

Effectiveness of Machine Learning Models for Entrepreneurship, Innovation, and Crop Growth Prediction: A Systematic Literature Review

María De Los Angeles Espejo-Castillo, Estudiante de Ingeniería de Sistemas e Informática¹, and Josué Israel Clemente-Avila, Estudiante de Ingeniería de Sistemas e Informática¹, and Christian Abraham Dios-Castillo, Dr en Administración de la Educación¹

Universidad Tecnológica del Perú, U20307213@utp.edu.pe, U19310008@utp.edu.pe, C16763@utp.edu.pe

Abstract– In recent years, the application of Machine Learning (ML) in agriculture has emerged as a crucial innovation for enhancing productivity. This systematic literature review (SLR) aims to evaluate the effectiveness of Machine Learning models in the quantitative analysis of crops. Utilizing the PICO methodology and adhering to PRISMA standards, research articles published between 2015 and 2024 were thoroughly analyzed. Results sourced from Scopus highlight that the most commonly employed techniques are Deep Learning, Random Forest, and Support Vector Machine (SVM) due to their high accuracy and capacity to manage large datasets. The review also addresses challenges related to data quality and model implementation, emphasizing the need for ongoing research and international collaboration to advance agricultural technology. The findings underscore the transformative potential of ML in agriculture, paving the way for enhanced crop precision and improved decision-making processes.

Keywords: quantitative analysis, machine learning, plant breeding, nonhuman, agriculture.

Efectividad de Modelos de Machine Learning para el Emprendimiento, la Innovación y Predicción del Crecimiento de Cultivos. Una Revisión Sistemática de Literatura

María De Los Angeles Espejo-Castillo, Estudiante de Ingeniería de Sistemas e Informática¹, and Josué Israel Clemente-Avila, Estudiante de Ingeniería de Sistemas e Informática¹, and Christian Abraham Dios-Castillo, Dr en Administración de la Educación¹

Universidad Tecnológica del Perú, U20307213@utp.edu.pe, U19310008@utp.edu.pe, C16763@utp.edu.pe

Abstract– *En los últimos años, la aplicación de Machine Learning (ML) en la agricultura ha emergido como una innovación crucial para mejorar la productividad. Esta revisión sistemática de literatura (RSL) tiene como objetivo determinar la efectividad de los modelos de Machine Learning en el análisis cuantitativo de sembríos. Empleando la metodología PICO y adhiriéndose a los estándares PRISMA, se analizó meticulosamente artículos de investigación publicados desde 2015 hasta 2024. Los hallazgos rescatados de la fuente de Scopus, destacan que las técnicas más comúnmente utilizadas son Deep Learning, Random Forest y Support Vector Machine (SVM). Debido a su alta precisión y capacidad para manejar grandes conjuntos de datos. La revisión también discute los desafíos en la calidad de los datos y la implementación de modelos, enfatizando la necesidad de investigación continua y colaboración internacional para avanzar en la tecnología agrícola. Los resultados subrayan el potencial transformador del ML en la agricultura, allanando el camino para una mejor precisión de cultivos y procesos de toma de decisiones.*

Keywords: *quantitative analysis, machine learning, plant breeding, nonhuman, agriculture.*

I. INTRODUCCIÓN

En el contexto actual de la agricultura y la tecnología, el uso de técnicas avanzadas de Machine Learning está ganando prominencia como herramienta para mejorar la eficiencia y la producción. Los modelos de Machine Learning ofrecen nuevas oportunidades para la predicción del crecimiento de cultivos, patrones y la gestión agrícola [1]. Estos modelos avanzados permiten analizar grandes volúmenes de datos agrícolas de manera eficiente, extrayendo conocimientos que no serían fácilmente accesibles con métodos tradicionales. Al aplicar algoritmos de aprendizaje automático, los agricultores y los investigadores pueden predecir rendimientos de cultivos con mayor precisión. La problemática reside en ¿Cuál es la efectividad de los modelos de Machine Learning en el análisis cuantitativo de sembríos? Ello se justifica en que estos

modelos tienen la capacidad de manejar y analizar grandes conjuntos de datos, lo que permite identificar patrones y tendencias que no serían evidentes a simple vista. Además, los algoritmos pueden entrenarse para mejorar continuamente sus predicciones según se disponen de más datos, lo que aumenta la precisión y la confiabilidad de las estimaciones [2].

La investigación tiene como objetivo determinar la efectividad de los modelos de Machine Learning en el análisis cuantitativo de sembríos.

Esta Revisión Sistemática de Literatura (RSL), se centra en investigar los modelos de Machine Learning más efectivos para el análisis cuantitativo de los cultivos, considerando su influencia y utilidad en el ámbito de la innovación agrícola, como parte de la estrategia de emprendimiento de muchos inversores.

Esta revisión no solo busca identificar y evaluar los modelos existentes, sino también entender los desafíos metodológicos y técnicos que enfrenta la comunidad científica en este campo emergente como parte fundamental para el emprendimiento en el sector agroindustrial. A través de un enfoque sistemático utilizando la metodología PICO y los estándares PRISMA, se pretende proporcionar una visión clara y crítica de cómo los modelos de Machine Learning están siendo aplicados y qué resultados efectivos están siendo alcanzados en el análisis cuantitativo de cultivos.

II. METODOLOGÍA

Una RSL, es un método estructurado para identificar, evaluar e interpretar el trabajo de investigadores, académicos y profesionales en un área específica, recopilando y analizando críticamente múltiples estudios o investigaciones a través de un proceso metódico [3]. Para ello se utilizó la estrategia PICO, la cual tiene como objetivo, buscar un diagnóstico, mediante la definición de procedimientos y operaciones para obtener la información necesaria y poder resolver una necesidad de información concreta [4]. La pregunta principal es: ¿Qué modelos de Machine Learning existen para la

identificación de patrones cuantitativos de cultivos? y como subpreguntas se tienen a:

P: ¿Qué tipos de problemas de sembríos son estudiados a través de investigaciones de Machine Learning? I: ¿Qué modelos de Machine Learning existen para el análisis cuantitativo de sembríos? O: ¿Qué nivel de efectividad muestran los casos de aplicación de modelos de Machine Learning en el análisis cuantitativo de sembríos? En la Tabla I se muestran los keywords por cada componente PICO.

