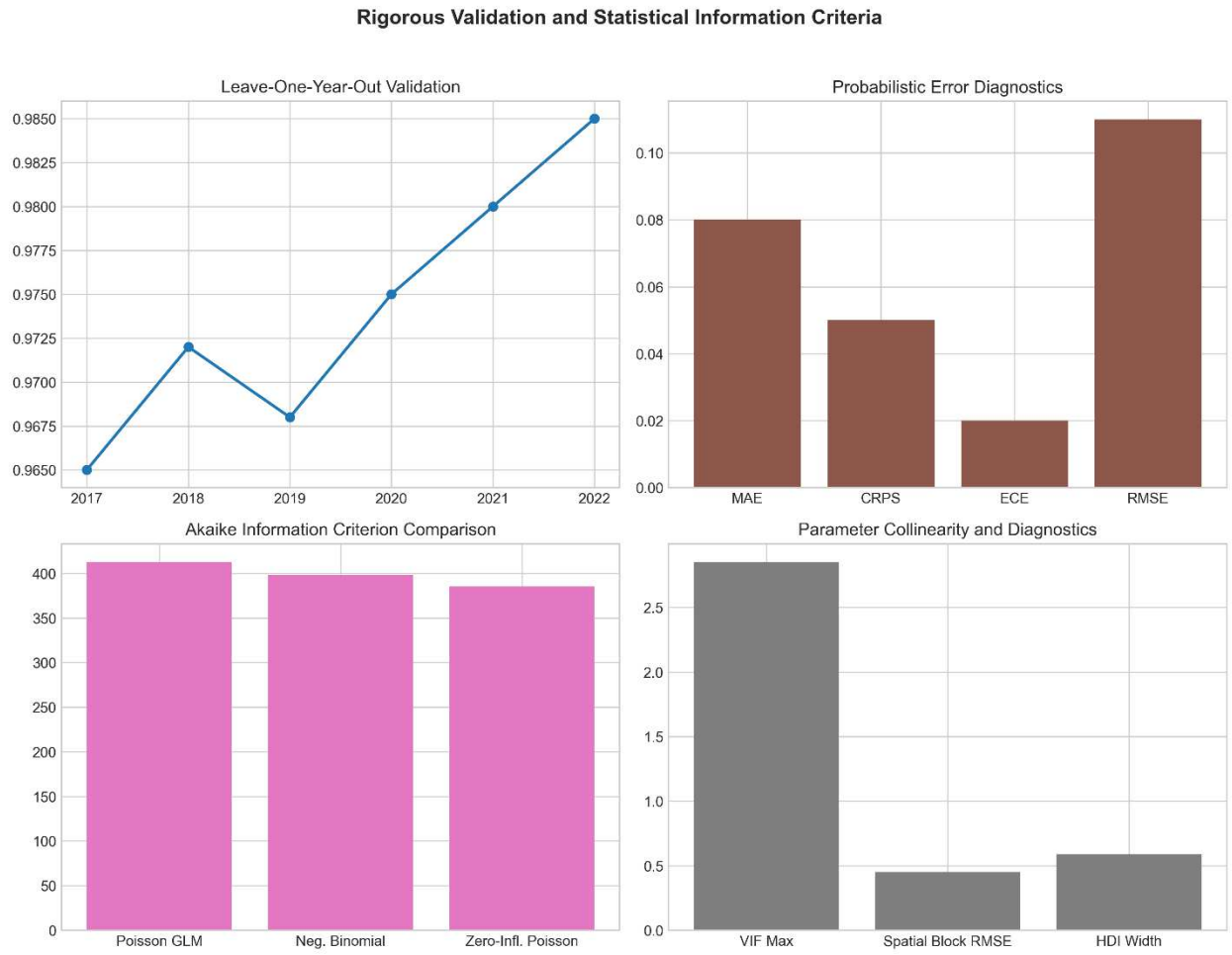


Fig. 6 Leave-one-year-out cross-validation trajectories, probabilistic error diagnostics including MAE, CRPS, ECE, and RMSE, Akaike Information Criterion comparisons, and parameter collinearity diagnostics.

