

Algoritmos de aprendizaje de máquina para la predicción de propiedades fisicoquímicas del suelo mediante información espectral: una revisión sistemática

Machine learning algorithms for prediction of physicochemical soil properties by spectral information: a systematic review

Mateo Vargas-Zapata¹

Marisol Medina-Sierra²

Luis Fernando Galeano-Vasco³

Mario Fernando Cerón-Muñoz⁴

Recibido: diciembre 01 de 2021

Aceptado: : diciembre 28 de 2021

Resumen

En la literatura científica actual se discute ampliamente acerca de la predicción de propiedades edáficas mediante información espectral. El objetivo de esta revisión fue encontrar algoritmos con el mayor potencial predictivo para las propiedades fisicoquímicas del suelo, basados en información espectral capturada con diferentes instrumentos. Se realizó una revisión sistemática en la cual se encontraron 121 artículos de los cuales se eligieron 19, que cumplieran con un coeficiente de determinación mayor a 0,80 o una raíz del error cuadrado medio cercana a 0. Se determinó que el rango espectral más utilizado corresponde al rango desde 350 hasta 2500 nm; los algoritmos mínimos cuadrados parciales, máquina de soporte vectorial y máquina de soporte vectorial ajustado son adecuadas para predecir pH, materia orgánica y carbono orgánico. Además, la regresión lineal solo es efectiva para predecir el carbonato de calcio, materia orgánica, humedad y contenido de agua mediante bandas individuales.

Palabras clave: algoritmos de predicción, aprendizaje de máquina, análisis químico, espectroscopía.

Abstract

The prediction of soil properties through spectral information is widely discussed in the current scientific literature. The objective of this review was to find algorithms with the highest predictive potential for soil physicochemical properties based on spectral information captured with different instruments. A systematic review was carried out in which 121 articles were found, and 19 of them were chosen which met a determination coefficient greater than 0.80 or a root mean square error close to 0. It was determined that the most used spectral range corresponds to the range from 350 to 2500 nm; the partial least squares, support vector machine, and adjusted support vector machine algorithms are suitable for predicting pH, organic matter, and organic carbon. Furthermore, linear regression is only effective in predicting calcium carbonate, organic matter, moisture, and water content using individual bands.

Keywords: prediction algorithms, machine learning, chemical analysis, spectroscopy.

1 Zootecnista, Universidad de Antioquia, Medellín, Colombia. E-mail: mateo.vargas@udea.edu.co

Orcid: <https://orcid.org/0000-0003-4140-088X>

2 Ingeniera Agrónoma, Doctora en Ciencias, Universidad de Antioquia, Medellín, Colombia. E-mail: marisol.medina@udea.edu.co

Orcid: <https://orcid.org/0000-0003-1929-8305>

3 Zootecnista, Doctor en Ciencias, Universidad de Antioquia, Medellín, Colombia. E-mail: luis.galeano@udea.edu.co

Orcid: <https://orcid.org/0000-0002-6842-3945>

4 Zootecnista, Doctor en Ciencias, Universidad de Antioquia, Medellín, Colombia. E-mail: mario.ceron@udea.edu.co

Orcid: <https://orcid.org/0000-0002-7233-6625>

1. Introducción

Existen diferentes razones por las que la determinación de los componentes fisicoquímicos del suelo a través de las metodologías tradicionales, resulta ser ineficiente y costosa (Demattê et al., 2015; Yunus et al., 2019; Guo et al., 2019; Shi et al., 2020; Dhawale et al., 2021; Olatunde et al., 2021). Debido a esto, el empleo de metodologías como la espectroscopia de infrarrojo ha cobrado relevancia como un método alternativo (no sustituto) a los procesos convencionales (Nawar et al., 2018; Jaconi et al., 2019; Liu et al., 2019; Barra et al., 2021), ya que la obtención de resultados se logra

a mayor velocidad, menos costos e impactos negativos al ambiente (Gourlay et al., 2017; Fang et al., 2018; Hong et al., 2018; Demattê et al., 2019; Dangal et al., 2019).

Adicionalmente, Guerrero et al. (2021), reportaron un amplio uso de sensores con capacidad de captar datos procedentes del espectro electromagnético (figura 1), como el visible (VIS), el infrarrojo cercano (NIR) y el infrarrojo medio (MIR). Los cuales hacen parte de las tecnologías de la información y la comunicación, que a su vez son un prerrequisito esencial para la implementación de la agricultura moderna (Benos et al., 2021).

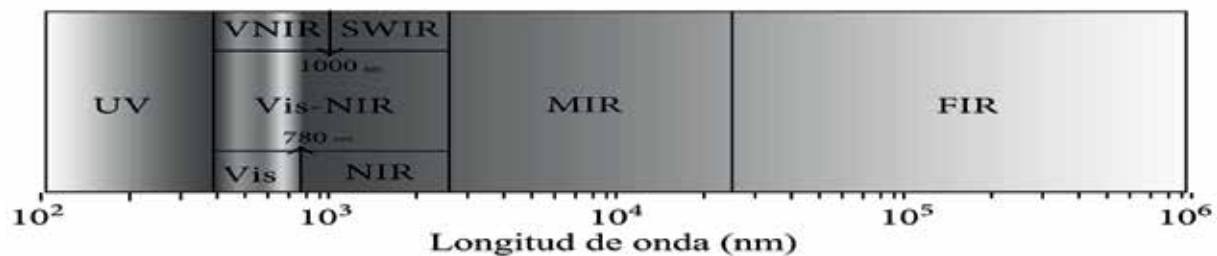


Figura 1. Ilustración de los rangos del espectro electromagnético (modificada de Fang et al., 2018).

Las problemáticas mencionadas se han abordado por medio de diferentes sensores, con la finalidad de optimizar recursos. Sin embargo, factores como el tipo de sensor, la resolución espacial, espectral y temporal, pueden influir en los resultados finales. De hecho, existe una amplia gama de sensores aeroespaciales, como: GeoEye-1, WorldView 2/3/4, DEIMOS-2, Skysat, Dove, RapidEye, etc., que según Inoue (2020), cuentan con al menos cuatro bandas espectrales comprendidas en las regiones VIS-NIR, resoluciones espaciales de 1 - 100 m y resoluciones temporales de 1 - 26 días. No obstante, Vohland et al. (2017), refieren que estos sensores están subordinados a variaciones espaciales y temporales, como: la cobertura vegetal, humedad, rugosidad, encostramiento y textura.

Por otra parte, los sensores aerotransportados al estar a menor altura que los aeroespaciales, podrían facilitar la obtención de imágenes con mayor resolución espacial y un aprovechamiento

del rango espectral que, según Angelopoulou et al. (2019), puede estar desde los 400 hasta los 2540 nm en sensores como: HyMap, ProSpecTIR V-S, AISA-Eagle, etc. Asimismo, sensores hiperspectrales como Headwall Micro, NanoHyperspec, UHD 185-Firefly etc., pueden ser incorporados a sistemas de aeronaves remotamente piloteados o drones y cuentan con más de 100 bandas en la región VIS-NIR (Lu et al., 2020). Sin embargo, podrían estar sujetos a las limitaciones mencionadas antes.