TABLA I
KEYWORDS POR CADA COMPONENTE PICO

Componente	KEYWORDS
P = problema	Quantitative patterns, quantitative traits, quantitative analysis
I = Intervención	Machine Learning, Artificial Intelligence, Data Mining
O = Resultados	Methods, Challenges, techniques

Nota: El componente C, no se ha considerado por ser opcional.

La base de datos científica utilizada para la búsqueda de información es Scopus, cuya cadena de búsqueda es la siguiente:

(TITLE-ABS-KEY ("quantitative patterns" OR "quantitative traits" OR "quantitative analysis")) AND TITLE-ABS-KEY ("machine learning" OR "artificial intelligence" OR "data mining") AND TITLE-ABS-KEY (challenges OR methods OR techniques) AND TITLE-ABS-KEY (crop OR agriculture)) AND PUBYEAR > 2014 AND PUBYEAR < 2025 AND (LIMIT-TO (OA , "all")) AND (LIMIT-TO (DOCTYPE , "ar") OR LIMIT-TO (DOCTYPE , "re")) AND (LIMIT-TO (LANGUAGE , "English"))

La cadena de búsqueda incluyó a los criterios de elegibilidad, los cuales están contemplados en la Tabla II y Tabla III.

TABLA II
CRITERIOS DE INCLUSIÓN

Ítem	Criterio
CI1	Publicaciones relacionadas con los modelos de Machine Learning efectivos para los patrones cuantitativos de cultivos.
CI2	Publicaciones relacionadas en gestión o producción agrícola.

Nota: Criterios implementados en la cadena de búsqueda utilizada.

TABLA III
CRITERIOS DE EXCLUSIÓN

Ítem	Criterio
CE1	Publicaciones anteriores al 2015
CE2	Publicaciones que no son de acceso abierto
CE3	Publicaciones que no son artículos originales y revisiones
CE4	Publicaciones distintas al idioma inglés

Nota: Criterios implementados en la cadena de búsqueda utilizada.

4th LACCEI International Multiconference on Entrepreneurship, Innovation and Regional Development - LEIRD 2024

“Creating solutions for a sustainable future: technology-based entrepreneurship” - Virtual Edition, December 2 – 4, 2024

Para un análisis más detallado, se utilizó el método PRISMA, proporcionando una documentación de manera transparente el propósito de la revisión, las acciones efectuadas y los resultados obtenidos [5].

La fecha y hora de la búsqueda y descarga de documentos fue el 24 de abril a las 4:30 pm. En la figura 1, se muestra que en la primera búsqueda se identificaron 121 documentos. De ellos, al pasar por el filtro de los últimos 10 años quedaron 117 artículos, luego se pasó por el filtro de tipo de artículo, idioma en inglés y acceso libre, de los cuales se excluyeron 15. De las siguientes 102 publicaciones recuperadas, 30 no se lograron descargar a texto completo. Entonces se tomó las 72 publicaciones elegidas, sin embargo, luego de su lectura, se descartaron 21 publicaciones por no cumplir con la relevancia del esquema PICO. Finalmente, 51 artículos fueron seleccionados para ser incluidos en el análisis.

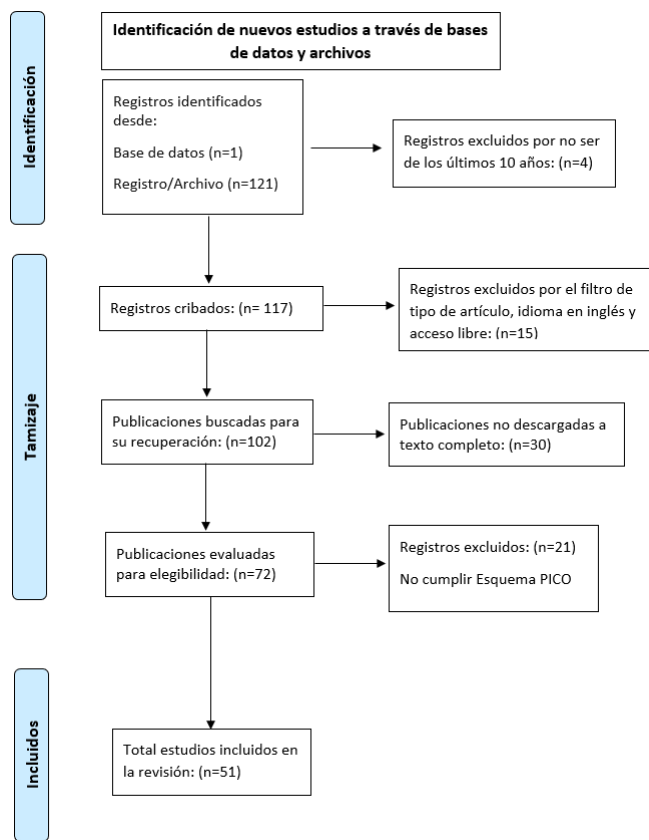


Fig. 1 Diagrama de flujo PRISMA

III. RESULTADOS

A. Bibliometría

La Fig.2 ilustra el número de documentos indizados anualmente desde 2015 hasta 2024. En 2015 y 2016, la cantidad de documentos publicados es limitada, con solo un

documento por año. Estos años marcan el inicio de la investigación en el campo, aunque de manera muy incipiente. En 2017 se observa un ligero aumento en las publicaciones, con 5 documentos, lo que sugiere un creciente reconocimiento y relevancia del tema, posiblemente impulsado por avances tecnológicos o descubrimientos clave. En los años 2018 y 2019 se registran 4 publicaciones por año, lo cual puede indicar una fase de estabilización en la investigación. En 2020, con 3 documentos, la ligera disminución podría estar influenciada por factores externos, como la pandemia de COVID-19, que afectó diversas actividades académicas y de investigación. En 2021 se observa un notable crecimiento con 13 documentos publicados, evidenciando una mayor producción de investigación en este campo. En 2022, con 12 publicaciones, aunque ligeramente inferior al año anterior, se muestra un compromiso sostenido con el área de investigación. En 2023 se registran 14 publicaciones, reafirmando la tendencia de crecimiento y sugiriendo una expansión continua y robusta del campo.

