

Estimación de propiedades de un suelo mediante teledetección y aprendizaje automático

Estimation of Soil Properties by Remote Sensing and Machine Learning

 Roberto Alejandro García-Reyes^{I*},  María Elena Ruíz-Pérez^{II},
 Sergio Rodríguez-Rodríguez^{III} and  Juan Alejandro Villazón-Gómez^{IV}

^IMinisterio de la Agricultura, Departamento Provincial de Suelos y Fertilizantes, Holguín, Cuba.

^{II}Universidad Agraria de La Habana, Facultad de Ciencias Técnicas, San José de las Lajas, Mayabeque, Cuba.

^{III}Universidad de Granma, Facultad de Ciencias Agrícolas, Centro de Estudios de Biotecnología Vegetal, Bayamo, Granma, Cuba.

^{IV}Universidad de Holguín. Departamento de Ciencias Agropecuarias, Holguín, Cuba.

*Autor para correspondencia: Roberto Alejandro García-Reyes, e-mail: ralejandro9409@gmail.com

RESUMEN: La aptitud del suelo para un cultivo, se puede determinar mediante indicadores de calidad, que pueden ser obtenidos por diferentes métodos. La investigación tuvo como objetivo la estimación de propiedades de un Vertisuelo dedicado al cultivo del arroz mediante teledetección y aprendizaje automático. Se determinaron “*in situ*” propiedades químicas y físicas del suelo que inciden en la fisiología del arroz, con un muestreo completamente aleatorizado de 100 puntos a una profundidad de 0 m a 0,20 m. Las propiedades se estimaron a partir del Índice de Vegetación por Diferencia Normalizada (NDVI) mediante herramientas de aprendizaje automático. La materia orgánica, el calcio, el magnesio y el fósforo, se pueden estimar mediante regresión lineal con el NDVI, teniendo en cuenta que tuvieron Errores Estándares de 0,27; 3,41; 4,12 y 1,68, respectivamente, y los coeficientes de determinación cercanos a 1. La técnica de bosques aleatorios mostró el mejor rendimiento, con valores en su determinación cercanos a 1 y un error en su estimación y validación cercanos a 0.

Palabras clave: índice de vegetación por diferencia normalizada, vertisol.

ABSTRACT: Soil suitability for a crop can be determined by quality indicators, which can be obtained by different methods. The objective of this research was to estimate the properties of a vertisol dedicated to rice cultivation by means of remote sensing and automatic learning. Chemical and physical properties of the soil that affect rice physiology were determined “*in situ*”, with a completely randomized sampling of 100 points at a depth of 0 m to 0.20 m. Properties were estimated from the Normalized Difference Vegetation Index (NDVI) using machine learning tools. Organic matter, calcium, magnesium and phosphorus can be estimated by linear regression with NDVI, considering that they had Standard Errors of 0.27, 3.41, 4.12 and 1.68, respectively, and coefficients of determination close to 1. The random forest technique showed the best performance, with values in its determination close to 1 and an error in its estimation and validation close to 0.

Keywords: Normalized Difference Vegetation Index, Vertisol.

INTRODUCCIÓN

El cultivo del arroz tiene alta demanda para la alimentación mundial y su producción alcanza más de 700 millones de toneladas. Cuba es una de las naciones altamente consumidoras de este cereal, que asciende a 80,38 kg persona⁻¹ año⁻¹ (Del Valle *et al.*, 2022). En el país se dedican al cultivo un total de 16 847 hectáreas para una producción de 266 596 toneladas. La producción del cultivo es de 3,35 t.ha⁻¹ y se encuentra en producciones de empresas estatales, cooperativas y el sector particular (Casanovas *et al.*, 2022).

En la provincia Holguín, Cuba, debido a la construcción del trasvase Este-Oeste se dispone de agua suficiente para dedicarlas al cultivo del arroz (*Oryza sativa* L.). En este contexto, es importante considerar que el crecimiento de las plantas de arroz depende de las condiciones físicas y químicas del suelo, que afectan la capacidad del sistema radicular del cultivo para crecer eficientemente. Operaciones agrícolas como la preparación del terreno para el cultivo, labranza, la fertilización, el manejo del riego y los métodos de plantación, alteran las propiedades del suelo a corto y largo plazo, impactando en la sustentabilidad como en el rendimiento del cultivo (Baroudy *et al.*, 2020).

Recibido: 31/07/2024

Aceptado: 24/02/2025

Los autores de este trabajo declaran no presentar conflicto de intereses

CONTRIBUCIONES DE AUTOR: **Conceptualización:** Roberto Alejandro García-Reyes. **Curación de datos:** Roberto Alejandro García-Reyes, María Elena Ruíz-Pérez y Sergio Rodríguez-Rodríguez. **Investigación:** Roberto Alejandro García-Reyes, María Elena Ruíz Pérez, Juan Alejandro Villazón-Gómez. **Metodología:** Roberto Alejandro García-Reyes, María Elena Ruíz-Pérez y Sergio Rodríguez-Rodríguez. **Supervisión:** María Elena Ruíz-Pérez y Sergio Rodríguez-Rodríguez. **Validación:** María Elena Ruíz Pérez y Sergio Rodríguez Rodríguez, Juan Alejandro Villazón-Gómez. **Papeles/Redacción, proyecto original:** Roberto Alejandro García-Reyes. **Redacción, revisión y edición:** Roberto Alejandro García-Reyes, María Elena Ruíz-Pérez, Sergio Rodríguez-Rodríguez, Juan Alejandro Villazón-Gómez.