En los últimos años se han utilizado de manera constante sensores de tipo portátil. Por ejemplo, en la revisión realizada por Ahmadi et al. (2021), se encontró que entre los años 1998 y 2019, el 60% de los artículos emplearon la serie ASD FieldSpec (39%) y FOSS NIR System (21%). Esto puede deberse a que permiten una mejor adecuación del sitio de muestreo para la captura de datos, además de la posibilidad de implementarse bajo condiciones controladas de laboratorio que, de

acuerdo con Kurz y Buckley (2016), supone una ventaja con respecto a los factores atmosféricos y de luminosidad.

De acuerdo a las caracterizaciones expuestas, se puede decir que cada clase de sensor provee información multidimensional, la cual, hoy en día es abordada con diferentes algoritmos pertenecientes al campo de la inteligencia artificial y del aprendizaje de máquina (ML). Autores como Padarian et al. (2020), mostraron cómo en los últimos 20 años ha variado proporcionalmente el uso de procedimientos predictivos en las ciencias de los suelos. Por ejemplo, el uso de modelos como los árboles de clasificación y regresión (CART) fue moderado ($> 40\%$) entre los años 2000 y 2005, en comparación con el periodo 2015-2020 donde fue bajo ($< 20\%$). Por el contrario, el bosque aleatorio (RF) registró un aumento ostensible desde 2005 ($< 10\%$) hasta el 2020 ($> 25\%$).

Otro dato relevante en la revisión de Padarian et al. (2020), es que métodos más avanzados como las redes neuronales artificiales (ANN), la máquina de soporte vectorial (SVM) y RF, tienen un mejor rendimiento que otros más simples, como la regresión por componentes principales (PCR), la regresión por mínimos cuadrados parciales (PLSR) y la regresión lineal múltiple (MLR). Esta afirmación es secundada por Diaz-Gonzalez et al. (2022), al indicar que modelos como ANN y RF no están sujetos a la linealidad de los datos.

Teniendo en cuenta el rendimiento de estos procedimientos, algunos investigadores como Ahmadi et al. (2021) consideraron que las métricas más comunes para medir el desempeño de los modelos usados en la predicción de las propiedades fisicoquímicas, son: el coeficiente de determinación (R^2), la ratio de desempeño de la desviación (RPD) y la raíz cuadrada del error medio (RMSE). Estos autores reportaron valores de R^2 desde 0,68 y 0,87 para propiedades como carbono orgánico (SOC), nitrógeno total (TN), materia orgánica (SOM), carbono total, salinidad, arena, humedad, carbono inorgánico y limo. Al mismo tiempo, encontraron valores de RMSE

promedio de 5,31% y 6,05% para arcilla y arena, respectivamente.

Otros investigadores como Malmir et al. (2019), encontraron que es posible predecir efectivamente nutrientes como el boro, calcio (Ca), cobre, magnesio (Mg) y zinc desde los 400 hasta los 1000 nm usando PLSR. Mientras que el aluminio (Al), hierro, potasio (K), manganeso, sodio (Na) y fósforo (P) no lo fueron. Adicionalmente, reportaron que la presencia de agua se puede identificar desde 1400 y 1900 nm, SOM desde los 350 hasta los 2500 nm y los minerales de la arcilla desde los 1400 - 1410 nm y 2160 - 2200 nm. No obstante, otros como el nitrógeno (N), P y K en su forma disponible, no tienen un rango específico definido (Qi et al., 2018), lo que puede ser debido a que los nutrientes edáficos tienen gran variación en diferentes longitudes de onda (Song et al., 2018).

En lo referente a la región MIR, diversos estudios de la última década muestran como este rango ha cobrado gran relevancia en la predicción de elementos clave del suelo. Concretamente, con R^2 mayores a 0,80 para parámetros como pH, arcilla, arena, limo, SOC, capacidad de intercambio catiónico (CIC), azufre (S), Al, Ca, Mg, Na, etc., mediante PLSR, ANN, PLSR - ANN, regresión por SVM y RF (Barra et al., 2021). Sin embargo, los mismos autores consideraron que el desarrollo de instrumentos portátiles que capten los rangos VIS-NIR-MIR, el mejoramiento de algoritmos de ML, preprocesamiento de información espectral y la transferibilidad de modelos entre sensores, deberían ser considerados como una oportunidad para mejorar la efectividad de las predicciones en campo.

Por estas razones, es posible que los avances tecnológicos en los sensores espectrales estén encaminados a integrarse directamente a los sistemas productivos donde esté involucrado el recurso suelo, con la finalidad de agilizar la toma de decisiones y disminuir la pérdida de recursos. Burton et al. (2020), mencionaron que la gama de sensores disponibles para el monitoreo del suelo debería ser incorporados con el internet de las cosas y la inteligencia artificial

para optimizar la producción y minimizar el uso de recursos.

La recolección de artículos científicos relacionados con la predicción de elementos edáficos y sus respuestas espectrales, ponen a disposición resultados relevantes acerca de procedimientos utilizados actualmente y cómo pueden ser empleados de forma individual para abordar variables multidimensionales mediante transformaciones. El objetivo de esta revisión fue encontrar algoritmos con el mayor potencial predictivo para las propiedades fisicoquímicas del suelo, basados en información espectral capturada con diferentes instrumentos. El presente trabajo partió de una revisión sistemática haciendo uso de bases de datos relevantes para la producción agrícola y el reporte de artículos científicos en general. Se identificaron las características de diferentes sensores espectrales, los rangos del espectro electromagnético que abarcan y en cuáles responden las propiedades fisicoquímicas del suelo. Adicionalmente, se mencionan los procedimientos de predicción utilizados, los resultados obtenidos con estos y finalmente, se comparan conjuntamente por medio de algunas de las métricas más utilizadas en años recientes para la evaluación de modelos estadísticos.

2. Metodología

La búsqueda de información se ejecutó entre los días 4 y 7 de enero de 2021, en las bases de datos: Scopus, ScienceDirect y Google Académico (figura 2), ya que son motores de búsqueda representativos para las ciencias agrarias y la academia en general. Se emplearon los términos: spectral, hyperspectral, soil, properties, prediction, algorithm, chemical y texture, con la finalidad de obtener investigaciones referentes a los sensores y las propiedades en cuestión. Asimismo, se usó el término machine learning junto con PLSR, PCR, Random Forest, BPNN (Redes neuronales de retropropagación) y CNN (redes neuronales convolucionales), para abarcar técnicas lineales y no lineales. Además de los operadores booleanos AND, OR y NOT mencionados por Aliyu (2017), para ayudar con el filtrado de resultados de manera más precisa. Luego de la búsqueda, se incluyeron artículos donde los sensores tuvieran alta resolución espectral y las propiedades estudiadas fueran componentes edáficas, más no metales pesados. Además, se tuvo en cuenta que las metodologías fueran similares, no hicieran uso de algoritmos de clasificación y que reportaran una métrica de evaluación, ya sea R2 o RMSE.

