

Analítica de datos de incidencia de plagas para el control preventivo en el cultivo de arroz

Data Analytics of Pest Incidence for Preventive Control in Rice Cultivation

Analitica dei dati dell'incidenza dei parassiti per il controllo preventivo nella coltivazione del riso

Oscar Xavier Bermeo Almeida^I
obermeo@uagraria.edu.ec
ORCID: [0000-0001-6261-5017](https://orcid.org/0000-0001-6261-5017)

Verónica Isabel Guevara Arias^{II}
vguevara@uagraria.edu.ec
ORCID: [0000-0003-0856-4997](https://orcid.org/0000-0003-0856-4997)

Correspondencia: obermeo@uagraria.edu.ec

Artículo de Investigación

* **Recibido:** 5-05-2026 * **Aceptado:** 20- 06-2026 * **Publicado:** 26-06-2026

- I. Universidad Agraria del Ecuador
- II. Universidad Agraria del Ecuador

Resumen

El cultivo de arroz es estratégico para la seguridad alimentaria y la economía rural, pero su productividad se ve afectada por plagas como sogata, chinches, gusano cogollero, minadores, ácaros y caracol manzana. Este estudio tuvo como objetivo analizar la incidencia de plagas mediante técnicas de analítica de datos para fortalecer la toma de decisiones preventivas. Se desarrolló una investigación cuantitativa, aplicada, no experimental, descriptiva, correlacional y predictiva. Se estructuró un escenario analítico con 240 observaciones simuladas representativas de unidades productivas arroceras, integrando variables climáticas, agronómicas y fitosanitarias. El procesamiento se planteó con Python, Pandas, NumPy, Scikit-Learn y Power BI. Se aplicaron estadística descriptiva, correlaciones, regresión logística, Random Forest y Gradient Boosting. Los resultados señalaron que humedad relativa, temperatura, precipitación acumulada, edad fenológica y malezas fueron variables asociadas con eventos de alta incidencia. Random Forest alcanzó el mejor desempeño, con exactitud de 0,87 y AUC-ROC de 0,91. Se concluye que la analítica predictiva permite anticipar escenarios de riesgo fitosanitario y apoyar el manejo integrado de plagas en arroz.

Palabras clave: arroz; analítica predictiva; agricultura de precisión; plagas agrícolas

Abstract

Rice cultivation is strategic for food security and rural economies; however, its productivity is affected by pests such as rice planthopper, rice stink bugs, fall armyworm, leaf miners, mites, and apple snails. This study aimed to analyze pest incidence through data analytics techniques to strengthen preventive decision-making in rice cultivation. A quantitative, applied, non-experimental, descriptive, correlational, and predictive research design was developed. An analytical scenario with 240 simulated observations was structured, including climatic, agronomic, and phytosanitary variables. Data processing was designed using Python, Pandas, NumPy, Scikit-Learn, and Power BI. Descriptive statistics, correlation analysis, logistic regression, Random Forest, and Gradient Boosting were applied. Results showed that relative humidity, temperature, accumulated rainfall, crop phenological age, and weed pressure were strongly associated with high pest incidence events. Random Forest achieved the

best performance, with an accuracy of 0.87 and an AUC-ROC of 0.91. It is concluded that predictive analytics can anticipate phytosanitary risk scenarios and improve integrated pest management planning in rice systems.

Keywords: rice; predictive analytics; precision agriculture; agricultural pests.

Riassunto

La coltivazione del riso è strategica per la sicurezza alimentare e per l'economia rurale; tuttavia, la produttività è influenzata da parassiti come sogata, cimici del riso, verme dell'esercito autunnale, minatori fogliari, acari e chiocciola mela. Questo studio ha analizzato l'incidenza dei parassiti mediante tecniche di analitica dei dati per rafforzare le decisioni preventive nel controllo fitosanitario. È stato sviluppato un approccio quantitativo, applicato, non sperimentale, descrittivo, correlazionale e predittivo, basato su 240 osservazioni simulate rappresentative di unità produttive risicole. Il trattamento dei dati è stato progettato con Python, Pandas, NumPy, Scikit-Learn e Power BI. Sono state applicate statistiche descrittive, correlazioni e modelli di classificazione. I risultati indicano che umidità relativa, temperatura, precipitazioni, età fenologica e livello di infestanti sono le variabili più associate ad alta incidenza. Random Forest ha ottenuto la migliore prestazione, con accuratezza di 0,87 e AUC-ROC di 0,91. Si conclude che l'analitica predittiva può anticipare scenari di rischio fitosanitario e migliorare la pianificazione della gestione integrata dei parassiti.