Finalmente, en 2024, aunque la cantidad de documentos disminuye a 8, esta reducción temporal puede deberse a que el año aún no ha finalizado, y se espera que el número de publicaciones aumente a medida que avance el año.

En conclusión, la tendencia general desde 2015 hasta 2024 muestra un crecimiento en el número de publicaciones, con un incremento más pronunciado a partir de 2021, lo que sugiere un interés creciente y sostenido en el área de modelos de Machine Learning para la identificación de patrones cuantitativos de sembríos.

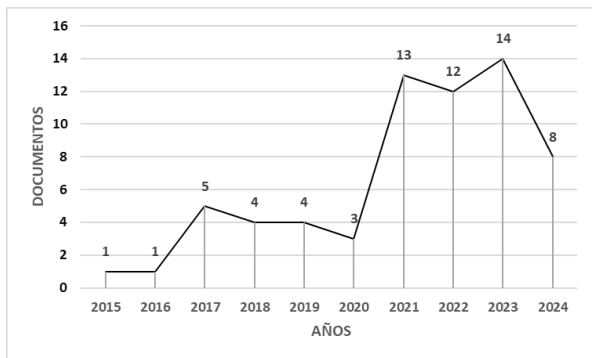


Fig.2 Documentos indizados por años.

La Fig. 3 muestra el número de documentos indizados por países desde 2015 hasta 2024. Estados Unidos, con 24 documentos, lidera con la mayor cantidad de publicaciones. La predominancia de Estados Unidos en el número de publicaciones refleja su liderazgo continuo y su compromiso significativo. China ocupa el segundo lugar con 15 documentos, lo que demuestra su creciente inversión en ciencia y tecnología y su rápida evolución como potencia en investigación. Australia y Canadá tienen una presencia

significativa con 6 publicaciones cada uno. Estos países cuentan con una robusta comunidad académica y de investigación, especialmente en áreas innovadoras y tecnológicas. Alemania, con 5 publicaciones, subraya su histórica fortaleza en investigación. India, con 4 publicaciones, muestra su creciente interés y capacidad en investigación. Brasil, con 3 publicaciones, indica un esfuerzo considerable en investigación, apoyando la comunidad científica en América Latina. Finalmente, Francia también cuenta con 3 publicaciones, lo que refleja su continua influencia y prestigio en el ámbito de la investigación europea.

En análisis refleja cómo diferentes países contribuyen al campo de investigación desde 2015 hasta 2024, mostrando una combinación de liderazgo, inversión y compromiso con el desarrollo científico.

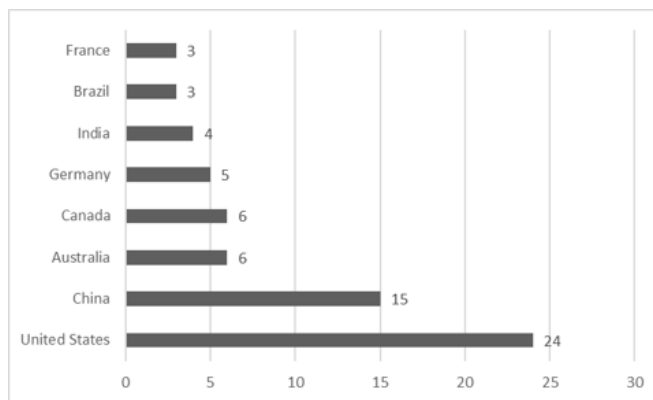


Fig. 3 Documentos indizados por país

La Tabla IV muestra que el área de Ciencias Agrícolas y Biológicas lidera con 25 documentos, reflejando un fuerte interés en la investigación relacionada con la agricultura y biología. En Bioquímica, Genética y Biología Molecular es prominente con 22 publicaciones, siendo campos fundamentales para entender la base molecular de la vida. Ciencias Ambientales cuenta con 13 publicaciones, indicando la urgencia global de abordar los desafíos ambientales. Química muestra 11 publicaciones, evidenciando la comprensión de fenómenos fundamentales en la ciencia. Ciencias de la Computación registran 9 publicaciones, reflejando avances en algoritmos, Inteligencia Artificial y aplicaciones informáticas que subrayan su impacto en la sociedad moderna. Ciencias de la Tierra y Planetarias tienen 8 publicaciones, destacando investigaciones desde la geología hasta la astrofísica. Medicina presenta 7 publicaciones, indicando investigaciones en prevención de enfermedades. Ingeniería, Multidisciplinaria y Ciencias Sociales muestran 5 publicaciones cada una, abordando la necesidad de enfoques

integrados para resolver problemas complejos y promover la innovación científica.

Finalmente, Ingeniería Química tiene 4 publicaciones, enfocándose en procesos y aplicaciones. En conclusión, muestra cómo diferentes áreas de investigación han contribuido desde 2015 hasta 2024, reflejando la diversidad y la importancia en avance científico específicamente con el análisis de cultivos usando modelos de innovación con Machine Learning.

TABLA IV
DOCUMENTO POR CADA ÁREA DE INVESTIGACIÓN

Área de Investigación	Documentos
Agricultural and Biological Sciences	25
Biochemistry, Genetics and Molecular Biology	22
Environmental Science	13
Chemistry	11
Computer Science	9
Earth and Planetary Sciences	8
Medicine	7
Engineering	5
Multidisciplinary	5
Social Sciences	5
Chemical Engineering	4

Nota: Elaborada a partir de los documentos analizados.

Las keywords “quantitative analysis”, “machine learning”, “plant breeding”, y “nonhuman” destacan por su alto nivel de ocurrencia en la literatura. Por otro lado, las keywords “agriculture”, “crop” y “data mining” muestran niveles más moderados de ocurrencia en los documentos analizados. Sin embargo, cabe destacar que, aunque “machine learning” y “quantitative analysis” presentan altos niveles de ocurrencia, su total link strength es menor en comparación con “plant breeding” y “nonhuman”. Lo cual significa que los enfoques de machine learning son variados en su campo de aplicación, siendo la agricultura y la mejora de cultivos solo una parte de ellos.