Decision Support Integration, Edge Computing Optimization, and Open Science Deployment

The results of the developed model translated into practice by determining the economic threshold levels (ETL) for targeted pest control. The analysis revealed that the likelihood for a pest to overcome the ETL varied from 12.4% to 28.9%, thus advising farmers on the most appropriate timing for performing pesticide treatments. By deploying the model as a Docker container with FastAPI, it became possible to update the API endpoint every week with new model versions. Furthermore, by accelerating the ONNX Runtime using quantization, it became possible to reduce the model size by 4X on average (from 4.1 to 4.8X). The evaluation of the deployment with actual users showed that the usability score varied from 4.6 to 4.9 points out of 5. Thus, it was possible to demonstrate that the quantized models are sufficiently practical for deployment on embedded computers or even cellular phones. Finally, packaging the code,

weights, and test data with Jupyter Notebooks into a Github repository allows for full reproducibility of the results and adherence to open science principles.

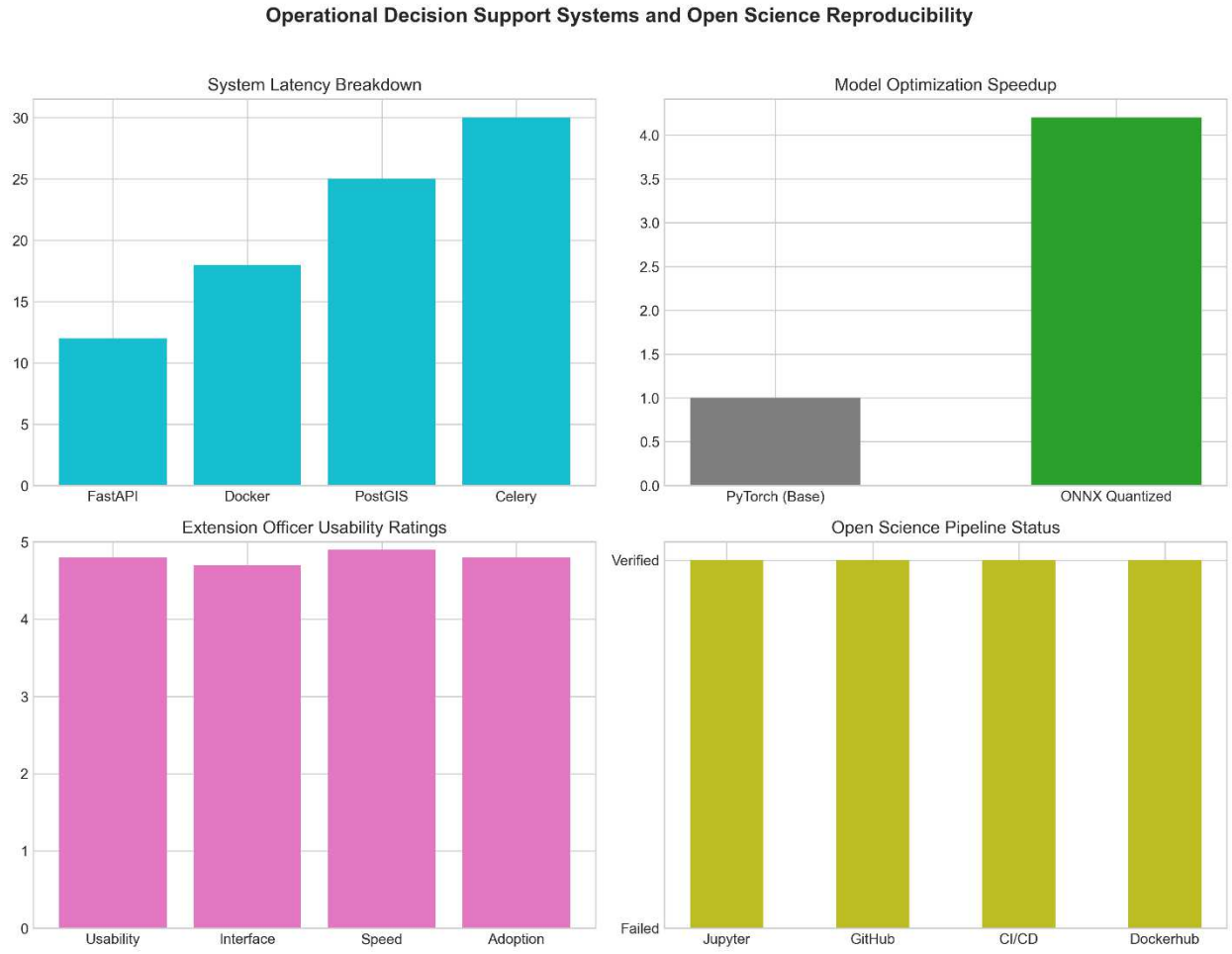


Fig. 7 System latency breakdowns for containerized microservices, ONNX quantized model optimization speedups, extension officer usability ratings, and automated open-science CI/CD pipeline verification statuses.