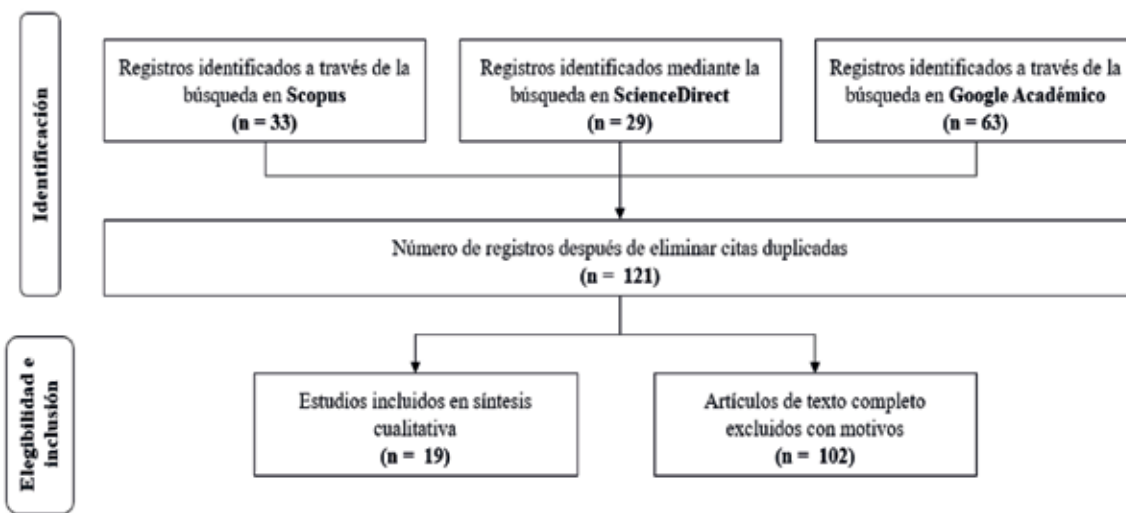


Figura 2. Diagrama de flujo para la selección de los artículos utilizados en la revisión.

3. Resultados y discusión

Inicialmente, se pudo identificar que el 60,87% de los sensores, cubren un rango desde los 350 hasta los 2500 nm (Lagacherie et al., 2008; Divya & Gopinathan, 2019; Crucil et al., 2019; Wei et al., 2020; Lei et al., 2016; Justin-George et al., 2020; He et al., 2004; Liu et al., 2009; Mustafa & Moursy, 2020; Gaikwad et al., 2020; Kurmai-Takács & Csorba, 2015; Santra et al., 2015; Sur & Chauhan, 2019; Vašát et al., 2017). El 30,43% cubre una porción menor del rango antes mencionado (Lagacherie et al., 2008; An et al., 2012; Li et al., 2019; Wu et al., 2020; Hong et al., 2020); y el 8,70% abarcan el infrarrojo lejano (FIR), infrarrojo de banda corta

(SWIR) y MIR (Deiss et al., 2020). Por tanto, podría decirse que el rango VIS-NIR-SWIR (350 - 2500) fue el más relevante en la revisión actual, independientemente de la referencia y resolución espectral de los instrumentos empleados (Tabla 1). Aunque el rango MIR (>2500 nm) y FIR serían un complemento esencial para el entrenamiento de modelos con bajo rendimiento en el rango VIS-NIR. Hay que mencionar que las diversas características de los sensores no permitieron realizar una comparación concreta de su efecto en la predicción de nutrientes fisicoquímicos, debido a que el volumen de datos con la misma variable, sensor y transformación fue muy bajo.

Tabla 1. Caracterización de los sensores utilizados en cada uno de los artículos de acuerdo con el rango y resolución espectral.

| Autores | Sensor | Rango espectral (nm) | Resolución espectral (nm) |
|---|----------------------|----------------------|---------------------------------|
| Liu et al. (2009) | ASD FIELDSPEC1 | 350 - 2500 | 1 |
| Mustafa & Moursy (2020); Santra et al. (2015); Vašát et al. (2017); Wei et al. (2020) | ASD FIELDSPEC 31 | 350 - 2500 | |
| Crucil et al. (2019) | ASD FIELDSPEC 3 FR1 | 350 - 2500 | 3 (350 - 1000) 10 (1000 - 2500) |
| Kurmai-Takács & Csorba (2015) | ASD FIELDSPEC 3 MAX1 | 350 - 2500 | 10 |
| Gaikwad et al. (2020) | ASD FIELDSPEC 41 | 350 - 2500 | |
| Lagacherie et al. (2008); He et al. (2004) | ASD PRO FR1 | 350 - 2500 | 1,4 - 2 |
| Wu et al. (2020) | GAIASORTER2 | 380 - 2500 | |
| Lagacherie et al. (2008) | HYMAP3 | 400 - 2500 | 19 |
| Sur y Chauhan (2019) | HYPERION4 | 350 - 2500 | 10 |
| Li et al. (2019) | IMSPECTOR N17E5 | 900 - 1700 | 3 |
| An et al. (2012) | MATRIX-1 FT-NIR6 | 800 - 2500 | 1 |
| Divya y Gopinathan (2019) | | 350 - 2500 | 3 (VIS), 8,5 (NIR) y 6,5 (SWIR) |
| Justin-George et al. (2020) | | 350 - 2500 | 1 |
| Hong et al. (2020) | SPECTIR LLC | 380 - 1700 | 1 |
| Crucil et al. (2019) | STS-VIS7 | 350 - 800 | 1,5 |
| | STS-NIR7 | 650 - 1100 | 1,5 |
| Lei et al. (2016) | SVC-HR-10248 | 350 - 2500 | |
| Deiss et al. (2020) | XY AUTOSAMPLER9 | 2500 - 25000 | |
| | NICOLET IS5010 | | |
| | FT- IR TENSOR 276 | 2500 - 16611,3 | 700 |

1. Analytical Spectral Devices, Boulder, CO, USA; 2. Zolix Instruments CO., LTD, Beijing; 3. Integrated Spectronics INC, Australia; 4. ATK Spacecraft Systems and Services, Minneapolis; 5. Spectral Imaging LTD., Oulu, Finland; 6. Bruker Optical Company, Germany.; 7. Ocean Optics, INC, USA; 8. Spectra Vista Corporation, USA.; 9. Pike Technologies INC., Madison, WI.; 10. Thermo Fisher Scientific INC., Waltham, MA.

3.1 Componentes del suelo reportados y su respuesta espectral

En los 19 artículos revisados, se encontraron reportadas 14 variables del suelo: carbonato de calcio (CaCO₃), permanganato de carbono oxidable (POXC), SOC, carbono orgánico total (TOC), SOM, TN, N, pH, arcilla, arena, limo, conductividad eléctrica (EC), humedad, contenido de agua en el suelo (SWC) y la respuesta espectral para la

mayoría de ellas (Tabla 2). Las variables humedad y N mostraron mayor respuesta espectral al ser transformadas. Otras como CaCO₃, arcilla, TN y SWC fueron reportadas sin transformación. Este procedimiento fue común en gran parte de los artículos revisados, aunque la identificación individual de las respuestas espectrales de los nutrientes fisicoquímicos, podría no ser suficiente para realizar predicciones acertadas.