Finalmente, en la Tabla V se muestra la distribución de los documentos seleccionados por tipo.

TABLA V
CO-ORRRENCIA DE KEYWORDS

Keyword	Occurrences	Total link strength
quantitative analysis	31	152
Machine learning	29	168
plant breeding	26	184
nonhuman	20	157
agriculture	11	54
crop	10	79
data mining	9	59
genotype	8	61

Nota: Elaborada a partir de los documentos analizados.

B. Preguntas Específicas

Los resultados se basan en las preguntas específicas que abordan a la pregunta principal por lo tanto estos son ordenados considerando cada una de las preguntas.

Para el componente “P”: ¿Qué tipos de problemas de sembríos son estudiados a través de investigaciones de Machine Learning?

La Tabla VI muestra la distribución los documentos analizados, según el tipo de diseño de la investigación, experimental o teórico; relacionados con el análisis cuantitativo de sembríos, utilizando modelos de Machine Learning. El análisis se centra en los porcentajes de cada tipo de diseño entre los artículos totales y por categoría de problema de sembríos. De un total de 51 documentos, 41 utilizaron el método experimental (80.39 %) y 10 son revisiones bibliográficas (19.61 %).

El 33.33 % de los documentos analizados, estudiaron el tipo de problema "Fitosanitario". De estos, el 82.35% son del tipo experimental y el 17.65% son revisiones de literatura. También muestra prevalencia el tipo de problema "Factor abiótico", donde el 66.67 % de los documentos son del tipo experimental y el 33.33 % son revisiones de literatura.

Por lo tanto, respondiendo a la sub pregunta ¿Qué tipos de problemas de sembríos son estudiados a través de investigaciones de Machine Learning? Se puede concluir, que el problema de sembríos más común es Fitosanitario, con métodos experimentales de Machine Learning.

TABLA VI
PROBLEMAS DE CULTIVO CON MACHINE LEARNING

Count of Title	Tipo de diseño de artículo (Experimental o teórico)	
¿Qué tipos de problemas de sembríos existen? TIPO DE PROBLEMA	Experimental	Teórico
Fitosanitario	82.35%	17.65%
Factor abiótico	66.67%	33.33%

Climático	71.43%	28.57%
Edáfico y climático	60.00%	40.00%
Factor biótico	100.00%	0.00%
Factor biótico y abiótico	100.00%	0.00%
Fertilización	100.00%	0.00%

Nota: Elaborado a partir de la totalidad de los documentos analizados.

Para el componente “I”: ¿Qué modelos de Machine Learning existen para el análisis cuantitativo de sembríos?

La Tabla VII muestra la distribución de modelos de Machine Learning relacionados con el tipo de problema que presenta. De una muestra de 51 documentos, los modelos de Machine Learning más estudiados son Deep Learning junto a Random forest y Support Vector Machine (SVM), y además de manera independiente “Redes neuronales”. Luego, el modelo Random Forest es estudiado tanto de manera individual como de manera colectiva.

El 11.76 % de los documentos analizados, estudian la combinación de modelos “Deep Learning, Random Forest, Support Vector Machine (SVM)” de los cuales, el 50% se enfocan en el problema Fitosanitario. También involucra que el 11.76 % de los documentos analizados, estudiaron el modelo "Redes neuronales", con un enfoque equitativo en problemas Factor abiótico y Fitosanitario, cada uno representando el 33.33 %. Además, el 5.88 % de los documentos analizan los modelos "Deep Learning y Random Forest", demostrando que el 33.33 % son del tipo de problema Fitosanitario. Luego con el mismo porcentaje, 5.88 %, estudia el modelo “Random Forest” abordando problemas Climáticos, con un 33.33 %.

Por lo tanto, respondiendo a la sub pregunta ¿Qué modelos de Machine Learning existen para el análisis cuantitativo de sembríos? Se puede deducir, que los modelos de Machine Learning más estudiados son Deep Learning, Random Forest, y Support Vector Machine (SVM); siendo el tipo de problema de mayor aplicación, el Fitosanitario.

TABLA VII
MODELOS DE MACHINE LEARNING SEGÚN SU TIPO DE PROBLEMA

¿Qué modelos de Machine Learning existen para el análisis cuantitativo de sembríos?	Climático	Edáfico y climático	Factor abiótico	Factor biótico	Fitosanitario	Otros	Total
Deep Learning, Random Forest, SVM	0.00%	16.67%	16.67%	0.00%	50.00%	16.67%	11.76%
Redes neuronales	16.67%	0.00%	33.33%	0.00%	33.33%	16.67%	11.76%

Deep Learning, Random Forest	0.00%	0.00%	0.00%	0.00%	33.33%	16.67%	5.88%
Bosques aleatorios (Random Forest)	33.33%	0.00%	0.00%	0.00%	0.00%	16.67%	5.88%
Redes neuronales, PLSR	0.00%	0.00%	0.00%	16.67%	0.00%	16.67%	3.92%
Deep Learning, Modelos de regresión, Random Forest	0.00%	16.67%	0.00%	0.00%	16.67%	0.00%	3.92%
Random Forest, SVM	0.00%	0.00%	0.00%	0.00%	33.33%	0.00%	3.92%
Random forest, SVM, GLM	0.00%	0.00%	16.67%	0.00%	0.00%	16.67%	3.92%
Otros	16.67%	33.33%	50.00%	50.00%	83.33%	50.00%	33.33%
Total	13.73%	9.80%	17.65%	7.84%	33.33%	17.65%	100.00%

Nota: Elaborado a partir de la totalidad de los documentos analizados.

Para el componente “O”: ¿Qué nivel de efectividad muestran los casos de aplicación de modelos de Machine Learning en el análisis cuantitativo de sembríos?

La Tabla VIII muestra la distribución de modelos de Machine Learning relacionados con el nivel de efectividad. De una muestra de 51 artículos, los modelos más estudiados son Deep Learning, Random Forest, Support Vector Machine (SVM), presentan de acuerdo con los artículos, un nivel de efectividad entre el 90% al 95%.