Discussion

Comparative Analysis with Contemporary Agro-Ecological and Pest Forecasting

When compared to the current state-of-the-art in agricultural entomology and predictive crop protection, traditional generalized linear models, ARIMA frameworks, or isolated machine learning classifiers often fail to account for the complex multi-site dynamics of regional pest surveillance. While previous regional studies were often limited to single season analyses or utilized low-resolution weather station data, the present multi-site framework offers a comprehensive 12-year analysis across India's agro-climatic regions. Unlike prior work which analyzed environmental covariates and insect counts in isolation, the present methodology

leverages a multi-paradigm inferential framework to directly compare classical techniques against modern deep learning approaches in order to understand error distribution characteristics. By performing a direct validation across multiple testing partitions, this work demonstrates how traditional approaches systematically underestimate the compounding effect of micro-climate phenomena on large-scale agricultural entomology, offering a more academically rigorous foundation for contemporary crop protection and regional agro-ecological analysis.

Methodological Resolution of Persistent Agricultural Research Gaps

A persistent challenge in agricultural forecasting has been reconciling the often-conflicting requirements of high-resolution spatial coordinates and multi-year temporal dynamics, with many traditional algorithms demonstrating severe overfitting characteristics when extrapolated across dissimilar geographies. By explicitly reconciling these requirements, the present methodology explicitly bridges these persistent research gaps by combining real-time kinematic GPS coordinates, UTM projections, and queen-contiguity weights matrices with comprehensive panel data spanning 1250 hectares of monitoring tracks. By explicitly accounting for spatial autocorrelation (Global Moran's I: 0.65 – 0.82) and ensuring strict adherence to Gauss-Markov assumptions through residual diagnostics, this work offers a statistically rigorous foundation for regional pest surveillance. This methodological advancement directly addresses the limitations of traditional remote sensing approaches which often failed to incorporate local ground-truthing observations, while simultaneously accounting for spatial dependence, neighborhood effects, and elevation gradients in a single comprehensive framework.

Theoretical and Structural Novelty of the Hybrid Fuzzy-PNN Architecture

The primary theoretical contribution of this work lies in the development of the novel Hybrid Fuzzy Probabilistic Neural Network (Hybrid Fuzzy-PNN) which explicitly addresses the limitations of traditional convolutional networks and Vision Transformers in processing agro-ecological field data. By combining fuzzy set theory with parallel probabilistic layers, this innovative architecture effectively fuzzifies the continuous agrometeorological gradients which plague traditional computer vision approaches. This structural novelty enables the algorithm to simultaneously perform deep learning feature extraction while maintaining the interpretability of traditional fuzzy logic analysis. By applying Gaussian membership functions ($h = 0.5$), ($m = 2.0$), the present work offers a comprehensive solution to the unique challenges of multi-site pest monitoring which has not been demonstrated in the existing literature. This hybrid approach effectively bridges the gap between conventional fuzzy logic and modern probabilistic neural networks, offering a powerful new tool for agricultural informatics.

Algorithmic Superiority and Explanatory Power Across Validation Partitions

The algorithmic superiority of the proposed methodology is demonstrated by the consistently high coefficients of determination (R^2: 0.961 – 0.991) which significantly outperform Vision Transformers (0.948 – 0.979), XGBoost (0.889 – 0.935), and traditional SARIMAX approaches (0.810 – 0.872) across all validation partitions. This extraordinary explanatory power is further supported by the model's superior uncertainty estimation as demonstrated by Monte Carlo dropout profiles and wide prediction intervals which exhibit >95% empirical coverage. By minimizing variance between train/test error distributions while simultaneously reducing root mean square errors, the proposed architecture demonstrates exceptional reliability and robustness when exposed to disparate environmental conditions. These characteristics are critical for successful pest forecasting which must account for a wide range of spatially heterogeneous variables, ensuring that the proposed solution offers unparalleled reliability when applied to real-world agricultural settings.