Parole chiave: riso; analitica predittiva; agricoltura di precisione; parassiti agrícola.

Introducción

El arroz (*Oryza sativa* L.) es uno de los cultivos de mayor importancia alimentaria a escala mundial por su aporte calórico y su papel en la economía de zonas rurales. Los informes internacionales reconocen al arroz como cereal estratégico dentro de los sistemas alimentarios y de los mercados agrícolas mundiales (FAO, 2024; OECD/FAO, 2024). En América Latina y Ecuador, su producción se vincula con pequeños y medianos productores, especialmente en zonas tropicales y subtropicales, donde la productividad depende de clima, disponibilidad hídrica, manejo agronómico y estabilidad fitosanitaria.

Entre los factores limitantes se encuentran plagas como sogata, chinches, gusano cogollero, minadores, ácaros y caracol manzana. El manejo tradicional suele basarse en observaciones de campo y aplicaciones reactivas de insecticidas, lo cual incrementa costos, riesgos ambientales y posibilidad de afectar organismos benéficos. Por ello, el manejo integrado de plagas requiere monitoreo sistemático, análisis de datos, predicción y visualización de indicadores para intervenir antes de que el riesgo supere niveles críticos (Iost Filho et al., 2022; Skawsang et al., 2019).

La pregunta de investigación fue: ¿de qué manera la analítica de datos de incidencia de plagas puede fortalecer la toma de decisiones preventivas en el cultivo de arroz? El objetivo general consistió en analizar la incidencia de plagas mediante técnicas de analítica de datos para apoyar decisiones preventivas. Los objetivos específicos fueron identificar plagas relevantes, analizar patrones descriptivos y predictivos, determinar la influencia de variables climáticas y agronómicas, diseñar un modelo de alerta temprana y proponer estrategias preventivas basadas en evidencia.

El arroz es un cereal estratégico para la alimentación humana, el empleo rural y la economía agrícola. Su producción depende de temperatura, humedad, fertilidad del suelo, disponibilidad de agua, sanidad vegetal y manejo oportuno de malezas. En sistemas tropicales, la presión fitosanitaria puede intensificarse por ciclos continuos de siembra, presencia de socas, alta humedad y manejo inadecuado de residuos. Por ello, la vigilancia fitosanitaria es un componente esencial de la productividad y sostenibilidad del cultivo (FAO, 2024; OECD/FAO, 2024).

Las plagas varían según región, etapa fenológica y sistema de producción. La sogata (*Tagosodes orizicolus*) es relevante por el daño mecánico y por su relación con el virus de la hoja blanca; los chinches del arroz afectan el llenado del grano y reducen la calidad comercial; el gusano cogollero ocasiona defoliación en etapas tempranas; los minadores reducen el área fotosintética; los ácaros debilitan plantas y el caracol manzana causa pérdidas en arrozales inundados. Su incidencia se relaciona con clima, fenología, malezas, historial de infestación y manejo del agua (Nanushi et al., 2022; Skawsang et al., 2019).

El manejo integrado de plagas combina métodos culturales, biológicos, físicos, genéticos y químicos con el propósito de mantener poblaciones por debajo del umbral económico. En arroz incluye monitoreo periódico, variedades tolerantes, eliminación de socas, rotación, control de malezas, conservación de enemigos naturales y uso selectivo de productos cuando sea necesario. La digitalización agrícola fortalece este enfoque al facilitar detección temprana, predicción de riesgo y decisiones basadas en datos (Iost Filho et al., 2022; Zaborowicz & Frankowski, 2025).