Este artículo se encuentra bajo los términos de la licencia Creative Commons Attribution-NonCommercial (CC BY-NC 4.0).
<https://creativecommons.org/licenses/by-nc/4.0/>



Dentro de los métodos más utilizados para evaluar el estado de los suelos dedicados a este cultivo, se encuentra el planteado por Baroudy *et al.* (2020) donde utilizan imágenes procedentes de satélites y determinan índices espectrales de suelo y cultivo, los cuales son correlacionados con información *in situ*. También Singh *et al.* (2024) plantean algoritmos de aprendizaje automático basado en modelos matemáticos para estimar propiedades mediante información *in situ* y de teledetección. Ambas investigaciones han utilizado distintos métodos de regresión que pueden ser usados para construir modelos que permitan estimar propiedades con un menor costo y brindar una información rápida en el tiempo (Siqueira *et al.*, 2024).

Los modelos que se derivan de la utilización del aprendizaje automático más empleados en la literatura, destacan las redes neuronales, bosques aleatorios y vectores de regresión entre otros (Ließ *et al.*, 2016). No obstante, en Cuba no se han publicado estudios que empleen estas herramientas para evaluar la aptitud de los suelos para el cultivo del arroz. De ahí las razones que hacen posible el uso de la teledetección espacial Luque (2023), como su capacidad para discriminar grandes áreas que tienen diferentes características en su composición físico-química de la superficie expuesta sobre el suelo. Por lo antes expuesto, el objetivo de la investigación es la estimación de propiedades de un suelo dedicado al cultivo del arroz mediante teledetección y herramientas del aprendizaje automático.

MATERIALES Y MÉTODOS

El área seleccionada pertenece a la Empresa Agropecuaria Guatemala, CCS “Tomás Machado” del poblado Cosme Herrera, ubicada en los 20°44'54,601"N y 75°50'43,743"W del municipio Mayarí, provincia Holguín (Figura 1).

En la zona de estudio según datos de la Estación Meteorológica de Guaro, ubicada en los 20°40'21"N y 75°46'57" W en el municipio Mayarí a 20,96 msnm, las precipitación anual es de 1 067,6 mm y la temperatura media es de 25,6 °C según los estudios realizado por Villazón *et al.* (2023).



FIGURA 1. Ubicación del área donde se realizó la investigación, perteneciente a la CCS “Tomás Machado”, poblado Cosme Herrera, Mayarí, Holguín, Cuba. **Fuente:** SAS PlanetNightly 200718.10081 (<http://www.geojamal.com>).

En cuanto a los datos meteorológicos desde el inicio del mes de abril hasta el día 26 de mayo de 2022, fecha en la que se realizó el muestreo, el total de precipitaciones fue de 168,7 mm, con una temperatura media de 25, 4 °C y una humedad relativa media de 73,8 %. El suelo característico del área es del tipo Vertisol Crómico según Hernández *et al.* (2015) con una pendiente < 2 % por lo que puede considerarse plana. En el área de 100 ha, se realizó un muestreo sistemático en 100 puntos georeferenciados con un GPS con apreciación de 3 m, a una distancia entre puntos de 100 m.

Las muestras fueron tomadas con una barrena para análisis agroquímicos en el rango de profundidad entre 0 m a 0,20 m por ser en esta profundidad donde se encuentra el mayor contenido de raíces del cultivo del arroz capaces de absorber el agua y los elementos nutritivos necesarios para su crecimiento y desarrollo (Angladette, 1969).

Las propiedades del suelo seleccionadas como indicadores de calidad para un suelo cultivado con arroz se muestran en la Tabla 1. Los métodos seguidos para su selección aparecen en García *et al.* (2025). Todas las propiedades fueron determinadas según las normas cubanas vigentes para la determinación de propiedades químicas de los suelos en la red Nacional de Laboratorios de Suelos del país.

TABLA 1. Determinación de propiedades del suelo más influyentes la calidad del suelo para el cultivo del arroz

Nombre de la propiedad del suelo	Propiedades analizadas	Unidad de medida	Técnica analítica empleada
pH en agua	pH _{H2O}	unidades	(NC 2001.2015)
Fósforo asimilable	P ₂ O ₅	mg kg ⁻¹	(NC 52.1999)
Potasio asimilable	K ₂ O	mg kg ⁻¹	
Nitrógeno total	Nt	%	(NC 11261: 2009)
Materia orgánica	MO	%	(NC 1043.2014)
Calcio	Ca	cmol kg ⁻¹	(NC 209:2002)
Magnesio asimilable	Mg	cmol kg ⁻¹	(NC 209:2002)
Sodio asimilable	Na	cmol kg ⁻¹	
Conductividad eléctrica	CE	dS m ⁻¹	(NC 776: 2010)

NC: Norma Cubana

Para la estimación de las propiedades seleccionadas, se determinó el Índice de Vegetación por Diferencia Normalizada (NDVI) del área de estudio, a partir de la imagen del día 26 de abril de 2022, perteneciente al satélite Landsat 9 OLI/TIRS 2 (LC09_L2SP_011046_20220426_20220428_02_T1) del Servicio Geológico de los Estados Unidos en el path 011 row 046 y fueron proyectados en el Sistema WGS 84 UTM Zona 18 Norte en el software QGIS 3.10 “A Coruña”.

Para la determinación del NDVI se utilizó la ecuación 1 según Rouse Jr *et al.* (1974), luego de realizar la corrección atmosférica para eliminar el efecto de las nubes sobre la imagen.

$$NDVI = \frac{(B_{NIR} - B_{red})}{(B_{NIR} + B_{red})} \quad (1)$$

donde: B_{NIR} es la banda infrarroja y B_{red} la banda roja del sensor.