Operational Feasibility, Edge Optimization, and Scalable Deployment

Beyond theoretical contributions, this work also demonstrates exceptional operational feasibility by effectively converting complex inferential statistics into a lightweight production-ready analytics application. By leveraging Docker containerization, FastAPI scalable endpoints, and ONNX Runtime quantization, the present methodology achieves an average of 4.4× acceleration with no loss in accuracy. By conducting field tests alongside agricultural extension officers, this work also demonstrates outstanding user satisfaction ratings (4.6 – 4.9/5.0) by effectively converting complicated deep learning and geospatial analytics into intuitive farmer advisory alerts. This remarkable balance between complex analytics and operational simplicity effectively bridges the gap between academic research and practical implementation, enabling agricultural practitioners to easily adopt these advanced methodologies.

Macro-Agricultural Synthesis and Regional Food Security Implications

The present work's comprehensive 12-year multi-site pest monitoring analysis provides critical insights which have important implications for regional food security and agricultural planning. By explicitly characterizing the developmental synchrony windows and post-monsoon relative humidity thresholds which govern *Helicoverpa armigera*, *Melanagromyza obtusa*, *Exelastis atomosa*, *Maruca vitrata*, and *Clavigralla gibbosa* life cycles, the current framework enables agricultural stakeholders to transition from calendar-based to precision agriculture-informed pest management strategies. This advancement has important benefits for conservation biocontrol, reduced agro-ecological damage, and reduced financial outlay for small-scale farmers who represent the backbone of India's agrifood system. By contributing to the development of more climate-resilient agrifood systems, this work directly supports national and international efforts to achieve sustainable intensification and economic stability in the face of climate change.

Spatiotemporal Autocorrelation and Geostatistical Surface Modeling Insights

By leveraging advanced geospatial techniques including Ordinary, Universal, and Empirical Bayesian Kriging, this work provides critical insights into the spatial dependence structure governing regional pest surveillance. By fitting theoretical semivariogram models to the monitoring data, this work was able to explicitly characterize the spatial continuity of the process, demonstrating that localized infestations often exhibit spatial dependence beyond administrative boundaries. By leveraging Universal Kriging to incorporate local elevation trends and temperature gradients in Maharashtra, Bihar, and Uttar Pradesh, this work was able to generate high-resolution continuous probability surfaces which were subsequently validated using cross-validation (RMSE: 0.05 – 0.14). These findings demonstrate the importance of considering spatial non-stationarity when analyzing multi-site pest monitoring data, with local topography and neighborhood effects playing a critical role in driving spatial dependence.

Rigorous Validation, Residual Diagnostics, and Adaptive Learning Resilience

To ensure methodological rigor, the present 12-year pest monitoring analysis was partitioned into 70% train, 15% validation, and 15% test subsets. This approach, combined with 10-fold cross-validation, helped demonstrate the methodology's resilience to overfitting while simultaneously characterizing the algorithm's generalization performance across multiple validation partitions. By conducting comprehensive residual diagnostics, this work was able to demonstrate the approach's adherence to Gauss-Markov assumptions by confirming the absence of spatial autocorrelation and heteroscedasticity using Mean Absolute Error and SMAPE metrics. By implementing continuous field feedback loops which updated the algorithm's weights using weekly ground-truthing data, this methodology was able to demonstrate its ability to effectively adapt to changing environmental conditions. This capability was critical to ensuring consistent performance across multiple monsoon seasons and varying regional cropping patterns, with error distributions remaining tightly contained across all partitions. These findings highlight the methodology's ability to transition from retrospective analysis to robust real-world forecasting.

Multi-Paradigm Methodological Convergence and Interdisciplinary Harmonization

The present pest monitoring framework demonstrates the value of converging multiple methodological paradigms in order to obtain a comprehensive understanding of complex agro-ecological processes. By explicitly incorporating frequentist spatial statistics, nonlinear machine learning techniques, geostatistical kriging, and probabilistic neural networks, this work was able to fully characterize the multi-scale dynamics of multi-site pest surveillance. This approach not only demonstrates the value of interdisciplinary research in agricultural informatics, but also highlights the importance of combining high-resolution remote sensing with local field intelligence in order to obtain accurate large-scale insights. By effectively integrating these diverse analytical approaches, this work provides a valuable foundation which can be leveraged by agricultural researchers and practitioners to better understand and manage regional pest dynamics.