Tabla 2. Componentes del suelo y su respuesta en el espectro electromagnético según los autores seleccionados.

| Variable del suelo | Respuesta espectral (nm) |
|--------------------|--|
| CaCO ₃ | 2275 y 2375 (Lagacherie et al., 2008) |
| Arcilla | 2143 - 2269 (Lagacherie et al., 2008); 2500 - 2647,60 y 7142,85 - 8333,33 (Deiss et al., 2020) |
| EC | R 400 - 1200, R 714-2500, 2200, FD 480 - 817, 962, Log (R) 645 - 2500 (Lei et al., 2016) |
| Humedad | CARS-SPA 695, 711, 736, 747, 767, 778 y 796 (Wu et al., 2020) |
| N | FD 400 - 2500 (Gaikwad et al., 2020) |
| TN | 940, 1050, 1100, 1200, 1300, 1550 (An et al., 2012); 975 - 1645 (Li et al., 2019) |
| pH | R 500, 350 - 2500, FD 407, 350 - 2500, Log (R) 350 - 2500 (Lei et al., 2016) y 400 - 2500 (Gaikwad et al., 2020) |
| SOC | 391,4 - 989,9 (Hong et al., 2020); 451, 520, 690, 907, 1720, 2180 y FD 2309 (Justin George et al., 2020) |
| SOM | 440 - 830, 550, 600, 650, 620 - 810, 710 (Liu et al., 2009); R 1501, 2137, Sq - R 1499, 2137, 1/R 2277, Log (R) 1504, 2149, FD - 1/R 863, 1145, SD - 1/R 672, 1740, 2222, FD - Log (R) 849, 1681, 2187, SD - Log (R) 587, 905, 1726, 2219, Sq - FD 846, 1685, 2180, Sq - SD 587, 905, 1725, 2199, FD 845, 2037, 2180, 1521, SD 529, 587, 1725 (He et al., 2004); 400 - 2500 (Gaikwad et al., 2020); MSC 497, 503, 504, SG-SD 1888, FD 1393, SD 1392, CR 479, 482, 487, R 974,1 - 1009,2, MC 974,1 - 1009,2, MA 994,8 - 1015,2, SG 993,9 - 1011,2, SG-FD 515,9 - 588, FD 541,2 - 595,5 y Log1/R 974,1 - 1006,5 (Wei et al., 2020) |
| SWC | 1400 - 1900 (Divya & Gopinathan, 2019) |

Nota. Las bandas o porciones individuales del espectro fueron separados por comas y los rangos por guiones.

MSC = Corrección de dispersión multiplicativa, SG = filtro Savitzky-Golay, FD = Primera derivada, SD = Segunda derivada, CR = Remoción Continua, R = Reflectancia, MC = Media centrada, MA = Media Móvil, Log (R) = Logaritmo de la reflectancia, CARS-SPA = algoritmo de muestreo reponderado adaptativo competitivo - algoritmo de proyecciones sucesivas, Sq = Raíz cuadrada.

3.2 Algoritmos de predicción, transformaciones y resultados para el análisis espectral del suelo

La información espectral puede requerir de transformaciones y de la implementación de algoritmos para construir modelos que aporten resultados acertados con datos no conocidos. Las técnicas, transformaciones y resultados encontrados en esta revisión, fueron:

Aprendizaje de máquina extremo (ELM): Fue utilizada por Li et al. (2019), para la predicción

de TN sobre los espectros en bruto con transformaciones como el algoritmo de proyecciones sucesivas (SPA) y la eliminación de variables no informativas (UVE). Obteniendo para todas ellas una RMSEP (raíz cuadrada del error medio de predicción) < 0,01.

RF: Fue empleada por Hong et al. (2020), para predecir SOC con transformaciones como el filtro Savitzky-Golay (filtro S-G) y la remoción continua (CR) junto con derivadas de orden fraccionario, donde reportaron R² desde 0,32 hasta 0,54, y RMSEcv (raíz cuadrada del error medio por

validación cruzada) de 0,28 hasta 0,34. Por su parte, Vašát et al. (2017), usando el filtro S-G, CR, la corrección por reflectancia máxima (CMR), la corrección de dispersión multiplicativa (MSC) y la absorbancia, obtuvieron R2 desde 0,09 hasta 0,75.

Regresión Lineal (LR): Fue aplicada para la predicción de SWC en suelos con diferente tonalidad por Divya y Gopinathan (2019), donde lograron $R^2 > 0,90$ en los 1400 nm y desde 0,52 hasta 0,68 en los 1900 nm. Concerniente a la arcilla, Lagacherie et al. (2008) obtuvieron R2 desde 0,40 hasta 0,64 y RMSEP desde 58 hasta 130 en los 2206 nm. Para la variable CaCO₃, los mismos autores usaron CR en los 2341 nm, donde obtuvieron R2 desde 0,61 hasta 0,86 y RMSEP desde 71 hasta 133.

Para la predicción de SOM, Liu et al. (2009), aplicaron el modelo lineal derivado de espectros normalizados ($R^2 = 0,91$ y $RMSE = 0,25$), el modelo lineal de curvatura de arco ($R^2 = 0,93$ y $RMSE = 0,21$) y, sobre el logaritmo de SOM, la relación de derivada de espectros normalizados exponencial ($R^2 = 0,77$ y $RMSE = 1,34$). Referente a EC, SOC, arena, limo y arcilla, Santra et al. (2015), utilizaron LR junto con análisis de componentes principales, donde alcanzaron $R^2 < 0,43$. Finalmente, Wu et al. (2020), utilizaron LR transformando los datos mediante el algoritmo de muestreo reponderado adaptativo competitivo y el SPA (CARS-SPA) consiguiendo un R2 de 0,83 y un RMSEP $< 0,01$ para humedad.

Regresión de cresta (RR): Fue utilizada por Wei et al. (2020), para predecir SOM junto con la regresión de cresta de núcleo (KRR) y la regresión de cresta bayesiana (BRR) sobre los espectros de muestras del Noroeste de China y Honghu transformados por MSC, filtro S-G, primera derivada (FD), segunda derivada (SD), CR, media centrada (MC), media móvil (MA), y absorbancia, con lo que obtuvieron R2 desde 0,77 hasta 0,90 y RMSEP desde 0,18 hasta 0,93.

MLR: Utilizando el filtro S-G, CR, CMR, MSC y absorbancia, Vašát et al. (2017), obtuvieron R2 desde 0,31 hasta 0,82 y RMSE desde 0,11 hasta 0,23 para SOC.

Regresión lineal escalonada (SLR): Se utilizó por He et al. (2004), con transformaciones logarítmicas y derivadas, donde obtuvieron R2 desde 0,67 hasta 0,88 y RMSE desde 0,36 hasta 0,60 para SOM. Por su parte Liu et al. (2009), utilizaron el espectro normalizado derivado multivariante paso a paso con el que obtuvieron R2 de 0,94 y RMSE de 0,18 sobre el logaritmo de SOM.