Dado el objetivo de la investigación, se utiliza un enfoque propuesto por Choudhury & Mandal, (2021) el cual consiste en la construcción de modelos para estimar propiedades del suelo a partir de mapas NDVI.

Para la estimación de las propiedades del suelo se utilizaron herramientas de aprendizaje automático, las cuales fueron: regresión lineal simple Khanal *et al.* (2018), Random forest (RF) o bosques aleatorios Park *et al.* (2024), Máquina de Vectores de Soporte para la regresión Shrestha & Shukla (2015), Descenso de Gradiente Estocástico (SGD por sus siglas en inglés) Nisbet *et al.* (2009) y k-vecinos más cercanos (kNN) (Taghizadeh *et al.*, 2022).

Para evaluar la efectividad de las capacidades de los modelos regresión y de aprendizaje automático utilizados para predecir las propiedades del suelo a partir del NDVI se determinaron los estadígrafos: Raíz Cuadrada del Error Cuadrático Medio (RMSE) Montgomery *et al.* (2003) Media del Error Absoluto (MAE) Panigrahi *et al.* (2023), el coeficiente de determinación (R^2) Gatera *et al.* (2023) coeficiente de correlación (r) Montgomery *et al.* (2003) y el estadístico Durbin-Watson (DW) (de Smith *et al.*, 2013). En todos los casos se utilizó el 70 % de los datos para realizar la estimación y el 30 % restante para validar el modelo según la metodología propuesta por Whetton *et al.* (2017).

Se realizó una clasificación jerárquica basada en las métricas de rendimiento de los modelos de aprendizaje automático (bosques aleatorios, vector de regresión, k-vecino más cercano y el gradiente estocástico) mediante los índices de eficiencia (Nash-Sutcliffe efficiency index, EF) propuesto por Nash & Sutcliffe (1970) y el de concordancia según Willmott (1982).

RESULTADOS Y DISCUSIÓN

El análisis descriptivo del NDVI calculado a partir de una imagen del satélite Landsat 9, proporciona una visión general completa de los datos, donde el valor promedio encontrado del NDVI fue de 0,26 como se muestre en la (Tabla 2).

Es importante señalar que la escala del NDVI oscila entre -1 y 1 y, en este caso particular, los valores observados se encuentran dentro de este rango. Esto concuerda con la proposición de Rawashdeh (2012).

de que, para el NDVI, los valores entre 0 y 0,5 indican una presencia limitada de vegetación, de acuerdo con las condiciones actuales del área de estudio. Estos conocimientos arrojan luz sobre el paisaje vegetal y contribuyen a una comprensión más profunda de los resultados de la investigación.

La Tabla 3 muestra los estadígrafos del análisis de regresión lineal de los modelos generados entre el NDVI y las propiedades del suelo utilizadas como índice de calidad. Con el uso de la técnica de regresión simple entre los valores de NDVI y las propiedades de suelo utilizadas como indicador de calidad se pueden evidenciar que existe una alta correlación de 0,98 entre el contenido de materia orgánica del suelo y el NDVI.

El coeficiente de determinación fue de 0,94, por lo que se pudo afirmar que el NDVI puede predecir el contenido de materia orgánica, con un error en su pronóstico en todos los casos dentro de los rangos permisibles en los que son medidos las variables determinadas Ayoubi *et al.* (2011), tanto el NDVI con valores que pueden oscilar desde -1 a 1 y la materia orgánica con valores máximos de 6,5 %.

Los resultados resaltan asociaciones positivas notables entre la concentración de magnesio presente en el suelo y el NDVI, que se sitúa en un 0,93 % de coeficiente de correlación. Estos resultados se alinean con los resultados reportados por Mazur *et al.* (2022), quienes también observaron una correlación del 0,95 entre el contenido de magnesio y el NDVI en un suelo destinado específicamente al cultivo de cereales. La relación entre el NDVI y el contenido de calcio del 0,94 se asocia significativamente con los hallazgos presentados por Abdalkarim *et al.* (2023), en los que sugiere que hay una ocurrencia simultánea de un aumento del contenido de calcio en los suelos y una reducción de la cobertura vegetal.

En el análisis de regresión lineal simple entre el NDVI y el potasio. El coeficiente de regresión indica que el modelo ajustado explica el 0,31 de la variabilidad del potasio. El coeficiente de correlación es igual a 0,56 lo que muestra una relación moderada entre las variables. El error estándar de la estimación muestra que la desviación estándar de los

TABLA 2. Análisis descriptivo del NDVI en el área de estudio

Índice espectral	Media	DE	EE	CV (%)	Mínimo	Máximo	Mediana
NDVI	0,26	0.06	0.01	21.74	0.11	0.43	0.25

DE: desviación estándar; EE: error estándar; CV: coeficiente de variación

TABLA 3. Estadígrafos del análisis de regresión lineal

Modelos de regresión lineal	R ²	r ²	E.E	MAE	D.W
NDVI vs. MO (%)	0,94	0,98	0,27	0,20	0,33
NDVI vs. Mg (cmol kg ⁻¹)	0,88	0,93	4,12	3,55	0,10
NDVI vs. Ca (cmol kg ⁻¹)	0,90	0,95	3,41	2,84	0,14
NDVI vs. P ₂ O ₅ (mg kg ⁻¹)	0,90	0,82	1,68	1,33	0,13
NDVI vs. Nt (%)	0,84	0,70	0,01	0,01	0,12
NDVI vs K ₂ O (mg kg ⁻¹)	0,31	0,56	12,26	9,67	0,23
NDVI vs. CE (dS m ⁻¹)	0,58	0,76	0,21	0,17	0,11
NVDI vs Na (cmol kg ⁻¹)	0,47	0,63	0,11	0,27	0,16
NVDI vs pH	0,34	0,31	0,04	0,17	0,15

R²: coeficiente de determinación; r²: coeficiente de correlación; E.E: error estándar; MAE: Media del Error Absoluto; D.W: Durwin-Watson.

residuos es 12,26 mg kg⁻¹ lo cual se puede utilizar para construir límites de predicción. El error absoluto medio es de 9,67 mg kg⁻¹ el cual es el valor promedio de los residuos.