Los modelos de Machine Learning que alcanzaron la mayor efectividad (95 %), destaca por el uso del modelo Redes Neuronales, así como Deep Learning combinado con Random Forest y Support Vector Machine (SVM). También con una efectividad media (90 %-92 %), incluye Deep Learning y Random Forest, también modelos de regresión junto a Random Forest, y combinaciones de Redes neuronales con Support Vector Machine SVM. Además, el modelo Redes Neuronales alcanza un nivel de efectividad del 93 %. Luego con efectividad moderada del 90 %, incluye modelos de regresión con XGBoost y modelos predictivos de árboles como Random Forest y Gradient Boosting.

Finalmente, con un nivel de efectividad más baja (85 %-88 %), combinaciones de Deep Learning con k-means, Deep Learning, Modelos de Regresión con Random Forest, asimismo, enfoques multidimensionales y modelos de predicción genética con el 86 %, y modelos de predicción genómica como BLUP y GBLUP con el nivel de efectividad del 86 %.

Por lo tanto, respecto a la tercera sub pregunta: ¿Qué nivel de efectividad muestran los casos de aplicación de modelos de Machine Learning en el análisis cuantitativo de sembríos? Se concluye, que este estudio demuestra que los modelos de Machine Learning más utilizados son (Deep Learning, Random Forest y Support Vector Machine (SVM) con un nivel de efectividad del 90 %.

TABLA VIII
MODELOS DE MACHINE LEARNING SEGÚN SU TIPO DE PROBLEMA

¿Qué modelos de Machine Learning existen para el análisis cuantitativo de sembríos?	95%	93%	92%	90%	88%	87%	86%	85%	Total
Redes neuronales	3	1		2	0				6
Deep Learning, Random Forest, SVM	2		1	2	1				6
Redes neuronales, PLSR	2				0				2
Bosques aleatorios (Random Forest)	1		1		1				3
PLSR Y SVM	1			1	0				2
Máquina de vectores de soporte (SVM)	1				0				1
Método de aprendizaje profundo basado en Transformer.	1				0				1
Modelo de regresión	1				0		1		2
Redes neuronales convolucionales (CNN)	1				0				1
Deep Learning, Random Forest			1	2	0				3
Total	13	1	7	15	6	1	5	3	51

Nota: Elaborado a partir de la totalidad de los documentos analizados.

La Tabla IX muestra la distribución de referencias por tipo de modelo de Machine Learning aplicado en análisis cuantitativo de cultivos. Se analiza cuántos artículos utilizaron cada tipo de modelo, lo que permite ver cuáles son los comúnmente investigados y utilizados.

TABLA IX
DISTRIBUCIÓN REFERENCIAS POR TIPO DE MODELO

Referencia	Tipo Modelo	No. de Papers
[6],[7],[8],[9],[10],[11]	Redes Neuronales	6
[12],[13],[14],[15],[16],[17]	Deep Learning, Random Forest, SVM	6
[18],[19]	Redes neuronales, PLSR	2

[20],[21],[22]	Bosques aleatorios (Random Forest)	3
[23], [24]	PLSR Y SVM	2
[25]	Máquina de vectores de soporte (SVM)	1
[26]	Método de aprendizaje profundo basado en Transformer.	1
[27],[28]	Modelo de regresión	2
[29]	Redes neuronales convolucionales (CNN)	1
[30],[31],[32]	Deep Learning, Random Forest	3
[33],[34]	Modelos de regresión, Random Forest	2
[35],[36]	Redes neuronales, SVM	2
[37]	Deep learning LSTM	1
[38],[39],[40],[41], [42] [43],[44],[45]	Modelos de Regresión	8
[46],[47],[48],[49],[50] [51],[52],[52],[54],[55]	Algoritmos genéticos de optimización de monitoreo	10

Nota: Elaborada a partir de los documentos analizados.

El uso de algoritmos de Machine Learning en la agricultura ha mostrado un potencial significativo para optimizar la producción y mejorar la gestión de cultivos. Sin embargo, un análisis comparativo exhaustivo de estos modelos, considerando su rendimiento y aplicabilidad en diferentes escenarios agrícolas, es crucial. Los estudios existentes, muestran que el uso de modelos como Random Forest y Support Vector Machines ha resultado exitoso en la predicción de variables críticas, como el rendimiento de cultivos [7].

Esta evidencia sugiere que estos algoritmos pueden ser aplicados eficazmente en la agricultura, pero es necesario profundizar en el porqué de su eficacia en ciertos contextos. La investigación debe enfocarse en una comparación más detallada de por qué algunos algoritmos, como el Deep Learning, superan a otros en tareas específicas.

Por ejemplo, el estudio de Pimenta et al. (2021) resalta la integración de datos genómicos y ambientales para identificar genes asociados con la resistencia de afecciones, lo que podría ser un aspecto por considerar en el desarrollo de modelos predictivos más precisos [5].

Además, es vital que el análisis no solo se limite al rendimiento técnico de los algoritmos, sino que también se

integre un enfoque hacia la sostenibilidad. La investigación de He et al. (2022) sobre el uso de Machine Learning para mejorar la distribución de gotas en la protección de cultivos puede inspirar nuevas metodologías que consideren el impacto ambiental de las prácticas agrícolas, fomentando un uso más responsable de los recursos.

Finalmente, la necesidad de colaboración entre distintas disciplinas se hace evidente. Incorporar conocimientos de biología, ecología y tecnología puede enriquecer el análisis de los modelos de Machine Learning aplicados a la agricultura, como lo sugiere el trabajo de Ospina-Hernández et al. (2022) en el contexto educativo. Esta sinergia podría llevar a la creación de modelos más robustos y adaptativos que respondan a las complejidades del entorno agrícola.

IV. CONCLUSIONES

El problema de sembríos más comúnmente abordado a través de investigaciones de Machine Learning es el fitosanitario. Los estudios han utilizado métodos experimentales de Machine Learning, y este enfoque representa un 33.33% de los artículos analizados. Esto indica que la aplicación de Machine Learning en la gestión de la salud de los cultivos es un área de gran interés y relevancia.