PLSR: Fue utilizada por el 52,60% de los autores. Entre estos, Santra et al. (2015), la usaron para predecir SOC, arena, limo y arcilla consiguiendo RMSEP desde 0,14 hasta 6,28. De igual forma, Mustafa y Moursy (2020), reportaron R2 desde 0,31 hasta 0,83 y RMSEP desde 0,27 hasta 2,72 en el conjunto de validación para pH, EC y CaCO₃. Asimismo, Kurmai-Takács y Csorba (2015), usaron esta metodología sobre los espectros brutos, absorbancia y derivadas con MSC y variante normal estándar (SNV), donde obtuvieron R2 desde 0,68 hasta 0,95 y RMSE desde 0,17 hasta 0,52 para pH.

Otros investigadores como Vašát et al. (2017), usaron PLSR junto con filtro S-G con CMR, CR, MSC, FD, SD, absorbancia y SNV para la variable SOC. Donde obtuvieron R2 desde 0,26 hasta 0,84 y RMSE desde 0,12 hasta 0,28. Por su parte, Lei et al. (2016), aplicaron CR, logaritmos y derivadas, obteniendo R2 desde 0,47 hasta 0,69 para EC y desde 0,78 hasta 0,89 para pH con $RMSE < 0,65$ en ambos casos.

La predicción de arena, arcilla, limo y SOM fue realizada por Sur y Chauhan (2019), donde obtuvieron RMSE de 7,81, 3,36, 2,32 y 0,64, respectivamente. Otros investigadores como Crucil et al. (2019), obtuvieron R2 desde 0,85 hasta 0,96 y RMSE desde 2,1 hasta 4,2 utilizando FD, CR y SNV como transformaciones para SOC. Por su parte Li et al. (2019), utilizaron la técnica sobre los espectros brutos, SPA y UVE para TN obteniendo $R^2 > 0,91$ y $RMSEP < 0,01$ en todos los casos.

Para los componentes N, SOM, y pH, Gaikwad et al. (2020), hicieron predicciones utilizando los espectros brutos donde obtuvieron R2 desde

0,63 hasta 0,69 para N, $R^2 > 0,95$ para SOM y $R^2 > 0,91$ para pH. Mediante la utilización de logaritmos y derivadas, consiguieron $R^2 > 0,89$ para SOM, valores de R^2 desde 0,59 hasta 0,99 para N y $R^2 > 0,92$ para pH.

Deiss et al. (2020), utilizaron las transformaciones filtro S-G, segmento-Gap (G-S), SNV y detrend (DT) y reportaron valores de R^2 y RMSEV (raíz cuadrada del error medio de validación), para suelos de Tanzania y Medio oeste de Estados Unidos. Para arcilla, encontraron R^2 de 0,78 y 0,82; RMSEV de 5,18 y 4,58 en ambos suelos, respectivamente. Similarmente, para pH utilizaron filtro S-G y SNV y encontraron R^2 de 0,87 y 0,80; RMSEV de 0,21 y 0,44. Así mismo, para POXC usaron promedio móvil y derivadas, donde encontraron R^2 de 0,46 y 0,74; RMSEV de 165 y 123. Para arena utilizaron filtro S-G, SNV, G-S y DT, donde encontraron R^2 de 0,81 y 0,75; RMSEV de 3,40 y 7,70. Para TOC, usaron la absorbancia, filtro S-G y DT obteniendo R^2 de 0,85 y 0,13; RMSEV de 0,29 y 0,86 para los suelos estudiados, respectivamente.

BPNN: Esta metodología fue usada por An et al. (2012), para predecir TN junto con el filtro S-G en tamaños de partícula sin modificar y desde 0,30 mm hasta 2 mm (RMSEP desde 0,10 hasta 0,15). Con el método de la derivada de primer orden obtuvieron RMSEP desde 0,06 hasta 0,16. Por medio de la absorbancia alcanzaron valores para RMSEP desde 0,07 hasta 0,16. Adicionalmente, con el método de corrección de la derivada de primer orden, consiguieron RMSEP desde 0,04 hasta 0,12.

SVM: Fue empleada por Vašát et al. (2017), con filtro S-G junto con CMR, CR, absorbancia, MSC, FD, SD y SNV, reportando R^2 desde 0,28 hasta 0,84 y RMSE desde 0,12 hasta 0,28 para SOC. Lei et al. (2016), usaron logaritmos, derivadas, espectros brutos y CR, donde consiguieron R^2 desde 0,51 hasta 0,78 y RMSE desde 0,03 hasta 0,04 para EC. Para pH reportan $R^2 > 0,9$ y RMSE desde 0,32 hasta 0,34.

Deiss et al. (2020), hicieron uso de SVM junto al filtro S-G, G-S, DT y SNV (en los mismos suelos

mencionados anteriormente), para predecir pH ($R^2 = 0,94$ y $0,84$ y RMSEV= 0,13 y 0,40), arena ($R^2 = 0,88$ y $0,76$ y RMSEV= 2,73 y 7,45), TOC ($R^2 = 0,54$ y $0,93$ y RMSEV= 0,59 y 0,18), POXC ($R^2 = 0,84$ y $0,93$ y RMSEV= 88 y 69) y arcilla ($R^2 = 0,81$ y $0,87$ y RMSEV= 4,58 y 3,81). Utilizando máquina de soporte vectorial ajustada (TSVM) junto con las mismas transformaciones, estimaron pH ($R^2 = 0,95$ y $0,87$ y RMSEV= 0,13 y 0,36), arena ($R^2 = 0,87$ en ambos casos y RMSE= 2,83 y 5,24), TOC ($R^2 = 0,57$ y $0,93$ y RMSE= 0,56 y 0,19), POXC ($R^2 = 0,84$ y $0,92$ y RMSE= 89 y 74) y arcilla ($R^2 = 0,81$ y $0,91$ y RMSE= 4,66 y 3,22).

Algoritmo AdaBoost (ABA): Fue usado por Wei et al. (2020), en muestras del Noroeste de China y Honghu, donde aplicaron transformaciones como MSC, filtro S-G, FD, SD, CR, MC, MA, y $\text{Log}(1/R)$, para predecir SOM obteniendo R^2 de 0,58 y 0,64; RMSEP de 0,36 y 0,93, respectivamente. Además, usando ABA en combinación con KRR y RR, hallaron R^2 de 0,91 y 0,86; RMSEP de 0,22 y 0,72 para cada zona.

ANN: Justin-George et al. (2020), obtuvieron $R^2 = 0,92$ y RMSE = 0,49 para SOC sin usar transformaciones.

En la Tabla 3 se evalúan los resultados por cada propiedad del suelo y procedimientos utilizados, de acuerdo con las métricas R^2 y RMSE.