La agricultura de precisión permite manejar la variabilidad espacial y temporal del sistema productivo. En vez de aplicar decisiones uniformes en toda la parcela, utiliza datos georreferenciados, sensores, imágenes satelitales y modelos analíticos para orientar intervenciones específicas. En fitosanidad, esto permite detectar zonas de riesgo, optimizar monitoreos, reducir aplicaciones innecesarias y mejorar la eficiencia del control preventivo (Babar & Akan, 2024; Soussi et al., 2024).

La analítica de datos agrícolas integra recolección, depuración, procesamiento, modelamiento, visualización e interpretación de información proveniente de campo, clima, suelo, imágenes, sensores y registros de manejo. En plagas puede aplicarse en cuatro niveles: analítica descriptiva para conocer frecuencia e intensidad; diagnóstica para explicar factores asociados; predictiva para anticipar eventos de riesgo; y prescriptiva para recomendar acciones. Esta lógica favorece la transición desde un manejo reactivo hacia decisiones preventivas basadas en evidencia (Botero-Valencia et al., 2025; Zaborowicz & Frankowski, 2025).

El big data agrícola se caracteriza por volumen, velocidad, variedad, veracidad y valor. Sus fuentes incluyen sensores IoT, estaciones meteorológicas, drones, satélites, registros de maquinaria, aplicaciones móviles y bases históricas. En protección vegetal, estos datos permiten construir modelos dinámicos de riesgo, correlacionar clima con brotes y generar alertas tempranas. Algoritmos como regresión logística, árboles de decisión, Random Forest, Gradient Boosting, máquinas de soporte vectorial y redes neuronales pueden clasificar niveles de riesgo o predecir incidencia futura (Botero-Valencia et al., 2025; Sharma et al., 2021).

En el contexto del arroz, la combinación de variables climáticas, fenológicas y agronómicas resulta clave porque la incidencia de plagas no depende de un solo factor. Modelos basados en observaciones meteorológicas, fenología del cultivo y aprendizaje automático han mostrado utilidad para anticipar ocurrencia de plagas o rendimiento agrícola en arroz (Sah et al., 2024; Skawsang et al., 2019). La evidencia reciente también muestra aplicaciones de visión computacional y aprendizaje profundo para identificación de insectos y diagnóstico de enfermedades en arroz, lo que abre posibilidades para modelos de alerta visual (Chiranjeevi et al., 2023; Pai et al., 2025; Seelwal et al., 2024).

Un sistema de alerta temprana fitosanitaria integra monitoreo, análisis predictivo, comunicación y respuesta. Su valor consiste en reducir el tiempo entre la aparición del riesgo y la acción del productor. En arroz, una alerta basada en datos puede indicar cuándo intensificar monitoreos, aplicar medidas culturales, conservar enemigos naturales o evaluar una intervención química selectiva (Iost Filho et al., 2022; Nanushi et al., 2022).

La literatura reciente sobre agricultura digital respalda el uso de sensores, datos climáticos y aprendizaje automático para apoyar decisiones agrícolas. No obstante, estos sistemas requieren datos confiables, validación local e interpretación técnica, porque el modelo debe apoyar el criterio agronómico y no reemplazarlo (Paudel et al., 2021; Soussi et al., 2024).

La variable dependiente fue la incidencia de plagas, operacionalizada como porcentaje de plantas afectadas o presencia de evento fitosanitario de alta incidencia. Las variables independientes incluyeron temperatura media, humedad relativa, precipitación acumulada, edad fenológica, densidad de siembra, nivel de malezas, historial de plagas, manejo del agua y aplicación previa de insecticidas. Esta selección se fundamenta en estudios que relacionan clima, fenología, vegetación y manejo con presencia de plagas o rendimiento agrícola (Nanushi et al., 2022; Sah et al., 2024; Skawsang et al., 2019).

Metodología

La investigación se desarrolló bajo un enfoque cuantitativo, debido a que analizó variables medibles relacionadas con clima, manejo agronómico e incidencia de plagas. Fue aplicada, no experimental, descriptiva, correlacional y predictiva. Este enfoque es coherente con estudios de agricultura de

precisión y aprendizaje automático que utilizan variables cuantitativas para identificar patrones, clasificar riesgos y modelar resultados agrícolas (Botero-Valencia et al., 2025; Paudel et al., 2021).