En todos los modelos de regresión lineal simple se observa que en el estadístico Durbin-Watson (DW) los valores oscilan entre 0,06 a 0,33 lo cual indica que existe una autocorrelación positiva entre los residuos. Estos valores del estadístico DW refieren que la autocorrelación espacial que existe se debe a que las propiedades tienen una tendencia a estar agrupadas con áreas donde el comportamiento espacial tiende a ser la principal fuente de errores.

Esta investigación sugiere que cuando se obtienen valores bajos del coeficiente de determinación (0,50) no significa que los modelos sean de mala calidad; más bien apunta a la presencia de una serie de factores esenciales, no tenidos en cuenta por el modelo y características cualitativas que son difíciles de determinar a partir de los datos del satélite Landsat (Gopp *et al.*, 2019).

No todos los modelos aportados por el análisis de regresión lineal son adecuados, debido a que en las propiedades anteriormente referidas no existe una estructura sólida de los modelos, para hacer inferencias imparciales con respecto a la dependencia funcional entre el NDVI y las propiedades del suelo. El modelo de regresión lineal es el enfoque más utilizado para estimar las propiedades del suelo con el uso de datos derivados de la teledetección (Vergopolan *et al.*, 2021). Sin embargo, tiene

limitaciones en el manejo de relaciones no lineales entre las variables de respuesta y predictores que generalmente existen en los diferentes usos de las tierras con fines agrícolas (Khanal *et al.*, 2018).

Como se vio no todas las relaciones entre el NVDI y propiedades del suelo que resultan indicadores de calidad satisfacen una relación lineal, por lo tanto, se hace necesario explorar otros tipos de modelos del aprendizaje automático. Los parámetros de calibración y validación de los modelos arrojados por la herramienta de bosques aleatorios se muestran en la Tabla 4. Se observó que el modelo RF estimó, adecuadamente, el contenido de potasio al presentar valores bajos de RMSE (2,34) y altos de R² (0,98).

El modelo RF tuvo un buen desempeño durante el proceso de validación, especialmente, en la estimación del pH, avalado por el valor de R² (0,93), lo cual indica un buen ajuste entre los datos medidos y los estimados, explicó, además, el 93% de la variabilidad en los datos. Por tanto, puede asumirse que RF es confiable y preciso para estimar esta variable. A diferencia del modelo que arroja la regresión lineal simple para estimar el nitrógeno total, no se predice esta propiedad. En el caso del uso del RF determina el 75 % de la variación total de los valores de NDVI, lo cual afirma una alta determinación de esta propiedad con errores cercanos a 0.

Diversos estudios han confirmado que el modelo RF predice las propiedades del suelo significativamente mejor que los métodos de regresión lineal. Los estudios realizados

TABLA 4. Calibración y validación de los modelos generados por bosques aleatorios (RF)

Modelos	Calibración			Validación		
	RMSE	MAE	R ²	RMSE	MAE	R ²
RF_Nt	0,02	0,01	0,38	0,01	0,01	0,75
RF_P ₂ O ₅	3,41	2,02	0,35	2,08	1,17	0,76
RF_K ₂ O	6,56	3,11	0,84	2,34	1,67	0,98
RF_Ca	10,32	7,92	0,15	7,94	5,97	0,48
RF_Mg	11,96	8,98	0,01	9,12	6,96	0,39
RF_MO	0,97	0,77	0,45	0,73	0,57	0,69
RF_CE	0,31	0,13	0,33	0,20	0,07	0,73
RF_Na	0,27	0,17	0,41	0,15	0,11	0,80
RF_pH	0,20	0,09	0,39	0,06	0,04	0,93

RMSE: Raíz cuadrada del Error Cuadrático Medio; MAE: Media del Error Absoluto; R²: Coeficiente de Determinación

ilustran el uso de las bandas originales del sensor y la determinación de índices espectrales para estimar las propiedades del suelo (Zhang *et al.*, 2018). Al igual que en esta investigación, existen referencia de trabajos con valores de NDVI en suelos desnudos los cuales han estimado las propiedades del suelo con modelos RF (Jiang *et al.*, 2018).

El modelo SVM el cual se muestra en la Tabla 5, tiene un rendimiento variado, con algunos resultados muy positivos y otros extremadamente negativos. En el caso del Na muestra un R^2 de 0,79 en la validación, lo que sugiere que SVM es bastante efectivo para esta variable, capturando la mayor parte de la variabilidad en los datos. Sin embargo, para el Nt presenta un R^2 de 0,41 en la validación, valor muy bajo según el rango en el que mejor se debe de desempeñar este tipo de coeficiente.

El kNN demuestra ser un modelo robusto según la Tabla 6, con un rendimiento consistente en la mayoría de las variables. Para el K_2O se obtiene un R^2 de 0,84 en la validación, indicando una excelente capacidad predictiva. En tanto para el pH muestra un R^2 de 0,67, lo que indica que el modelo es adecuado, aunque no tan preciso como RF en esta variable.