Los modelos de Machine Learning más estudiados para el análisis cuantitativo de sembríos son Deep Learning, Random Forest y Support Vector Machine (SVM). Estos modelos han sido aplicados principalmente en problemas fitosanitarios, demostrando su capacidad para manejar y analizar grandes conjuntos de datos y extraer patrones útiles para la agricultura.

Respecto a la efectividad de los modelos de Machine Learning en el análisis cuantitativo de sembríos, se concluye que Deep Learning, Random Forest y Support Vector Machine (SVM) son los modelos más efectivos, alcanzando un nivel de efectividad del 90% al 95%. Esto resalta la robustez y precisión de estos modelos en aplicaciones agrícolas.

Finalmente, esta Revisión Sistemática de Literatura (RSL) demuestra que los modelos de Machine Learning, especialmente Deep Learning, Random Forest y SVM, son herramientas valiosas para el análisis cuantitativo de cultivos. Su aplicación no solo mejora la precisión del análisis y la toma de decisiones en la agricultura, sino que también promueve la innovación y el emprendimiento en el sector agrícola. La metodología PICO utilizada en esta revisión ha permitido identificar y evaluar de manera sistemática los modelos más efectivos, proporcionando una guía clara para futuras investigaciones e inversiones en tecnología agrícola.

REFERENCIAS

[1] J. D. Rojo Hernández, Ó. J. Mesa, y U. Lall, «ENSO Dynamics, Trends, and Prediction Using Machine Learning», *Weather and Forecasting*, vol. 35, n.º 5, pp. 2061-2081, oct. 2020, doi: 10.1175/WAF-D-20-0031.1.

[2] C. Ospina-Hernández, Y. F. Ceballos, J. Moreno-Cadavid, C. Ospina-Hernández, Y. F. Ceballos, y J. Moreno-Cadavid, «Revisión Sistemática de Literatura (RSL) sobre la aplicación de los juegos serios en cursos de ciencias básicas para la modalidad virtual como estrategia para mejorar el índice de permanencia estudiantil», *DYNA*, vol. 89, n.º SPE222, pp. 136-144, sep. 2022, doi: 10.15446/dyna.v89n222.101834.

[3] Universidad Loyola, «Estrategia PICO», Estrategia de búsqueda PICO. Accedido: 11 de junio de 2024. [En línea]. Disponible en: <https://www.loyola.es/investigacion/biblioteca/biblioguias/biblioguia-estrategia-pico>

[4] UNESCO, «Construyendo la Paz a través de la Educación, la Ciencia y la Cultura, la comunicación y la información», UNESCO. Accedido: 4 de junio de 2024. [En línea]. Disponible en: <https://www.unesco.org/es>

[5] R. J. G. Pimenta et al., «Genome-wide approaches for the identification of markers and genes associated with sugarcane yellow leaf virus resistance», *Sci Rep*, vol. 11, n.º 1, p. 15730, ago. 2021, doi: 10.1038/s41598-021-95116-1.

[6] T. Xu, F. Wang, Q. Yi, L. Xie, y X. Yao, «A Bibliometric and Visualized Analysis of Research Progress and Trends in Rice Remote Sensing over the Past 42 Years (1980–2021)», *Remote Sensing*, vol. 14, n.º 15, p. 3607, jul. 2022, doi: 10.3390/rs14153607.

[7] E. K. Juarez y M. R. Petersen, «A Comparison of Machine Learning Methods to Forecast Tropospheric Ozone Levels in Delhi», *Atmosphere*, vol. 13, n.º 1, p. 46, dic. 2021, doi: 10.3390/atmos13010046.

[8] Z. Zhu et al., «A Long Short-Term Memory Neural Network Based Simultaneous Quantitative Analysis of Multiple Tobacco Chemical Components by Near-Infrared Hyperspectroscopy Images», *Chemosensors*, vol. 10, n.º 5, p. 164, abr. 2022, doi: 10.3390/chemosensors10050164.

[9] Z. Zununjan, M. A. Turghan, M. Sattar, N. Kasim, B. Emin, y A. Abliz, «Combining the fractional order derivative and machine learning for leaf water content estimation of spring wheat using hyper-spectral indices», *Plant Methods*, vol. 20, n.º 1, p. 97, jun. 2024, doi: 10.1186/s13007-024-01224-0.

[10] M. Yoosefzadeh-Najafabadi, S. Torabi, D. Tulpan, I. Rajcan, y M. Eskandari, «Application of SVR-Mediated GWAS for Identification of Durable Genetic Regions Associated with Soybean Seed Quality Traits», *Plants*, vol. 12, n.º 14, p. 2659, jul. 2023, doi: 10.3390/plants12142659.

[11] S. Maleki, N. Baghdadi, S. Najem, C. F. Dantas, H. Bazzi, y D. Ienco, «Determining Effective Temporal Windows for Rapeseed Detection Using Sentinel-1 Time Series and Machine Learning Algorithms», *Remote Sensing*, vol. 16, n.º 3, p. 549, ene. 2024, doi: 10.3390/rs16030549.

[12] H. Zhu y Q. Wu, «Artificial-Intelligence-Enhanced Study on the Optimization of the Responsibility and Compensation Mechanism for Provincial Cultivated Land Retention from a Fairness Perspective», *Land*, vol. 12, n.º 12, p. 2118, nov. 2023, doi: 10.3390/land12122118.

[13] L. Shumilo, A. Okhrimenko, N. Kussul, S. Drozd, y O. Shkalikov, «Generative adversarial network augmentation for solving the training data imbalance problem in crop classification», *Remote Sensing Letters*, vol. 14, n.º 11, Art. n.º 11, nov. 2023, doi: 10.1080/2150704X.2023.2275551.

[14] I. Kecoglu, M. Sirkeci, M. B. Unlu, A. Sen, U. Parlattan, y F. Guzelcimen, «Quantification of salt stress in wheat leaves by Raman spectroscopy and machine learning», *Sci Rep*, vol. 12, n.º 1, p. 7197, may 2022, doi: 10.1038/s41598-022-10767-y.

[15] D. Singh et al., «Fab Advances in Fabaceae for Abiotic Stress Resilience: From 'Omics' to Artificial Intelligence», *IJMS*, vol. 22, n.º 19, p. 10535, sep. 2021, doi: 10.3390/ijms221910535.