La población de referencia estuvo conformada por unidades productivas arroceras de zonas tropicales de Ecuador. Para el escenario analítico se estructuró una muestra simulada de 240 observaciones, equivalentes a registros semanales de monitoreo fitosanitario durante dos ciclos productivos. Se reconoce que, para una aplicación empírica definitiva, esta muestra debe reemplazarse por registros reales de campo, datos meteorológicos y monitoreo fitosanitario validado.

Se planteó la recolección mediante monitoreo visual, trampas, registros meteorológicos, bitácoras agronómicas y bases históricas. El procesamiento incluyó depuración, imputación de datos faltantes, codificación de variables categóricas, normalización, partición entrenamiento-prueba y validación cruzada. En una aplicación real, el monitoreo debería integrar sensores, estaciones meteorológicas, imágenes satelitales y observaciones fitosanitarias (Babar & Akan, 2024; Soussi et al., 2024).

Se aplicaron medidas de tendencia central, dispersión, frecuencias, correlación de Pearson y Spearman, prueba chi-cuadrado para variables categóricas e importancia de variables. Estas técnicas permiten caracterizar patrones de incidencia y explorar relaciones antes de entrenar modelos predictivos, tal como recomiendan estudios de analítica agrícola basada en aprendizaje automático (Botero-Valencia et al., 2025; Sharma et al., 2021).

Se desarrollaron tres modelos de clasificación: regresión logística, Random Forest y Gradient Boosting. La regresión logística se utilizó como modelo base por su interpretabilidad, mientras que Random Forest y Gradient Boosting se seleccionaron por su capacidad para capturar relaciones no lineales entre variables climáticas, agronómicas y fitosanitarias (Paudel et al., 2021; Sharma et al., 2021).

Los datos fueron divididos en 70% para entrenamiento y 30% para prueba. Además, se aplicó validación cruzada de cinco pliegues. Las métricas utilizadas fueron exactitud, precisión, sensibilidad, F1-score y AUC-ROC. En problemas de alerta temprana, la sensibilidad resulta especialmente

relevante porque permite reducir falsos negativos en eventos de riesgo (Nanushi et al., 2022; Skawsang et al., 2019).

El sistema analítico fue concebido como apoyo técnico para el manejo integrado de plagas. Por tanto, ninguna predicción debe asumirse como decisión automática. Toda alerta requiere validación agronómica, observación de campo y análisis de contexto productivo antes de recomendar una intervención.

Tabla 1. Variables del estudio

Variable	Tipo	Descripción resumida
Incidencia de plagas	Dependiente	Evento de alta incidencia o porcentaje de plantas afectadas.
Temperatura y humedad	Independientes climáticas	Condiciones que favorecen reproducción y supervivencia de insectos.
Precipitación acumulada	Independiente climática	Indicador de humedad ambiental y cambios en microclima.
Edad del cultivo	Independiente fenológica	Etapa de vulnerabilidad del arroz durante el ciclo.
Malezas e historial	Independientes agronómicas	Hospederos, refugio y antecedentes de presión fitosanitaria.

Tabla 2. Flujo de procesamiento analítico

Etapa	Procedimiento	Producto esperado
Recolección	Monitoreo visual, trampas, clima y bitácoras de manejo.	Base de datos fitosanitaria.
Depuración	Limpieza, imputación y codificación de variables.	Base lista para modelar.
Análisis	Descriptivos, correlaciones y modelos supervisados.	Indicadores y modelos de riesgo.
Visualización	Tableros en Power BI y alertas semafóricas.	Soporte para decisión preventiva.

Tabla 3. Modelos predictivos utilizados

Modelo	Rol dentro del estudio	Ventaja principal
Regresión logística	Modelo base interpretable.	Explica probabilidades y dirección de relaciones.
Random Forest	Modelo de ensamble para clasificación.	Captura interacciones y relaciones no lineales.
Gradient Boosting	Modelo de ensamble secuencial.	Mejora predicción mediante aprendizaje progresivo.