Los modelos derivados del SGD se presentan los resultados en la Tabla 7. De acuerdo a los valores obtenidos de los estadígrafos evaluados, los resultados que arroja son moderados en todas las variables, sin destacar en ninguna en particular. La MO durante la validación alcanza un R^2 de 0,13, lo que sugiere un desempeño limitado, mientras que el K_2O logra un R^2 de 0,41, lo que indica que el modelo puede captar algunas relaciones, pero es superado por RF y kNN.

En relación con los resultados obtenidos, Forkuor *et al.* (2017) plantean que obtener valores bajos de R^2 generalmente pueden atribuirse a una interacción compleja y una alta variabilidad de factores ambientales y la alta variabilidad en las prácticas agrícolas como son la gestión del suelo, la aplicación de nutrientes y la cubierta vegetal. De acuerdo con la opinión de Chai y Draxler (2014), es poco probable que RMSE proporcione una evaluación sólida de los modelos. Cabe

señalar especialmente que el índice de eficiencia y de concordancia pueden proporcionar información más eficaz sobre el rendimiento del modelo. Estas diferencias en los estimadores para la evaluación se deben a que, entre los valores observados y estimados al calcularse al cuadrado, pueden sobreestimar los valores más grandes en los datos estimados mientras que los valores más pequeños pueden desprejarse (Willmott, 1982).

En la Tabla 8 se presentan los índices de eficiencia y de concordancia de cada estimación de propiedades del suelo y el indicador de calidad a partir de los modelos de aprendizaje automático (bosques aleatorios, vector de regresión, k-vecino más cercano y el gradiente estocástico). Resulta evidente que el modelo RF demuestra el nivel más alto de eficiencia, como lo indica la prevalencia de valores que oscilan constantemente entre 0,77 y 1,0 para el índice de eficiencia, junto con un rango igualmente impresionante de valores de 0,93 a 1,0 para el índice de concordancia. En una interpretación más amplia, esto sugiere que el modelo RF posee una capacidad encomiable para hacer predicciones precisas, particularmente en el contexto de alinear los valores de varias propiedades del suelo con las condiciones promedio observadas, mientras que el alto grado de concordancia significa que los resultados predictivos son notablemente congruentes con los datos reales observados en escenarios reales, lo que presenta una perspectiva muy favorable para la aplicación de metodologías de aprendizaje automático en este dominio.

Siguiendo el modelo RF en la clasificación jerárquica basada en las métricas de rendimiento de los índices de eficiencia y concordancia, el modelo k-NN también muestra un rendimiento favorable, ya que arroja resultados que se sitúan dentro de un rango de 0,78 a 1,0 para el índice de eficiencia, y un índice de concordancia de alta validez, que oscila entre 0,92 y 1,0, lo que indica su eficacia para hacer predicciones confiables. En este contexto de métricas comparativas de desempeño, cabe destacar que el modelo SGD muestra resultados superiores cuando se yuxtapone con el modelo SVM, lo que destaca la eficacia relativa de estos diferentes enfoques de modelado predictivo en el análisis de los datos de las propiedades del suelo.

TABLA 5. Calibración y validación de los modelos generados por la Máquina de Vectores de Soporte (SVM, por sus siglas en inglés)

Modelos	Calibración			Validación		
	RMSE	MAE	R^2	RMSE	MAE	R^2
SVM_Nt	0,03	0,03	0,72	0,03	0,03	0,41
SVM_P ₂ O ₅	3,61	2,64	0,42	3,2	2,16	0,53
SVM_K ₂ O	15,19	9,57	0,12	13,65	8,53	0,28
SVM_Ca	10,6	8,67	0,11	9,49	7,62	0,26
SVM_Mg	11,86	9,17	0,01	11,07	8,35	0,1
SVM_MO	0,99	0,82	0,42	0,89	0,71	0,53
SVM_CE	0,29	0,18	0,42	0,25	0,15	0,56
SVM_Na	0,22	0,15	0,6	0,16	0,12	0,79
SVM_pH	0,2	0,12	0,37	0,16	0,09	0,56

RMSE: Raíz cuadrada del Error Cuadrático Medio; MAE: Media del Error Absoluto; R^2 : Coeficiente de Determinación.

TABLA 6. Calibración y validación de los modelos generados con k-vecinos más cercanos (kNN)

Modelos	Calibración			Validación		
	RMSE	MAE	R ²	RMSE	MAE	R ²
kNN_Nt	0,01	0,01	0,57	0,01	0,01	0,69
kNN_P ₂ O ₅	2,89	1,97	0,53	2,15	1,37	0,74
kNN_K ₂ O	5,38	3,29	0,89	3,67	2,18	0,95
kNN_Ca	9,09	7,01	0,34	6,8	5,03	0,62
kNN_Mg	10,05	7,48	0,29	8,21	5,98	0,51
kNN_MO	0,8	0,63	0,62	0,65	0,5	0,75
kNN_CE	0,27	0,15	0,48	0,18	0,1	0,77
kNN_Na	0,24	0,15	0,55	0,17	0,11	0,75
kNN_pH	0,19	0,1	0,4	0,14	0,06	0,67

RMSE: Raíz cuadrada del Error Cuadrático Medio; MAE: Media del Error Absoluto; R²: Coeficiente de Determinación**TABLA 7.** Calibración y validación de los modelos generados por Descenso de Gradiente Estocástico (SGD por sus siglas en inglés)