[16] A. J. DeSalvio, A. Adak, S. C. Murray, S. C. Wilde, y T. Isakeit, «Phenomic data-facilitated rust and senescence prediction in maize using machine learning algorithms», *Sci Rep*, vol. 12, n.º 1, p. 7571, may 2022, doi: 10.1038/s41598-022-11591-0.

[17] D. Vincent et al., «Mining the Wheat Grain Proteome», *IJMS*, vol. 23, n.º 2, p. 713, ene. 2022, doi: 10.3390/ijms23020713.

[18] J. Ou, Z. Wu, Q. Yan, X. Feng, y Z. Zhao, «Improving soil organic carbon mapping in farmlands using machine learning models and complex cropping system information», *Environ Sci Eur*, vol. 36, n.º 1, p. 80, abr. 2024, doi: 10.1186/s12302-024-00912-x.

- [19] V. Sejian, C. G. Shashank, M. V. Silpa, A. P. Madhusoodan, C. Devaraj, y S. Koenig, «Non-Invasive Methods of Quantifying Heat Stress Response in Farm Animals with Special Reference to Dairy Cattle», *Atmosphere*, vol. 13, n.º 10, p. 1642, oct. 2022, doi: 10.3390/atmos13101642.
- [20] K. V. Prasad, H. Vaidya, C. Rajashekhar, K. S. Karekal, R. Sali, y K. S. Nisar, «Multiclass Classification of diseased grape leaf identification using deep convolutional neural network(DCNN) classifier», *Sci Rep*, vol. 14, n.º 1, p. 9002, abr. 2024, doi: 10.1038/s41598-024-59562-x.
- [21] Y. He, J. Wu, H. Fu, Z. Sun, H. Fang, y W. Wang, «Quantitative Analysis of Droplet Size Distribution in Plant Protection Spray Based on Machine Learning Method», *Water*, vol. 14, n.º 2, p. 175, ene. 2022, doi: 10.3390/w14020175.
- [22] H. Liu *et al.*, «Driving Force Analysis of Natural Wetland in Northeast Plain Based on SSA-XGBoost Model», *Sensors*, vol. 23, n.º 17, p. 7513, ago. 2023, doi: 10.3390/s23177513.
- [23] S. Manickam, V. R. Rajagopalan, R. Kambale, R. Rajasekaran, S. Kanagarajan, y R. Muthurajan, «Plant Metabolomics: Current Initiatives and Future Prospects», *CIMB*, vol. 45, n.º 11, pp. 8894-8906, nov. 2023, doi: 10.3390/cimb45110558.
- [24] A. Gupta, T. Sharma, S. P. Singh, A. Bhardwaj, D. Srivastava, y R. Kumar, «Prospects of microgreens as budding living functional food: Breeding and biofortification through OMICS and other approaches for nutritional security», *Front. Genet.*, vol. 14, p. 1053810, ene. 2023, doi: 10.3389/fgene.2023.1053810.
- [25] D. Jeon *et al.*, «Digitalizing breeding in plants: A new trend of next-generation breeding based on genomic prediction», *Front. Plant Sci.*, vol. 14, p. 1092584, ene. 2023, doi: 10.3389/fpls.2023.1092584.
- [26] C. Y. Hamany Djande, F. Tugizimana, P. A. Steenkamp, L. A. Piater, y I. A. Dubery, «Metabolomic Reconfiguration in Primed Barley (*Hordeum vulgare*) Plants in Response to *Pyrenophora teres f. teres* Infection», *Metabolites*, vol. 13, n.º 9, p. 997, sep. 2023, doi: 10.3390/metabo13090997.
- [27] H. Liu *et al.*, «Driving Force Analysis of Natural Wetland in Northeast Plain Based on SSA-XGBoost Model», *Sensors*, vol. 23, n.º 17, p. 7513, ago. 2023, doi: 10.3390/s23177513.
- [28] C. Wang *et al.*, «Deep learning based high-throughput phenotyping of chalkiness in rice exposed to high night temperature», *Plant Methods*, vol. 18, n.º 1, p. 9, dic. 2022, doi: 10.1186/s13007-022-00839-5.
- [29] R. R. Mir, H. Kudapa, S. Pramod, y R. S. Lewis, «Editorial: Biotechnological and genomic approaches for enhancing agronomic performance of crops», *Front. Genet.*, vol. 13, p. 991630, oct. 2022, doi: 10.3389/fgene.2022.991630.
- [30] K. Yee, U. Tantipongpipat, y S. Mishra, «Image Cropping on Twitter: Fairness Metrics, their Limitations, and the Importance of Representation, Design, and Agency», *Proc. ACM Hum.-Comput. Interact.*, vol. 5, n.º CSCW2, pp. 1-24, oct. 2021, doi: 10.1145/3479594.
- [31] W. Tang *et al.*, «Improving the performance of a spectral model to estimate total nitrogen content with small soil samples sizes», *Chem. Biol. Technol. Agric.*, vol. 11, n.º 1, p. 33, mar. 2024, doi: 10.1186/s40538-024-00552-6.
- [32] «Recreational Water Quality Criteria».
- [33] S. Chen, «Advances in Molecular Breeding of Forage Crops: Technologies, Applications and Prospects», *Agriculture*, vol. 14, n.º 2, p. 279, feb. 2024, doi: 10.3390/agriculture14020279.
- [34] M. Yoosefzadeh-Najafabadi, D. Tulpan, y M. Eskandari, «Application of machine learning and genetic optimization algorithms for modeling and optimizing soybean yield using its component traits», *PLoS ONE*, vol. 16, n.º 4, p. e0250665, abr. 2021, doi: 10.1371/journal.pone.0250665.
- [35] U. C. Jha, K. D. Sharma, H. Nayyar, S. K. Parida, y K. H. M. Siddique, «Breeding and Genomics Interventions for Developing Ascochyta Blight Resistant Grain Legumes», *IJMS*, vol. 23, n.º 4, p. 2217, feb. 2022, doi: 10.3390/ijms23042217.
- [36] Z. Zununjan, M. A. Turghan, M. Sattar, N. Kasim, B. Emin, y A. Abliz, «Combining the fractional order derivative and machine learning for leaf water content estimation of spring wheat using hyper-spectral indices», *Plant Methods*, vol. 20, n.º 1, p. 97, jun. 2024, doi: 10.1186/s13007-024-01224-0.