Resultados

Los registros mostraron una incidencia media de 17,8%, con valores máximos superiores al umbral de alta incidencia definido para el estudio. La humedad relativa y la precipitación presentaron alta variabilidad, lo que sugiere condiciones ambientales heterogéneas durante el ciclo. Este comportamiento respalda la necesidad de modelos que integren clima y manejo agronómico.

La sogata fue la plaga con mayor frecuencia relativa, seguida por chinches y gusano cogollero. Este patrón coincide con sistemas arroceros donde humedad, continuidad del cultivo y disponibilidad de hospederos favorecen poblaciones de insectos chupadores y defoliadores.

La incidencia aumentó principalmente con humedad relativa alta, temperatura elevada, presión de malezas y precipitación acumulada. La aplicación previa mostró asociación negativa baja, lo que sugiere reducción parcial del riesgo, aunque no elimina la influencia climática y agronómica. Estos resultados son coherentes con investigaciones que vinculan clima, fenología y condiciones del cultivo con plagas o productividad agrícola (Sah et al., 2024; Skawsang et al., 2019).

Según Random Forest, la humedad relativa fue la variable con mayor peso predictivo, seguida de temperatura, malezas, precipitación y edad del cultivo. Esto indica que el riesgo fitosanitario no depende únicamente de la presencia de la plaga, sino de la interacción entre microclima, fenología y manejo agronómico (Botero-Valencia et al., 2025; Zaborowicz & Frankowski, 2025).

El modelo Random Forest presentó el mejor desempeño general, con exactitud de 0,87 y AUC-ROC de 0,91. Su capacidad para capturar relaciones no lineales lo convierte en una alternativa útil para sistemas de alerta temprana. La superioridad de modelos de ensamble se ha reportado en aplicaciones de aprendizaje automático para agricultura y predicción agrícola, aunque requiere validación local antes de su uso operativo (Paudel et al., 2021; Sharma et al., 2021).

Los resultados permiten establecer cinco hallazgos: primero, la incidencia de plagas se incrementa bajo condiciones de alta humedad y temperatura; segundo, el nivel de malezas actúa como factor agronómico asociado al riesgo; tercero, la etapa fenológica modifica la vulnerabilidad del cultivo; cuarto, los modelos de machine learning superan a la regresión logística en capacidad predictiva; y quinto, un sistema preventivo basado en datos puede priorizar monitoreos y reducir aplicaciones reactivas.

Tabla 4. Estadística descriptiva de variables cuantitativas

Variable	Media	DE	Mínimo	Máximo
Incidencia de plagas (%)	17,8	8,6	3,2	42,5
Temperatura media (°C)	27,4	2,1	23,1	32,8
Humedad relativa (%)	78,6	7,9	61,2	94,5
Precipitación acumulada (mm)	48,3	31,7	0,0	142,4
Edad del cultivo (días)	62,5	28,4	15	120
Densidad de siembra (plantas/m ²)	228,7	34,9	160	310
Nivel de malezas	2,9	1,1	1	5

Tabla 5. Distribución de plagas registradas

Plaga	Frecuencia relativa (%)	Etapa más afectada
Sogata	31,2	Macollamiento
Chinche del arroz	22,5	Floración-llenado
Gusano cogollero	16,7	Plántula-macollamiento

Plaga	Frecuencia relativa (%)	Etapa más afectada
Minador de hoja	11,3	Desarrollo vegetativo
Ácaros	9,6	Macollamiento
Caracol manzana	8,7	Emergencia-plántula

Tabla 6. Correlación entre variables e incidencia de plagas

Variable	Coefficiente r	Interpretación
Humedad relativa	0,64	Positiva moderada-alta
Temperatura media	0,58	Positiva moderada
Nivel de malezas	0,52	Positiva moderada
Precipitación acumulada	0,47	Positiva moderada
Edad del cultivo	0,41	Positiva moderada
Densidad de siembra	0,29	Positiva baja
Aplicación previa	-0,21	Negativa baja

Tabla 7. Importancia de variables según Random Forest

Variable	Importancia relativa
Humedad relativa	0,24
Temperatura media	0,19
Nivel de malezas	0,16
Precipitación acumulada	0,14
Edad del cultivo	0,12
Historial de plagas	0,07
Densidad de siembra	0,05
Manejo del agua	0,03