Modelos	Calibración			Validación		
	RMSE	MAE	R ²	RMSE	MAE	R ²
SGD_Nt	0,02	0,01	0,14	0,02	0,01	0,23
SGD_P ₂ O ₅	4,28	3,64	0,01	4,01	3,44	0,09
SGD_K ₂ O	13,39	11,31	0,61	12,43	10,57	0,73
SGD_Ca	11,7	9,46	0,09	10,82	8,87	0,03
SGD_Mg	12,2	10,02	0,05	11,37	9,52	0,06
SGD_MO	1,26	1,04	0,06	1,21	0,99	0,13
SGD_CE	0,36	0,29	0,08	0,33	0,27	0,22
SGD_Na	0,33	0,24	0,09	0,29	0,22	0,28
SGD_pH	0,21	0,16	0,27	0,19	0,14	0,35

RMSE: Raíz cuadrada del Error Cuadrático Medio; MAE: Media del Error Absoluto; R²: Coeficiente de Determinación**TABLA 8.** Índice de eficiencia y de concordancia para la estimación de las propiedades del suelo

Propiedades del suelo	Índice de eficiencia				Índice de concordancia			
	RF	SVM	KNN	SGD	RF	SVM	KNN	SGD
Nt	0,98	0,21	0,98	0,94	1,00	0,69	1,00	0,99
P	0,97	0,70	0,91	0,88	0,99	0,99	0,95	0,98
K	1,00	0,41	0,99	0,39	1,00	0,92	1,00	0,95
Ca	0,77	0,68	0,84	0,52	0,97	0,15	0,92	0,43
Mg	0,60	0,02	0,57	0,76	0,93	0,91	0,94	0,88
Na	0,99	0,99	0,99	0,97	1,00	1,00	1,00	0,99
MO	0,86	0,68	0,78	0,79	0,98	0,96	0,80	0,94
CE	0,44	0,83	0,68	0,22	0,98	0,95	0,97	0,86
pH	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00

Ko *et al.* (2024) plantean que un Índice de Eficiencia alto sugiere que el modelo captura bien la variabilidad de los datos observados, siendo superior los valores en este estudio que los reportados por esta investigación en el cultivo de la soya. Arrojan que la eficiencia de los modelos de aprendizaje automático puede estar dada además por el área geográfica o el tipo de suelo.

Dharumarajan *et al.* (2017) obtuvieron en la predicción del carbono orgánico del suelo y del pH a partir de índices espectrales de vegetación obtenido por imágenes satelitales, valores del Índice de Concordancia de 0,37 y 0,38 respectivamente, y solo fue aceptable el modelo

arrojado en la predicción de la conductividad eléctrica con 0,70 para este Índice, el cual en este estudio fue superior con 0,86.

CONCLUSIONES

- Es posible estimar la materia orgánica, el magnesio, el calcio y el fósforo a partir del NVDI con regresión lineal simple dado que al utilizar el estadístico Durbin-Watson los valores oscilan entre 0,06 a 0,33 lo cual indica que existe una autocorrelación positiva entre los residuos, existiendo propiedades que tienen una tendencia a estar agrupados con áreas donde

el comportamiento espacial es la principal fuente de errores.

- El modelo de bosques aleatorios es el más adecuado para estimar el potasio y el pH, avalado con valores de R^2 (cerca de 1), RMSE (próximos a 0) y los Índices de Eficiencia y Concordancia (cerca de 1).
- Los restantes algoritmos lograron mejor ajuste en el siguiente orden decreciente de acuerdo a los valores de los índices de eficiencia y concordancia calculados: el kNN, SGD y SVM.