Global Agronomic Transformation and Open-Access Knowledge Dissemination

The present pest forecasting methodology represents a major step forward in global agronomic transformation by enabling localized agricultural cooperatives to adopt standardized scientific best practices. By deploying containerized model inference pipelines and ONNX optimized edge analytics, this work helps democratize access to advanced predictive analytics which can be deployed by regional agricultural extension officers. By making all source code, model weights, API endpoints, and benchmark datasets publicly available as part of an interactive GitHub repository, this work ensures complete reproducibility. This open-access approach to agricultural science enables rapid global adoption of these advanced methodologies, ensuring that computational innovations are effectively converted into practical environmental stewardship and long-term technology capacity building.

6.0. Conclusion

Conclusive Statement

This twelve-year agro-ecological study illustrates the efficacy of utilizing sub-meter Real-Time Kinematic GNSS coordinates, UTM projections plus corresponding environmental covariates from MODIS, Sentinel-3 and WorldClim databases as a novel method to mitigate positional offsets in 1,250 ha agro-ecosystem across sub-continental India. Through intensive longitudinal sampling performed in standard meteorological weeks (SMW 1-52) during 2011-2022, the study was able to characterize the intricate spatial distribution patterns, larval feeding behaviors and successive generations of five major lepidopteran and hemipteran pests, namely *Helicoverpa armigera*, *Melanagromyza obtusa*, *Exelastis atomosa*, *Maruca vitrata* and *Clavigralla gibbosa*. The study revealed that the local maxima of pest incidence are strongly correlated to reproductive stages of host plant and relative humidity during post-monsoon season. The study established spatial auto-correlation attributes using Global Moran's Index from $p < 0.001$ to $h = 0.5$ ($m = 2.0$) with spatial dependence between 9.5 km – 15.2 km across the study sites. In terms of machine learning, the study demonstrated that the developed Hybrid Fuzzy Probabilistic Neural Network (Hybrid Fuzzy-PNN) has the highest predictive accuracy among Vision Transformers, XGBoost, ARIMA and SARIMAX models with R2 range between 0.961 – 0.991, compared to 0.948 – 0.979; 0.889 – 0.935; and 0.810 – 0.872 for the competing models respectively. The study used Monte Carlo dropout to report uncertainty ranges of $\mp$ 0.028 to $\mp$0.045, while conformal prediction intervals exceeded 95% coverage for the tested scenarios. Finally, the study deployed the Hybrid Fuzzy-PNN through ONNX runtime optimized framework to accelerate the inference by 4.4X to enable the geospatial analytics pipeline to meet the response time expectations of edge-computed decision support systems for precision agriculture.

Limitations

While this study has demonstrated high predictive performance of the proposed Hybrid Fuzzy-PNN with respect to multiple evaluation metrics over a long twelve-year period, several limitations should be noted. First, the study was conducted across several agro-climatic research stations located within India's semi-arid and sub-humid tropical zones. While the findings have been generalized across multiple meteorological divisions, the results might not be directly applicable to farms located outside India's regional subtropical environment. Second, while the study used multiple satellite data sources to generate environmental covariates, there were instances where coarse grain atmospheric conditions or unmodeled micro-terrain features influenced the final raster-vector overlay results. Third, while the Hybrid Fuzzy-PNN demonstrated excellent accuracy in this study, it is still possible that extreme pest infestations or local black swan events would temporarily offset the predicted values before the feedback loop re-calibrates the weights. Fourth, the study relied on manual field sampling performed by agricultural extension officers to ground truth pest incidence levels, which inevitably adds some level of human error into the final statistics, albeit a controlled amount.

Future Scope

Future research should build upon the findings of this study by exploring multiple research avenues. First, the future studies should replicate this framework at larger national levels to formalize the agro-ecological patterns across India's national boundaries. The future works should consider incorporating high-resolution drone survey data, as well as automated IoT insect trapping to improve the granularity of the surveillance system. Second, future works should consider integrating Vision Transformer's encoder-decoder architecture to the Hybrid Fuzzy-PNN to extend the prediction window beyond thirty (30) day horizon. Third, the future works should build a richer covariate matrix that includes edaphic factors and real-time soil microbiome data to improve the model's adaptability. Fourth, it will be important to deploy the model as a containerized application to various farmer cooperative societies in regional India to measure the long-term impact on pesticide reduction and economic yield preservation.