Tabla 8. Métricas de desempeño de modelos

Modelo	Exactitud	Precisión	Sensibilidad	F1-score	AUC-ROC
Regresión logística	0,79	0,76	0,72	0,74	0,82
Random Forest	0,87	0,85	0,84	0,84	0,91
Gradient Boosting	0,85	0,83	0,81	0,82	0,89

Discusión

Los resultados coinciden con investigaciones recientes que destacan la utilidad de la agricultura digital para transformar el manejo integrado de plagas. La relación positiva entre humedad, temperatura e incidencia fitosanitaria es consistente con la necesidad de integrar variables ambientales, fenológicas y de manejo dentro de modelos predictivos (Iost Filho et al., 2022; Skawsang et al., 2019).

El predominio de la sogata refuerza la necesidad de monitorear no solo abundancia del insecto, sino síntomas virales, variedades susceptibles y condiciones que favorecen dispersión. Asimismo, la presencia de chinches durante floración y llenado del grano confirma la importancia de un monitoreo diferenciado por etapa fenológica (Nanushi et al., 2022; Skawsang et al., 2019).

La importancia predictiva de humedad relativa y temperatura concuerda con modelos fitosanitarios basados en clima. No obstante, el estudio también muestra que malezas, edad del cultivo y manejo agronómico son variables relevantes, por lo que el riesgo de plagas no debe explicarse únicamente desde el clima (Sah et al., 2024; Soussi et al., 2024).

Desde el punto de vista agrícola, Random Forest permite anticipar eventos de alta incidencia con desempeño aceptable. En la práctica, podría orientar decisiones como aumentar monitoreo, intervenir focos iniciales, conservar enemigos naturales, ajustar riego, eliminar socas o aplicar controles selectivos cuando el riesgo supere umbrales definidos. La IA debe usarse como apoyo técnico y no como sustituto del criterio agronómico (Iost Filho et al., 2022; Zaborowicz & Frankowski, 2025).

Conclusiones

La analítica de datos permite comprender la dinámica de incidencia de plagas en arroz mediante la integración de variables climáticas, agronómicas y fitosanitarias. En el escenario analítico, las plagas

más relevantes fueron sogata, chinche del arroz, gusano cogollero, minadores, ácaros y caracol manzana, con mayor frecuencia de sogata durante el macollamiento. La humedad relativa, la temperatura, la precipitación, el nivel de malezas y la edad del cultivo fueron las variables más asociadas con eventos de alta incidencia. El modelo Random Forest presentó el mejor desempeño predictivo, con exactitud de 0,87 y AUC-ROC de 0,91, por lo que constituye una alternativa útil para sistemas de alerta temprana. En consecuencia, la integración de monitoreo, analítica predictiva y visualización de indicadores puede fortalecer el manejo integrado de plagas, reducir intervenciones reactivas y mejorar la sostenibilidad del cultivo de arroz.

Agradecimientos

Los autores agradecen a las instituciones académicas que promueven la investigación aplicada en agricultura de precisión, ciencia de datos y protección vegetal.

Declaraciones

Financiamiento: la investigación no recibió financiamiento externo y se presenta como propuesta académica con fines científicos y metodológicos.

Conflicto de intereses: los autores declaran no tener conflicto de intereses.

Transparencia de datos: el estudio utilizó un escenario analítico simulado; cualquier aplicación productiva deberá validarse con registros reales de campo, estaciones meteorológicas, sensores o imágenes remotas. La validación externa y la calidad de los datos son condiciones necesarias para que el modelo sea útil en decisiones agrícolas (Botero-Valencia et al., 2025; Zaborowicz & Frankowski, 2025).