REFERENCIAS BIBLIOGRÁFICAS

- ABDALKARIM, K.; GAZNAYEE, H.A.A.; AL-QURAISHI, A.M.F.; ABDALLA, Z.O.: "Predictive Digital Mapping of Surface Soil Properties using Remote Sensing and Multivariate Statistical Analysis.", *Zanco Journal of Pure and Applied Sciences*, 35(6): 189-203, 2023, ISSN: 2412-3986.
- ANGLADETTE, A.: *El arroz. Agricultura Tropical, Colección Agricultura Tropical*, Ed. Editorial Blume, 867 p., 1969, ISBN: 84-7313-835-X.
- AYOUBI, S.; SHAHRI, A.P.; KARCHEGANI, P.M.; SAHRAWAT, K.L.: "Application of artificial neural network (ANN) to predict soil organic matter using remote sensing data in two ecosystems", *Biomass and remote sensing of biomass*, 10: 181-196, 2011.
- BAROUDY, A.A.L.; ALI, A.M.; MOHAMED, E.S.; MOGHANM, F.S.; SHOKR, M.S.; SAVIN, I.; PODDUBSKY, A.M.; DING, Z.; KHEIR, A.; ALDOSARI, A.A.: "Modeling land suitability for rice crop using remote sensing and soil quality indicators: The case study of the Nile delta", *Sustainability*, 12(22): 9653, 2020, ISSN: 2071-1050.
- CASANOVAS, C.E.; SUÁREZ DEL VILLAR, L.E.; ÁLVAREZ, S.A.; AVILLEIRA, C.I.: "Valoración de la seguridad alimentaria cubana a partir de la superficie agrícola explotada y los rendimientos agrícolas", *Revista Universidad y Sociedad*, 14(5): 304-314, 2022, ISSN: 2218-3620.
- CHAI, T.; DRAXLER, R.R.: "Root mean square error (RMSE) or mean absolute error (MAE)?-Arguments against avoiding RMSE in the literature", *Geoscientific model development*, 7(3): 1247-1250, 2014, ISSN: 1991-9603.
- CHOUDHURY, B.U.; MANDAL, S.: "Indexing soil properties through constructing minimum datasets for soil quality assessment of surface and profile soils of intermontane valley (Barak, North East India)", *Ecological Indicators*, 123: 107369, 2021, ISSN: 1470-160X.
- DE SMITH, M.; GOODCHILD, M.; LONGLEY, P.: *Geospatial analysis. A comprehensive guide to principles, techniques and software tools*, Ed. The Winchester Press, Winchester, UK, Winchester, UK, 2013.
- DEL VALLE, M.J.; GONZÁLEZ, V.; RAFAEL, P.L.; SÁNCHEZ, A.O.R.; DELGADO, T.C.: "Efecto de las variables climáticas sobre el rendimiento agrícola del arroz (*Oryza sativa* L.)", *Ingeniería Agrícola*, 12(1), 2022, ISSN: 2227-8761.
- DHARUMARAJAN, S.; HEGDE, R.; SINGH, S.: "Spatial prediction of major soil properties using Random Forest techniques-A case study in semi-arid tropics of South India, *Geoderma Reg.*, 10, 154-162", 2017.
- FORKUOR, G.; HOUNKPATIN, O.K.; WELP, G.; THIEL, M.: "High resolution mapping of soil properties using remote sensing variables in south-western Burkina Faso: a comparison of machine learning and multiple linear regression models", *PloS one*, 12(1): e0170478, 2017, ISSN: 1932-6203.
- GARCÍA, R.R.A.; RUÍZ, P.M.E.; SERGIO, R.R.: "Indicador de calidad de un Vertisol dedicado al arroz en la provincia Holguín, Cuba", *Revista Ciencias Técnicas Agropecuarias*, 34, 2025.
- GATERA, A.; KURADUSENGE, M.; BAJPAI, G.; MIKEKA, C.; SHRIVASTAVA, S.: "Comparison of random forest and support vector machine regression models for forecasting road accidents", *Scientific African*, 21: e01739, 2023, ISSN: 2468-2276.
- GOPP, N.; SAVENKOV, O.; NECHAEVA, T.; SMIRNOVA, N.; SMIRNOV, A.: "Application of NDVI in digital mapping of phosphorus content in soils and phosphorus supply assessment in plants", *Izvestiya, Atmospheric and Oceanic Physics*, 55: 1322-1328, 2019, ISSN: 0001-4338.
- HERNÁNDEZ, J.A.; PÉREZ, P.; BOSCH, I.D.; CASTRO, S.N.: *Clasificación de los suelos de Cuba.*, Ed. Instituto Nacional de Ciencias Agrícolas, Instituto Nacional de Ciencias Agrícolas, ed., San José de las Lajas, Mayabeque, Cuba, 91 p., 2015, ISBN: 959-7023-77-6.
- JIANG, Y.; RAO, L.; SUN, K.; HAN, Y.; GUO, X.: "Spatio-temporal distribution of soil nitrogen in Poyang lake ecological economic zone (South-China)", *Science of the total environment*, 626: 235-243, 2018, ISSN: 0048-9697.
- KHANAL, S.; FULTON, J.; KLOPFENSTEIN, A.; DOURIDAS, N.; SHEARER, S.: "Integration of high resolution remotely sensed data and machine learning techniques for spatial prediction of soil properties and corn yield", *Computers and electronics in agriculture*, 153: 213-225, 2018, ISSN: 0168-1699.
- KO, J.; SHIN, T.; KANG, J.; BAEK, J.; SANG, W.G.: "Combining machine learning and remote sensing-integrated crop modeling for rice and soybean crop simulation", *Frontiers in Plant Science*, 15: 1320969, 2024, ISSN: 1664-462X.
- LISS, M.; SCHMIDT, J.; GLASER, B.: "Improving the spatial prediction of soil organic carbon stocks in a complex tropical mountain landscape by methodological specifications in machine learning