Tabla 3. Número de reportes totales para cada parámetro y su evaluación de acuerdo con el ajuste alcanzado.

| Parámetro | Reportes | Métodos | | | | | | | | | | | |
|-------------------|----------|---------|------|------|------|------|------|------|------|------|------|------|------|
| | | PLSR | LR | TSVM | SVM | MLR | SLR | ANN | BPNN | RF | RR* | ABA | ELM |
| CaCO ₃ | 2 | 1(1) | 1(1) | | | | | | | | | | |
| EC | 3 | 2(0) | 1(0) | | 1(0) | | | | | | | | |
| pH | 6 | 5(5) | 1(0) | 1(1) | 2(2) | | | | | | | | |
| Arena | 3 | 3(1) | 1(0) | 1(1) | 1(1) | | | | | | | | |
| Arcilla | 4 | 3(1) | 2(0) | 1(1) | 1(1) | | | | | | | | |
| Limo | 2 | 2(0) | 1(0) | | | | | | | | | | |
| SOC | 5 | 3(3) | 1(0) | | 1(1) | 1(1) | | 1(1) | | 2(2) | | | |
| TOC | 1 | 1(1) | | 1(1) | 1(1) | | | | | | | | |
| POXC | 1 | 1(0) | | 1(1) | 1(1) | | | | | | | | |
| SOM | 5 | 2(2) | 1(1) | | | | | 1(1) | | | 5(5) | 1(1) | |
| log(SOM) | 1 | | 1(1) | | | | 1(1) | | | | | | |
| SWC | 1 | | 1(1) | | | | | | | | | | |
| Humedad | 1 | | 1(1) | | | | | | | | | | |
| N | 1 | 1(1) | | | | | | | | | | | |
| TN | 2 | 1(1) | | | | | | | 1(1) | | | | 1(1) |

Nota. Valores entre paréntesis indican resultados adecuados según R² y/o RMSE.

PLSR = regresión por mínimos cuadrados parciales, LR = regresión lineal, TSVM = máquina de soporte vectorial ajustada, SVM = máquina de soporte vectorial, MLR = regresión lineal múltiple, SLR = regresión lineal escalonada, ANN = redes neuronales artificiales, BPNN = redes neuronales de retropropagación, RF = bosque aleatorio, BRR = regresión de cresta bayesiana, KRR = regresión de cresta de núcleo, RR = regresión de cresta, ABA = algoritmo AdaBoost, ELM = aprendizaje de máquina extremo.

* RR; BRR; KRR; ABA-KRR; ABA-RR.

Teniendo en cuenta la información suministrada, se considera que las ecuaciones de búsqueda permitieron la consecución de artículos relacionados al objetivo propuesto, junto con la información necesaria para determinar la capacidad predictiva de algunas metodologías afines al ML. Se encontró que los algoritmos de aprendizaje supervisado fueron los más reportados para la predicción de propiedades del suelo, siendo los más comunes: SOM, SOC, y pH. En concreto, PLSR fue la más usada (25), aunque solo el 64% de las ocasiones tuvo resultados positivos. Por el contrario, SVM y TSVM mostraron mejor rendimiento (92,30%) en menos artículos (13).

Por otra parte, la técnica LR solo tuvo resultados positivos el 41,66% de las veces que se empleó.

Lo que muestra que no es adecuada para la construcción de modelos predictivos, a no ser que se implemente por medio de bandas individuales, como en los estudios referentes a las variables: CaCO₃, SOM, humedad y SWC. Otras metodologías como MLR, SLR, ANN, BPNN, RF, RR, BRR, KRR y ABA, obtuvieron buenos resultados, aunque no contaron con un número considerable de reportes para afirmar su efectividad. Por lo tanto, se sugiere hacer búsquedas individuales para confirmar su usabilidad.

Los resultados obtenidos en cuanto a la textura del suelo (arena, arcilla y limo) en general, mostraron que la predicción de estas variables es compleja desde aproximaciones lineales, como PLSR y LR. Debido a que solamente el 16,67% de

las aproximaciones totales, fueron acertadas. Por el contrario, cuando se emplearon SVM y TSVM para predecir arena y arcilla, todos los resultados fueron adecuados, pero, en pocas observaciones. Por estas razones, se sugiere que en futuras investigaciones se haga énfasis en métodos de predicción no lineales, como: CART, ANN y RF, entre otros (Jaconi et al., 2019).

La utilización de transformaciones sobre la reflectancia como apoyo al análisis de datos espectrales, como el filtro S-G, derivadas y logarítmicas, son necesarias y pueden verse modificadas con múltiples combinaciones de acuerdo con la respuesta espectral original y la variable de interés. Debido a esto, en el presente trabajo no fue posible realizar comparaciones entre los tipos de transformaciones, ya que difícilmente se repetía el mismo tipo entre diferentes estudios.

4. Conclusiones

Se puede afirmar que la variedad de la información encontrada por medio de las ecuaciones de búsqueda, permiten comparar parámetros edáficos de interés agrícola en la actualidad. Adicionalmente, los rangos del espectro electromagnético más utilizados están comprendidos en el rango de los 350 nm hasta los 2500 nm, ya sea completo o con porciones menores. Por otra parte, los rangos MIR (>2500 nm) y FIR, no son utilizados en gran proporción de las aproximaciones encontradas en este estudio.

Los algoritmos PLSR, SVM y TSVM son adecuados para predecir diversas variables del suelo como pH, SOM y SOC. Asimismo, la LR puede ser una opción viable, para la predicción de variables como CaCO₃, SOM, humedad y SWC, haciendo uso de bandas espectrales con mayor respuesta. Sin embargo, la LR posee un rendimiento bajo generalizado para los demás parámetros. Otros algoritmos como MLR, SLR, ANN, BPNN, RF, RR, BRR, KRR y ABA, requieren de mayor cantidad de investigaciones para evaluar su efectividad predictiva.

De acuerdo con los resultados encontrados para los componentes de la variable textura, las téc-

nicas como PLSR y LR podrían no ser adecuadas para la construcción de modelos predictivos. Sin embargo, el uso de diversas transformaciones en el entrenamiento y evaluación de modelos por medio de técnicas no lineales, posiblemente sean una opción para obtener mejores ajustes en las predicciones de validación.

Por último, se puede decir que gradualmente los diferentes tipos de sensores expuestos tendrán mejoras en la resolución espectral, con la incorporación de los rangos MIR y posiblemente FIR. Esto debido a que, con su implementación, se han encontrado resultados alentadores para la predicción de variables de difícil determinación. Adicionalmente, la investigación y desarrollo de dispositivos con alta capacidad computacional facilitará el manejo de información espectral e hiperespectral, por medio del avance en algoritmos pertenecientes al campo de la inteligencia artificial como el ML y el aprendizaje profundo.

Agradecimientos

Los autores agradecen el apoyo económico brindado al proyecto del Sistema General de Regalías: "Desarrollo y establecimiento del Centro de Desarrollo Agrobiotecnológico de Innovación e Integración Territorial, el Carmen de Viboral, Antioquia, Occidente (CEDAIT)", cód. BPIN 2016000100060.