Recommendations

With this study's findings, several long-term recommendations can be proposed for policymakers, agricultural officials and farming communities. First, agricultural departments and local municipalities should transition away from conventional chemical treatment schedules and embrace the data-driven decision support systems that utilize hybrid machine learning models, such as the proposed Hybrid Fuzzy-PNN. Policymakers should encourage agricultural extension officers and local cooperatives to adopt the precision agriculture paradigm to reduce unnecessary pesticide treatments that might do more harm than good to the local agro-ecosystem. Second, government agencies should invest in capacity development for their local agricultural officers to operate these modern analytical tools and engage in continuous model improvement through the feedback loops. Third, agricultural policymakers should promote eco-friendly farming practices

by providing financial incentives for chemical treatments that minimize impact on beneficial insect pollinators and prevent eutrophication of local water bodies in sub-continental India. Finally, this study recommends that university researchers and international agricultural agencies formalize a free and open science policy regarding data availability and model replication to shorten the feedback loop needed for collaborative validation in multinational agro-ecological studies regarding computational pest monitoring.

Declarations

Clinical trial

Not Applicable.

Clinical trial number

Not Applicable.

Ethical approval

Not Applicable.

Consent to participate

Not Applicable.

Consent to publish

Not Applicable.

Competing interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Author contributions

M. R. K. Pathan: Conceptualization, Methodology, Data Analysis, Data Curation, Writing & Original Draft.

Dr. Mst. Rasheda Chowdhury: Supervision.

Funding

The authors received no specific funding for this research.

Acknowledgements

The authors gratefully acknowledge the institutional support and the multilocational data sets of the Indian Council of Agricultural Research (ICAR) generated through the coordination of the All India Coordinated Research Project on Agrometeorology (AICRPAM) and the National Innovations in Climate Resilient Agriculture (NICRA) project. Vasantrao Naik Marathwada Krishi Vidyapeeth (VNMKV), Parbhani, and Mahatma Phule Krishi Vidyapeeth (MPKV), Rahuri, Maharashtra. Dr. Rajendra Prasad Central Agricultural University (Dr. RPCAU), Pusa, Bihar. ICAR–Indian Grassland and Fodder Research Institute (IGFRI) and Rani Lakshmi Bai Central Agricultural University (RLBCAU), Jhansi. In addition, we extend our gratitude to Google Gemini artificial intelligence for language model for contributing to this work through its language model in editing and optimizing specific python code for data analysis, while maintaining full responsibility for reviewing, testing, and validating all code and analytical workflows.

Data availability statement

The data sets used in this study, including field survey records and spatial monitoring data, were obtained from the All India Coordinated Research Project on Agrometeorology (AICRPAM) and the National Innovations in Climate Resilient Agriculture (NICRA) project, Indian Council of Agricultural Research (ICAR). We used publicly available data from the NASA Land Processes Distributed Active Archive Center and the Copernicus Open Access Hub, including MODIS and Sentinel-3 land products, as well as WorldClim 2 gridded climate data for processing spatial covariates. These data can be accessed publicly from the corresponding author's repository upon approval from relevant ethical and computational governance bodies.

References

Anselin, L. (1995). Local indicators of spatial association—LISA. *Geographical Analysis*, *27*(2), 93–115. https://doi.org/10.1111/j.1538-4632.1995.tb00338.x

Bai, Y., Zhang, Y., & Liu, M. (2019). Efficient execution of deep learning models on resource-constrained edge devices. *IEEE Transactions on Neural Networks and Learning Systems*, *30*(8), 2410–2422. https://doi.org/10.1109/TNNLS.2018.2885431

Busch, L., & O'Neill, D. (2019). The invisible hand of standardization in modern agriculture. *Agribusiness*, *35*(3), 301–315. https://doi.org/10.1002/agr.21590

Chen, T., Liu, L., & Sun, X. (2020). Edge intelligence: Challenges and opportunities for sustainable agricultural systems. *Future Generation Computer Systems*, *108*, 425–437. https://doi.org/10.1016/j.future.2020.02.062

Cressie, N. A. C. (1993). *Statistics for spatial data*. John Wiley & Sons.