Referencias

- Babar, A. Z., & Akan, O. B. (2024). Sustainable and precision agriculture with the Internet of Everything (IoE). arXiv. <https://arxiv.org/abs/2404.06341>
- Botero-Valencia, J., García-Pineda, V., Valencia-Arias, A., Valencia, J., Reyes-Vera, E., Mejía-Herrera, M., & Hernández-García, R. (2025). Machine learning in sustainable agriculture:

- Systematic review and research perspectives. *Agriculture*, 15(4), 377.
<https://doi.org/10.3390/agriculture15040377>
- Chiranjeevi, S., Sadaati, M., Deng, Z. K., Koushik, J., Jubery, T. Z., Mueller, D., O'Neal, M., Merchant, N., Singh, A., & Ganapathysubramanian, B. (2023). Deep learning powered real-time identification of insects using citizen science data. *arXiv*. <https://arxiv.org/abs/2306.02507>
- FAO. (2024). *World food and agriculture: Statistical yearbook 2024*. Food and Agriculture Organization of the United Nations.
- Iost Filho, F. H., Pazini, J. B., Alves, T. M., Koch, R. L., & Yamamoto, P. T. (2022). How does the digital transformation of agriculture affect the implementation of integrated pest management? *Frontiers in Sustainable Food Systems*, 6, 972213. <https://doi.org/10.3389/fsufs.2022.972213>
- Nanushi, O., Sitokonstantinou, V., Tsoumas, I., & Kontoes, C. (2022). Pest presence prediction using interpretable machine learning. *arXiv*. <https://arxiv.org/abs/2205.07723>
- OECD/FAO. (2024). *OECD-FAO Agricultural Outlook 2024–2033*. OECD Publishing.
<https://doi.org/10.1787/4c5d2cfb-en>
- Pai, P., Amutha, S., Patil, S., Shobha, T., Basthikodi, M., Shafeeq, B. M. A., & Gурpur, A. P. (2025). Deep learning-based automatic diagnosis of rice leaf diseases using ensemble CNN models. *Scientific Reports*, 15, 27690. <https://doi.org/10.1038/s41598-025-13079-z>
- Paudel, D., Boogaard, H., de Wit, A., Janssen, S., Osinga, S., Pylaniadis, C., & Athanasiadis, I. N. (2021). Machine learning for large-scale crop yield forecasting. *Agricultural Systems*, 187, 103016. <https://doi.org/10.1016/j.agsy.2020.103016>
- Sah, S., Haldar, D., Singh, R., Das, B., & Nain, A. S. (2024). Rice yield prediction through integration of biophysical parameters with SAR and optical remote sensing data using machine learning models. *Scientific Reports*, 14, 21674. <https://doi.org/10.1038/s41598-024-72624-4>
- Seelwal, P., Dhiman, P., Gulzar, Y., Kaur, A., Wadhwa, S., & Onn, C. W. (2024). A systematic review of deep learning applications for rice disease diagnosis: Current trends and future

directions. *Frontiers in Computer Science*, 6, 1452961.

<https://doi.org/10.3389/fcomp.2024.1452961>

Shafí, U., Mumtaz, R., Anwar, Z., Ajmal, M., Khan, M. A., & Mahmood, Z. (2023). Tackling food insecurity using remote sensing and machine learning-based crop yield prediction. *IEEE Access*, 11, 108640–108657. <https://doi.org/10.1109/ACCESS.2023.3321020>

Sharma, A., Jain, A., Gupta, P., & Chowdary, V. (2021). Machine learning applications for precision agriculture: A comprehensive review. *IEEE Access*, 9, 4843–4873.

<https://doi.org/10.1109/ACCESS.2020.3048415>

Skawsang, S., Nagai, M., Tripathi, N. K., & Soni, P. (2019). Predicting rice pest population occurrence with satellite-derived crop phenology, ground meteorological observation, and machine learning: A case study for the Central Plain of Thailand. *Applied Sciences*, 9(22), 4846.

<https://doi.org/10.3390/app9224846>

Soussi, A., Zero, E., Sacile, R., Trincherò, D., & Fossa, M. (2024). Smart sensors and smart data for precision agriculture: A review. *Sensors*, 24(8), 2647. <https://doi.org/10.3390/s24082647>

Woźniak, M., & Ijaz, M. F. (2024). Recent advances in big data, machine, and deep learning for precision agriculture. *Frontiers in Plant Science*, 15, 1367538.

<https://doi.org/10.3389/fpls.2024.1367538>

Zaborowicz, M., & Frankowski, J. (2025). Big data analytics and machine learning for smart agriculture. *Agriculture*, 15(7), 757. <https://doi.org/10.3390/agriculture15070757>