Referencias

- Ahmadi, A., Emami, M., Daccache, A., & He, L. (2021). Soil properties prediction for precision agriculture using visible and near-infrared spectroscopy: A systematic review and meta-analysis. *Agronomy*, 11 (433), 1-14. <https://doi.org/10.3390/agronomy11030433>
- Aliyu, M. B. (2017). Efficiency of Boolean search strings for information retrieval. *American Journal of Engineering Research*, 6 (11), 216-222.
- An, X., Li, M., Zheng, L., & Liu, Y. (2012). Effect of particle size on prediction of soil TN with

- remote sensing based on NIR spectroscopy. Multispectral, Hyperspectral, and Ultra spectral Remote Sensing Technology, Techniques and Applications IV. *International Society for Optics and Photonics*, 8527, 852710-1 - 852710-7. <https://doi.org/10.1117/12.977349>
- Angelopoulou, T., Tziolas, N., Balafoutis, A., Zalidis, G., & Bochtis, D. (2019). Remote sensing techniques for soil organic carbon estimation: A review. *Remote Sensing*, 11 (6), 1-18. <https://doi.org/10.3390/rs11060676>
- Barra, I., Haefele, S. M., Sakrabani, R., & Kebede, F. (2021). Soil spectroscopy with the use of chemometrics, machine learning and pre-processing techniques in soil diagnosis: Recent advances—A review. *Trends in Analytical Chemistry*, 135, 1-13. <https://doi.org/10.1016/j.trac.2020.116166>
- Benos, L., Tagarakis, A. C., Dolias, G., Berruto, R., Kateris, D., & Bochtis, D. (2021). Machine learning in agriculture: A comprehensive updated review. *Sensors*, 21 (11), 1-55. <https://doi.org/10.3390/s21113758>
- Burton, L., Jayachandran, K., & Bhansali, S. (2020). The “Real-Time” revolution for in situ soil nutrient sensing. *Journal of The Electrochemical Society*, 167 (3), 1-9. <https://doi.org/10.1149/1945-7111/ab6f5d>
- Crucil, G., Castaldi, F., Aldana, E., van Wesemael, B., Macdonald, A., & Van Oost, K. (2019). Assessing the Performance of UAS-Compatible Multispectral and Hyperspectral Sensors for Soil Organic Carbon Prediction. *Sustainability*, 11 (7), 1-18. <https://doi.org/10.3390/su11071889>
- Dangal, S. R., Sanderman, J., Wills, S., & Ramirez-Lopez, L. (2019). Accurate and precise prediction of soil properties from a large mid-infrared spectral library. *Soil Systems*, 3 (1), 11. <https://doi.org/10.3390/soilsystems3010011>
- Deiss, L., Margenot, A. J., Culman, S. W., & Demyan, M. S. (2020). Tuning support vector machines regression models improves prediction accuracy of soil properties in MIR spectroscopy. *Geoderma*, 365, 1-12. <https://doi.org/10.1016/j.geoderma.2020.114227>
- Demattê, J. A., Alves, M. R., Gallo, B. C., Fongaro, C. T., Romero, D. J., & Sato, M. V. (2015). Hyperspectral remote sensing as an alternative to estimate soil attributes. *Revista Ciência Agronômica*, 46, 223-232. <https://doi.org/10.5935/1806-6690.20150001>
- Demattê, J. A., Dotto, A. C., Paiva, A. F., Sato, M. V., Dalmolin, R. S., Maria do Socorro, B., & do Couto, H. T. Z. (2019). The Brazilian Soil Spectral Library (BSSL): A general view, application and challenges. *Geoderma*, 354, 1-21. <https://doi.org/10.1016/j.geoderma.2019.05.043>
- Dhawale, N. M., Adamchuk, V. I., Prasher, S. O., & Viscarra Rossel, R. A. (2021). Evaluating the Precision and Accuracy of Proximal Soil vis-NIR Sensors for Estimating Soil Organic Matter and Texture. *Soil Systems*, 5(3), 1-19.
- Diaz-Gonzalez, F. A., Vuelvas, J., Correa, C. A., Vallejo, V. E., & Patino, D. (2022). Machine learning and remote sensing techniques applied to estimate soil indicators—Review. *Ecological Indicators*, 135, 1-12. <https://doi.org/10.1016/j.ecolind.2021.108517>
- Divya, Y., & Gopinathan, P. (2019). Soil water content measurement using hyper-spectral remote sensing techniques - A case study from north-western part of Tamil Nadu, India. *Remote Sensing Applications: Society and Environment*, 14, 1-7. <https://doi.org/10.1016/j.rsase.2019.01.005.45739>
- Fang, Q., Hong, H., Zhao, L., Kukolich, S., Yin, K., & Wang, C. (2018). Visible and near-infrared reflectance spectroscopy for investigating soil mineralogy: A review. *Journal of Spectroscopy*, 2018, 1-14. <https://doi.org/10.1155/2018/3168974>
- Gaikwad, C. M., Kakarwal, S. N., Sapkal, S., & Uke, N. (2020). Estimation of Soil Attributes using Spectroscopy. *CSI Journal of Computing*, 3 (2), 12-15.
- Gourlay, S., Aynekulu, E., Carletto, C., & Shepherd, K. (2017). Spectral Soil Analysis & Household Surveys: *Guidebook for Integration*. <https://cgspace.cgiar.org/handle/10568/99331>