- [37] L. Liu *et al.*, «Knowledge-guided machine learning can improve carbon cycle quantification in agroecosystems», *Nat Commun*, vol. 15, n.º 1, p. 357, ene. 2024, doi: 10.1038/s41467-023-43860-5.
- [38] D. Taliadoros *et al.*, «Emergence and spread of the barley net blotch pathogen coincided with crop domestication and cultivation history», *PLoS Genet.*, vol. 20, n.º 1, p. e1010884, ene. 2024, doi: 10.1371/journal.pgen.1010884.
- [39] F. Strobl, M. F. Schetelig, y E. H. K. Stelzer, «In toto light sheet fluorescence microscopy live imaging datasets of *Ceratitis capitata* embryonic development», *Sci Data*, vol. 9, n.º 1, p. 340, jun. 2022, doi: 10.1038/s41597-022-01443-x.
- [40] S. Jubair *et al.*, «GPTransformer: A Transformer-Based Deep Learning Method for Predicting Fusarium Related Traits in Barley», *Front. Plant Sci.*, vol. 12, p. 761402, dic. 2021, doi: 10.3389/fpls.2021.761402.
- [41] X. Wang, X. Lu, X. Zhang, y Z. Yang, «Inversion of the hybrid machine learning model to estimate leaf area index of winter wheat from GaoFen-6 WFV imagery», *Geocarto International*, vol. 37, n.º 27, Art. n.º 27, feb. 2024, doi: 10.1080/10106049.2022.2108906.
- [42] F. E. Nasir, M. Tufail, M. Haris, J. Iqbal, S. Khan, y M. T. Khan, «Precision agricultural robotic sprayer with real-time Tobacco recognition and spraying system based on deep learning», *PLoS ONE*, vol. 18, n.º 3, p. e0283801, mar. 2023, doi: 10.1371/journal.pone.0283801.
- [43] Z. Zununjan, M. A. Turghan, M. Sattar, N. Kasim, B. Emin, y A. Abliz, «Combining the fractional order derivative and machine learning for leaf water content estimation of spring wheat using hyper-spectral indices», *Plant Methods*, vol. 20, n.º 1, p. 97, jun. 2024, doi: 10.1186/s13007-024-01224-0.
- [44] F. Shahzad *et al.*, «Comparing machine learning algorithms to predict vegetation fire detections in Pakistan», *fire ecol*, vol. 20, n.º 1, p. 57, jun. 2024, doi: 10.1186/s42408-024-00289-5.
- [45] E. Rabiayan, R. Darvishzadeh, y H. Alipour, «Correction: Identification and estimation of lodging in bread wheat genotypes using machine learning predictive algorithms», *Plant Methods*, vol. 20, n.º 1, p. 86, jun. 2024, doi: 10.1186/s13007-024-01203-5.
- [46] L. Drees *et al.*, «Data-driven crop growth simulation on time-varying generated images using multi-conditional generative adversarial networks», *Plant Methods*, vol. 20, n.º 1, p. 93, jun. 2024, doi: 10.1186/s13007-024-01205-3.
- [47] L. Shumilo, A. Okhrimenko, N. Kussul, S. Drozd, y O. Shkalikov, «Generative adversarial network augmentation for solving the training data imbalance problem in crop classification», *Remote Sensing Letters*, vol. 14, n.º 11, pp. 1129-1138, nov. 2023, doi: 10.1080/2150704X.2023.2275551.
- [48] M. Hagen, R. Dass, C. Westhues, J. Blom, S. J. Schultheiss, y S. Patz, «Interpretable machine learning decodes soil microbiome's response to drought stress», *Environmental Microbiome*, vol. 19, n.º 1, p. 35, may 2024, doi: 10.1186/s40793-024-00578-1.
- [49] S. M. H. Shah *et al.*, «New strategy based on Hammerstein-Wiener and supervised machine learning for identification of treated wastewater salinization in Al-Hassa region, Saudi Arabia», *Environ Sci Eur*, vol. 36, n.º 1, p. 114, jun. 2024, doi: 10.1186/s12302-024-00914-9.
- [50] B. Chen, Y. Tu, J. An, S. Wu, C. Lin, y P. Gong, «Quantification of losses in agriculture production in eastern Ukraine due to the Russia-Ukraine war», *Commun Earth Environ*, vol. 5, n.º 1, p. 336, jun. 2024, doi: 10.1038/s43247-024-01488-3.
- [51] Y. Wang *et al.*, «A hybrid approach for rice crop disease detection in agricultural IoT system», *Discov Sustain*, vol. 5, n.º 1, p. 99, may 2024, doi: 10.1007/s43621-024-00285-4.
- [52] D. Ibrar *et al.*, «Application of machine learning for identification of heterotic groups in sunflower through combined approach of phenotyping, genotyping and protein profiling», *Sci Rep*, vol. 14, n.º 1, p. 7333, mar. 2024, doi: 10.1038/s41598-024-58049-z.
- [53] O. S. Nagesh *et al.*, «Boosting enabled efficient machine learning technique for accurate prediction of crop yield towards precision agriculture», *Discov Sustain*, vol. 5, n.º 1, p. 78, abr. 2024, doi: 10.1007/s43621-024-00254-x.

- [54] H. Li, N. Fotouhi, F. Liu, H. Ji, y Q. Wu, «Early detection of dark-affected plant mechanical responses using enhanced electrical signals», *Plant Methods*, vol. 20, n.º 1, p. 49, mar. 2024, doi: 10.1186/s13007-024-01169-4.
- [55] A. E. Cleves, A. N. Jain, D. A. Demeter, Z. A. Buchan, J. Wilmot, y E. N. Hancock, «From UK-2A to florylpicoxamid: Active learning to identify a mimic of a macrocyclic natural product», *J Comput Aided Mol Des*, vol. 38, n.º 1, p. 19, dic. 2024, doi: 10.1007/s10822-024-00555-3.