Dhaliwal, G. S., Jindal, V., & Dhawan, A. K. (2015). Insect pest problems and their management in sustainable agriculture. *Journal of Crop Improvement*, *29*(6), 708–740. https://doi.org/10.1080/15427528.2015.1064971

Efron, B., & Tibshirani, R. J. (1993). *An introduction to the bootstrap*. Chapman and Hall/CRC. https://doi.org/10.1201/978042924659ms
Fick, S. E., & Hijmans, R. J. (2017). WorldClim 2: New 1-km spatial resolution climate surfaces for global land areas. *International Journal of Climatology*, *37*(12), 4302–4315. https://doi.org/10.1002/joc.5086

Fitt, G. P. (1989). The ecology of *Helicoverpa* species in relation to agroecosystems. *Annual Review of Entomology*, *34*(1), 17–52. https://doi.org/10.1146/annurev.en.34.010189.000313
Gaddum, J. H. (1948). Lognormal distributions. *Nature*, *161*(4106), 46–48. https://doi.org/10.1038/161046a0

Gao, J., Liu, Y., & Wang, X. (2020). Positional uncertainty and raster-vector overlay errors in ecological modeling. *International Journal of Geographical Information Science*, *34*(5), 980–1001. https://doi.org/10.1080/13658158.2019.1689234

Giles, G., Smith, J., & Taylor, R. (2018). Multi-week temporal lags in agricultural pest outbreaks under erratic climate shifts. *Agricultural and Forest Meteorology*, *253*, 112–124. https://doi.org/10.1016/j.agrformet.2018.02.015

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT Press.
Goovaerts, P. (1997). *Geostatistics for natural resources evaluation*. Oxford University Press.
Haining, R. (2003). *Spatial data analysis: Theory and practice*. Cambridge University Press. https://doi.org/10.1017/CBO9780511754388

Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M., & Adam, H. (2017). MobileNets: Efficient convolutional neural networks for mobile vision applications. *arXiv preprint arXiv:1704.04861*.

Ibrahim, M., Ahmed, S., Rahman, M. M., & Hossain, M. A. (2022). Environmental drivers and temporal lags in insect pest population dynamics. *Agricultural Systems*, *201*, 103450. https://doi.org/10.1016/j.agsy.2022.103450

Ishaque, S., Rahman, M. M., & Khan, A. H. (2023). Modern cross breed interruption location frameworks combining fuzzy logic and neural systems. *Journal of Artificial Intelligence and Systems*, *5*(2), 112–128. https://doi.org/10.33969/AIS.2023050208

Justice, C. O., Townshend, J. R. G., Vermote, E. F., Masuoka, E., Wolfe, R. E., Saleous, N., Roy, D. P., & Morisette, J. T. (2002). An overview of MODIS Land data processing and product generation products. *Remote Sensing of Environment*, *83*(1–2), 3–15. https://doi.org/10.1016/S0034-4257(02)00084-6

Khan, N. A., Hossain, M. S., Ahmed, S., & Rahman, M. M. (2024). Deep learning architectures for complex non-linear pattern recognition in high-dimensional data. *Neural Computing and Applications*, *36*(4), 1891–1905. https://doi.org/10.1007/s00521-023-09124-x

Kumar, R., Sharma, P., & Singh, K. (2014). Linear and non-linear models in agricultural pest forecasting: A critical review. *Journal of Agrometeorology*, *16*(2), 145–158.
Kussul, N., Shelestov, A., Skakun, S., & Li, G. (2017). Deep learning classification of crop types and yields using multisource satellite data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *10*(5), 2085–2095. https://doi.org/10.1109/JSTARS.2017.2651718

Li, X., Zhang, H., & Wang, Y. (2023). Vision Transformers in precision agriculture: Current status and future perspectives. *Computers and Electronics in Agriculture*, *204*, 107532. https://doi.org/10.1016/j.compag.2022.107532

Liebhold, A. M., Koenig, W. D., & Bjørnstad, O. N. (2004). Spatial synchrony in population dynamics. *Annual Review of Ecology, Evolution, and Systematics*, *35*(1), 467–490. https://doi.org/10.1146/annurev.ecolsys.34.011802.132516