- Guerrero, A., De Neve, S., & Mouazen, A. M. (2021). Current sensor technologies for in situ and on-line measurement of soil nitrogen for variable rate fertilization: A review. *Advances in Agronomy*, 168, 1-38. <https://doi.org/10.1016/bs.agron.2021.02.001>
- Guo, L., Shi, T., Linderman, M., Chen, Y., Zhang, H., & Fu, P. (2019). Exploring the influence of spatial resolution on the digital mapping of soil organic carbon by airborne hyperspectral VNIR imaging. *Remote Sensing*, 11 (9), 1-16. <https://doi.org/10.3390/rs11091032>
- He, T., Wang, J., Guo, X., & Chen, Y. (2004). Study on applying hyperspectral remote sensing technology in land quality monitoring. *Remote Sensing for Environmental Monitoring, GIS Applications and Geology IV. International Society for Optics and Photonics*, 5574, 410-419. <https://doi.org/10.1117/12.565987>
- Hong, Y., Chen, S., Zhang, Y., Chen, Y., Yu, L., Liu, Y., ... & Liu, Y. (2018). Rapid identification of soil organic matter level via visible and near-infrared spectroscopy: Effects of two-dimensional correlation coefficient and extreme learning machine. *Science of the Total Environment*, 644, 1232-1243. <https://doi.org/10.1016/j.scitotenv.2018.06.319>
- Hong, Y., Guo, L., Chen, S., Linderman, M., Mouazen, A. M., Yu, L., Chen, Y., Liu, Y., Liu, Y., Cheng, H., & Liu, Y. (2020). Exploring the potential of airborne hyperspectral image for estimating topsoil organic carbon: Effects of fractional-order derivative and optimal band combination algorithm. *Geoderma*, 365, 1-12. <https://doi.org/10.1016/j.geoderma.2020.114228>
- Inoue, Y. (2020). Satellite-and drone-based remote sensing of crops and soils for smart farming—a review. *Soil Science and Plant Nutrition*, 66 (6), 798-810. <https://doi.org/10.1080/00380768.2020.1738899>
- Jaconi, A., Vos, C., & Don, A. (2019). Near infrared spectroscopy as an easy and precise method to estimate soil texture. *Geoderma*, 337, 906-913.
- Justin-George, K., Kumar, S., & Raj, R. A. (2020). Soil organic carbon prediction using visible-near infrared reflectance spectroscopy employing artificial neural network modelling. *Current Science*, 119 (2), 377-381.
- Kurmai-Takács, Z., & Csorba, Á. (2015). A reflektancia spektroszkópia alkalmazása talajok kémhatásának becslésében. *Journal of Central European Green Innovation*, 3 (3), 113-132.
- Kurz, T. H., & Buckley, S. J. (2016). A review of hyperspectral imaging in close range applications. XXIII ISPRS The International Archives of the Photogrammetry, *Remote Sensing and Spatial Information Sciences Congress, XLI-B5*, 865-870.
- Lagacherie, P., Baret, F., Feret, J. B., Netto, J. M., & Robbez-Masson, J. M. (2008). Estimation of soil clay and calcium carbonate using laboratory, field and airborne hyperspectral measurements. *Remote Sensing of Environment*, 112 (3), 825-835. <https://doi.org/10.1016/j.rse.2007.06.014.44133>
- Lei, S., Bao, N., Liu, S., & Liu, X. (2016). Modelling and Predicting of Soil Electrical Conductivity and PH from Semi-arid Grassland Using VIS-NIR Spectroscopy Technology. *International Conference on Computer and Computing Technologies in Agriculture*, 442-453.
- Li, H., Jia, S., & Le, Z. (2019). Quantitative Analysis of Soil Total Nitrogen Using Hyperspectral Imaging Technology with Extreme Learning Machine. *Sensors*, 19 (20), 1-13. <https://doi:10.3390/s19204355.47245>
- Liu, H., Zhang, Y., & Zhang, B. (2009). Novel hyperspectral reflectance models for estimating black-soil organic matter in Northeast China. *Environmental monitoring and assessment*, 154 (1), 147-154. <https://doi:10.1007/s10661-008-0385-4>
- Liu, S., Shen, H., Chen, S., Zhao, X., Biswas, A., Jia, X., ... & Fang, J. (2019). Estimating forest soil organic carbon content using vis-NIR spectroscopy: Implications for large-scale soil carbon spectroscopic assessment. *Geoderma*, 348, 37-44.
- Lu, B., Dao, P. D., Liu, J., He, Y., & Shang, J. (2020). Recent advances of hyperspectral imaging tech-

- nology and applications in agriculture. *Remote Sensing*, 12 (16), 1-44. <https://doi.org/10.3390/rs12162659>
- Malmir, M., Tahmasbian, I., Xu, Z., Farrar, M. B., & Bai, S. H. (2019). Prediction of soil macro-and micro-370elements in sieved and ground air-dried soils using laboratory-based hyperspectral imaging371 technique. *Geoderma*, 340, 70-80. <https://doi.org/10.1016/j.geoderma.2018.12.049>
- Mustafa, A. E. R. A., & Moursy, A. R. (2020). *Using a multivariate regression model and hyperspectral reflectance data to predict soil parameters of Agra, India. International Journal of Geography, Geology and Environment*, 2 (1), 4-9. <https://doi.org/10.22271/27067483.2020.v2.i1a.12>
- Nawar, S., & Mouazen, A. M. (2018). Optimal sample selection for measurement of soil organic carbon using on-line vis-NIR spectroscopy. *Computers and Electronics in Agriculture*, 151, 469-477.
- Olatunde, K. A. (2021). Estimation of soil organic carbon using chemometrics: a comparison between mid-infrared and visible near infrared diffuse reflectance spectroscopy. *West African Journal of Applied Ecology*, 29 (2), 1-11.
- Padarian, J., Minasny, B., & McBratney, A. B. (2020). Machine learning and soil sciences: A review aided by machine learning tools. *Soil*, 6 (1), 35-52. <https://doi.org/10.5194/soil-6-35-2020>
- Qi, H., Paz, T., Karnieli, A., Jin, X., & Li, S. (2018). Evaluating calibration methods for predicting soil available nutrients using hyperspectral VNIR data. *Soil and Tillage Research*, 175, 267-275. <https://doi.org/10.1016/j.still.2017.09.006.447>
- Santra, P., Singh, R., Sarathjith, M., Panwar, N., Varghese, P., & Das, B. (2015). Reflectance spectroscopic approach for estimation of soil properties in hot arid western Rajasthan, India. *Environmental Earth Sciences*, 74 (5), 4233-4245. <https://doi.org/10.1007/s12665-015-4383-x>
- Shi, P., Castaldi, F., van Wesemael, B., & Van Oost, K. (2020). Vis-NIR spectroscopic assessment of soil aggregate stability and aggregate size distribution in the Belgian Loam Belt. *Geoderma*, 357, 1-9. <https://doi.org/10.1016/j.geoderma.2019.113958>
- Song, Y. Q., Zhao, X., Su, H. Y., Li, B., Hu, Y. M., & Cui, X. S. (2018). Predicting Spatial Variations in Soil Nutrients with Hyperspectral Remote Sensing at Regional Scale. *Sensors*, 18 (9), 1-18. <https://doi:42410.3390/s18093086.42526>
- Sur, K., & Chauhan, P. (2019). Imaging spectroscopic approach for land degradation studies: a case study from the arid land of India. *Geomatics, Natural Hazards and Risk*, 10 (1), 898-911. <https://doi:51710.1080/19475705.2018.1552629.51864>
- Vašát, R., Kodešová, R., & Borvka, L. (2017). Ensemble predictive model for more accurate soil organic carbon spectroscopic estimation. *Computers & Geosciences*, 104, 75-83. <https://doi.org/10.1016/j.cageo.2017.04.008>
- Vohland, M., Ludwig, M., Sören, T.B., & Ludwig, B. (2017). Quantification of Soil Properties with Hyperspectral Data: Selecting Spectral Variables with Different Methods to Improve Accuracies and Analyze Prediction Mechanisms. *Remote Sensing*, 9 (11), 1-24. <https://doi:10.3390/rs9111103.40719>.
- Wei, L., Yuan, Z., Wang, Z., Zhao, L., Zhang, Y., Lu, X., & Cao, L. (2020). Hyperspectral inversion of soil organic matter content based on a combined spectral index model. *Sensors*, 20 (10), 1-17. <https://doi.org/10.3390/s20102777>
- Wu, T., Yu, J., Lu, J., Zou, X., & Zhang, W. (2020). Research on inversion model of cultivated soil moisture content based on hyperspectral imaging analysis. *Agriculture*, 10 (7), 1-14. <https://doi.org/10.3390/agriculture10070292>
- Yunus, Y., Satriyo, P., & Munawar, A. A. (2019). Rapid Prediction of Soil Quality Indices Using Near Infrared Spectroscopy. *IOP Conference Series: Earth and Environmental Science*, 365, 1-7. IOP Publishing. <https://doi.org/10.1088/1755-1315/365/1/012043>