- approaches", *PLoS One*, 11(4): e0153673, 2016, ISSN: 1932-6203.
- LUQUE, R.L.: "Revisión Sistemática de Literatura de Imágenes Satelitales en Hidrología y Agricultura", *Revista Ibérica de Sistemas e Tecnologías de Informação*, (E55): 264-278, 2023, ISSN: 1646-9895.
- MAZUR, P.; GOZDOWSKI, D.; WNUK, A.: "Relationships between soil electrical conductivity and sentinel-2-derived NDVI with pH and content of selected nutrients", *Agronomy*, 12(2): 354, 2022, ISSN: 2073-4395.
- MONTGOMERY, D.C.; SKINNER, K.R.; RUNGER, G.C.: "Process monitoring for multiple count data using generalized linear model-based control charts", *International Journal of Production Research*, 41(6): 1167-1180, 2003, ISSN: 0020-7543.
- NASH, J.E.; SUTCLIFFE, J.V.: "River flow forecasting through conceptual models part I-A discussion of principles", *Journal of hydrology*, 10(3): 282-290, 1970, ISSN: 0022-1694.
- NORMA CUBANA (NC): "Calidad del suelo-determinación de los componentes orgánicos", NC: 1043.2014, Oficina Nacional de Normalización, Cuba. 2014.
- NORMA CUBANA (NC): "Calidad del suelo-determinación de los componentes orgánicos", NC: 1043.2014, Oficina Nacional de Normalización, Cuba. 2014.
- NORMA CUBANA (NC): "Determinación de las formas móviles de Fósforo y Potasio", NC: 52.1999, Oficina Nacional de Normalización, Cuba. 1999.
- NORMA CUBANA (NC): "Calidad del Suelo. Determinación de la capacidad de intercambio catiónico y de los cationes intercambiables del suelo", NC: 209.2002, Oficina Nacional de Normalización, Cuba. 2002.
- NORMA CUBANA (NC): "Calidad del Suelo. Determinación del Nitrógeno total Método Kjeldahl", NC: 11261.2009, Oficina Nacional de Normalización, Cuba. 2009.
- NORMA CUBANA (NC): "Calidad del Suelo. Evaluación de la afectación por salinidad", NC: 776.2010, Oficina Nacional de Normalización, Cuba. 2010.
- NORMA CUBANA (NC): "Calidad del suelo-determinación de pH", NC: 10390.2014, Oficina Nacional de Normalización, Cuba. 2014.
- PANIGRAHI, B.; KATHALA, K.C.R.; SUJATHA, M.: "A machine learning-based comparative approach to predict the crop yield using supervised learning with regression models", *Procedia Computer Science*, 218: 2684-2693, 2023, ISSN: 1877-0509.
- PARK, H.J.; BAEK, N.; SEO, B.S.; JEONG, Y.J.; YANG, H.I.; LEE, S.I.; YOON, K.S.; KIM, H.Y.; CHOI, W.J.: "Estimation of the electrical conductivity of saturated paste from soil-water extracts of coastal saline paddy soils using random forest and multiple regression models", *Journal of Soils and Sediments*, 24(3): 1250-1259, 2024, ISSN: 1439-0108.
- ROUSE JR, J.W.; HAAS, R.H.; DEERING, D.; SCHELL, J.; HARLAN, J.C.: *Monitoring the vernal advancement and retrogradation (green wave effect) of natural vegetation*, 1974.
- SHRESTHA, N.; SHUKLA, S.: "Support vector machine based modeling of evapotranspiration using hydro-climatic variables in a sub-tropical environment", *Agricultural and forest meteorology*, 200: 172-184, 2015, ISSN: 0168-1923.
- SINGH, G.; SINGH, J.; WANI, O.A.; EGBUERI, J.; AGBASI, J.C.: "Assessment of groundwater suitability for sustainable irrigation: a comprehensive study using indexical, statistical, and machine learning approaches", *Groundwater for Sustainable Development*, 24: 101059, 2024, ISSN: 2352-801X.
- SIQUEIRA, R.G.; MOQUEDACE, C.M.; FERNANDES-FILHO, E.I.; SCHAEFER, C.E.R.; FRANCELINO, M.R.; SACRAMENTO, I.F.; MICHEL, R.F.: "Modelling and prediction of major soil chemical properties with Random Forest: Machine learning as tool to understand soil-environment relationships in Antarctica", *Catena*, 235: 107677, 2024, ISSN: 0341-8162.
- TAGHIZADEH, M.R.; KHADEMI, H.; KHAYAMIM, F.; ZERAATPISHEH, M.; HEUNG, B.; SCHOLTEN, T.: "A comparison of model averaging techniques to predict the spatial distribution of soil properties", *Remote S*, 14(3): 472, 2022, ISSN: 2072-4292.
- VERGOPOLAN, N.; CHANEY, N.W.; PAN, M.; SHEFFIELD, J.; BECK, H.E.; FERGUSON, C.; TORRES-, R.L.; SADRI, S.; WOOD, E.F.: "SMAP-HydroBlocks, a 30-m satellite-based soil moisture dataset for the conterminous US", *Scientific data*, 8(1): 264, 2021, ISSN: 2052-4463.
- VILLAZÓN, J.A.; NORIS, P.; GARCÍA, R.A.; CRUZ, M.: "Análisis temporal de la agresividad y concentración de las precipitaciones en áreas agropecuarias de la provincia de Holguín, Cuba", *Idesia (Arica)*, 41(3): 77-86, 2023.
- WHETTON, R.; ZHAO, Y.; SHADDAD, S.; MOUAZEN, A.M.: "Nonlinear parametric modelling to study how soil properties affect crop yields and NDVI", *Computers and electronics in agriculture*, 138: 127-136, 2017, ISSN: 0168-1699.
- WILLMOTT, C.J.: "Some comments on the evaluation of model performance", *Bulletin of the American Meteorological Society*, 63(11): 1309-1313, 1982, ISSN: 0003-0007.
- ZHANG, Y.; SUI, B.; SHEN, H.; WANG, Z.: "Estimating temporal changes in soil pH in the black soil region of Northeast China using remote sensing", *Computers and Electronics in Agriculture*, 154: 204-212, 2018, ISSN: 0168-1699.

Roberto Alejandro García-Reyes, Ing., Inv., Ministerio de la Agricultura, Departamento Provincial de Suelos y Fertilizantes, provincia Holguín, Cuba. e-mail: ralejandro9409@gmail.com.

María Elena Ruiz-Pérez, Dr.C., Profesora Titular, Universidad Agraria de La Habana “Fructuoso Rodríguez Pérez”. Carretera Tapaste y Autopista Nacional km 231/2, San José de Las Lajas, Mayabeque, Cuba. CP 32700. e-mail: mruiiz@unah.edu.cu.

Sergio Rodríguez-Rodríguez, Dr.C., Universidad de Granma, Facultad de Ciencias Agrícolas, Centro de Estudios de Biotecnología Vegetal, Bayamo, Granma, Cuba. e-mail: ralejandro9409@gmail.com

Juan Alejandro Villazón-Gómez, MSc., Profesor, Universidad de Holguín. Departamento de Ciencias Agropecuarias, Holguín, Cuba. e-mail: ralejandro9409@gmail.com

La mención de marcas comerciales de equipos, instrumentos o materiales específicos obedece a propósitos de identificación, no existiendo ningún compromiso promocional con relación a los mismos, ni por los autores ni por el editor.