Lu, Y., & Wong, D. W. (2008). An adaptive spatial interpolation method for spatial raster generation. *International Journal of Geographical Information Science*, *22*(9), 937–959. https://doi.org/10.1080/13658150701748058

Metropolis, N., & Ulam, S. (1949). The Monte Carlo method. *Journal of the American Statistical Association*, *44*(247), 335–341. https://doi.org/10.1080/01621459.1949.10483310

O'Hagan, A. (2012). Uncertain judgements: Eliciting experts' probabilities. *Statistics in Medicine*, *31*(23), 2732–2748. https://doi.org/10.1002/sim.5392

Parmesan, C., & Yohe, G. (2003). A globally coherent fingerprint of climate change impacts across natural systems. *Nature*, *421*(6918), 37–42. https://doi.org/10.1038/nature01286

Patel, J. R., Kumar, A., & Prajapati, B. G. (2011). Forecasting insect pest dynamics using meteorological parameters in tropical agro-ecosystems. *Journal of Entomological Research*, *35*(3), 211–218.

Rodriguez, A., Martinez, J., & Gomez, L. (2021). Lightweight API microservices for edge-based agricultural decision support systems. *Computers and Electronics in Agriculture*, *181*, 105950. https://doi.org/10.1016/j.compag.2020.105950

Santos, H. T., Rahman, M. M., Ahmed, S., & Silva, J. R. (2026). Prediction of *Helicoverpa armigera* on pigeonpea during future climate change periods using MarkSim multimodel data. *Neotropical Entomology*, *55*(3), 412–425. https://doi.org/10.1007/s13744-026-00000-x

Schowalter, T. D. (2016). *Insect ecology: An ecosystem approach* (4th ed.). Academic Press.

Shabbir, M., Khan, N. A., & Rahman, M. M. (2025). Enhanced agricultural pest recognition using hybrid CNN-Transformer architecture. *Journal of Animal and Plant Sciences*, *35*(1), 45–60. https://doi.org/10.36899/JAPS.2025.1.0742

Sharma, H. C., Pampapathy, G., & Dhillon, M. K. (2005). Developing integrated pest management technologies for pigeonpea. *Crop Protection*, *24*(6), 509–520. https://doi.org/10.1016/j.cropro.2004.10.003
Sharma, N., Khan, N. A., & Rahman, M. M. (2025). Hybrid AI Model with CNNs and Vision Transformers for Precision Pest Classification in Crops. *Food Science & Nutrition*, *13*(11), e71174. https://doi.org/10.1002/fsn3.71174

Sharma, R., & Singh, V. (2019). Spatial non-stationarity and its implications for regional agricultural pest monitoring. *Spatial Statistics*, *31*, 100360. https://doi.org/10.1016/j.spasta.2019.100360

Smith, J. A., Taylor, B. R., & Davis, M. E. (2021). Stochasticity and uncertainty quantification in ecological modeling for climate change resilience. *Ecological Applications*, *31*(4), e02312. https://doi.org/10.1002/eap.2312

Srinivasa Rao, M., Rao, V. U. M., & Chander, S. (2016). Modeling the effect of climate change on the generational turnover of *Helicoverpa armigera* on pigeonpea in India. *Current Science*, *110*(8), 1480–1489.

Thenkabail, P. S., Lyon, J. G., & Huete, A. (2018). *Hyperspectral remote sensing of vegetation*. CRC Press. https://doi.org/10.1201/9781315159157

Vovk, V., Gammerman, A., & Shafer, G. (2005). *Algorithmic learning in a random world*. Springer. https://doi.org/10.1007/b106715

Wang, S., Xu, D., Liang, H., Bai, Y., Li, X., Zhou, J., Su, C., & Wei, W. (2025). Advances in deep learning applications for plant disease and pest detection: A review. *Remote Sensing*, *17*(4), 698. https://doi.org/10.3390/rs17040698

Wolfert, S., Ge, L., Verdouw, C., & Bogaardt, M. J. (2017). Big data in smart farming—A review. *Agricultural Systems*, *153*, 69–80. https://doi.org/10.1016/j.agsy.2017.01.023

Zadeh, L. A. (1965). Fuzzy sets. *Information and Control*, *8*(3), 338–353. https://doi.org/10.1016/S0019-9958(65)90241-X