19th Iranian Soil Science Congress
2-4 December, 2025

نوزدهمین کنگره علوم خاک ایران
۱۱ تا ۱۳ آذرماه ۱۴۰۴

مدیریت جامع‌نگر و هوشمند خاک و آب
Holistic and Smart Soil and Water Management
دانشکدگان کشاورزی و منابع طبیعی دانشگاه تهران
College of Agriculture & Natural Resources, University of Tehran

۰۴۲۵۰-۳۲۰۳۱

# مروری بر تحلیل چندمقیاسی داده‌های خاک با هوش مصنوعی و یادگیری ماشین

**احسان قزلباش[1*]**

۳- استادیار گروه علوم و فناوری خاک، دانشکده کشاورزی، دانشگاه تربیت مدرس، تهران، ایران*.
**ghezelbash@modares.ac.ir**

## چکیده

در دهه‌های اخیر، داده‌محوری به یکی از ارکان اصلی علوم خاک بدل شده است. هم‌زمان با افزایش حجم و تنوع داده‌های خاک (از داده‌های زیستی در مقیاس ریزوسفر تا تصاویر ماهواره‌ای در مقیاس منطقه‌ای)، نیاز به روش‌های نوین تحلیل بیش از پیش احساس می‌شود. هوش مصنوعی[1] و یادگیری ماشین[2] (AI/ML) از اوایل دهه ۱۹۹۰ میلادی به بعد رشد چشمگیری در این حوزه داشته‌اند و اکنون ابزارهای کلیدی در ترسیم نقشه‌های دیجیتال خاک، توسعه توابع انتقالی، پایش نزدیک و دورسنجی، و حتی تحلیل متون علمی خاک‌شناسی محسوب می‌شوند. مرور حاضر با هدف بررسی نظام‌مند کاربردهای اصلی AI/ML در علوم خاک طی بازه ۱۹۹۰ تا ۲۰۲۵ انجام شد. نتایج نشان داد که الگوریتم‌هایی نظیر جنگل تصادفی (RF)، ماشین بردار پشتیبان (SVM) و شبکه‌های عصبی عمیق (DNN) بیشترین کاربرد را در مقیاس مزرعه و منطقه‌ای داشته‌اند، در حالی که فناوری‌های نوین پایش (طیف‌سنجی، تصویربرداری CT و سنجش‌ازدور چندزمانه) فرصت‌های تازه‌ای برای تحلیل داده‌های خاک فراهم کرده‌اند. چالش‌هایی همچون نیاز به شفافیت در ارزیابی مدل‌ها، کمبود داده‌های استاندارد، و ضرورت تلفیق دانش خاک‌شناسی با رویکردهای داده‌محور همچنان باقی است. این مقاله با معرفی چارچوبی مفهومی برای پیوند مقیاس‌پذیری و فناوری‌های نوین، مسیرهای آینده شامل بهره‌گیری از مدل‌های زبانی بزرگ و هوش ترکیبی را نیز ترسیم می‌کند.

**واژگان کلیدی:** داده‌محوری، چند مقیاسی، مدل‌های زبانی بزرگ، هوش مصنوعی تبیین‌پذیر.

1 - Artificial Inteligense
2 - Machine Learning

**مقدمه**

علوم خاک در دهه‌های اخیر با تحولات اساسی در تولید و گردآوری داده‌ها روبه‌رو بوده است. حجم و تنوع داده‌های حاصل از سنجش‌های آزمایشگاهی، حسگرهای نزدیک، سامانه‌های دورسنجی و حتی داده‌های زیستی ریزوسفر ریشه، به‌گونه‌ای افزایش یافته که روش‌های سنتی تحلیل دیگر پاسخ‌گوی نیازهای پژوهشی و مدیریتی نیستند (Arrouays et al., 2020; Heuvelink et al., 2021). در چنین شرایطی، داده‌محوری به یک رویکرد غالب در خاک‌شناسی بدل شده و مسیر توسعه بسیاری از حوزه‌های نوین این علم را تعیین کرده است (Minasny & McBratney, 2016).

از حدود سال ۲۰۱۰ به بعد، ورود هوش مصنوعی و یادگیری ماشین به عرصه علوم خاک جهشی بزرگ در توان تحلیل داده‌ها ایجاد کرد. الگوریتم‌هایی همچون RF، SVM، و DNN در مدت کوتاهی جایگاه ویژه‌ای در ترسیم نقشه‌های دیجیتال خاک (DSM)، توسعه توابع انتقالی خاک (PTFs)، و پردازش داده‌های سنجش‌ازدور پیدا کردند (Padarian et al., 2020; Tziolas et al., 2021; Wu et al., 2025). بررسی منابع نشان می‌دهد که هوش مصنوعی نه‌تنها به بهبود دقت مدل‌ها کمک کرده، بلکه توانسته است کارایی روش‌های کلاسیک را نیز ارتقا دهد (Viscarra Rossel et al., 2019; Wadoux et al., 2025).

همچنین، کاربرد فناوری‌های نوین مانند طیف‌سنجی درجا، تصویربرداری سه‌بعدی و داده‌های چندزمانه ماهواره‌ای فرصت‌های تازه‌ای برای تحلیل داده‌های خاک فراهم کرده است (Lobsey & Viscarra Rossel, 2016; Taghizadeh-Mehrjardi et al., 2020). در سال‌های اخیر، گسترش رویکردهایی چون یادگیری عمیق و مدل‌های زبانی بزرگ (LLMs[۳]) نشان می‌دهد که روند بهره‌گیری از AI/ML در خاک‌شناسی نه‌تنها ادامه دارد بلکه در حال شتاب گرفتن است (Minasny & McBratney, 2025; Wadoux et al., 2025). هدف

جدول (۱) روند تاریخی کاربرد هوش مصنوعی و یادگیری ماشین در علوم خاک را از پیش از دهه ۱۹۹۰ تا سال ۲۰۲۵ نشان می‌دهد. همان‌طور که مشخص است، استفاده از شبکه‌های عصبی و سیستم‌های خبره در دهه‌های ابتدایی محدود و پراکنده بوده، اما از اواسط دهه ۱۹۹۰ و به ویژه پس از سال ۲۰۰۰ با افزایش قدرت محاسبات و دسترسی به داده‌ها، یادگیری ماشین به‌طور چشمگیری رشد کرده است. پس از ۲۰۱۰، کاربرد یادگیری عمیق و الگوریتم‌های پیشرفته شدت گرفت و پژوهش‌های پایش خاک با محوریت سنجش از دور در دوره ۲۰۱۹–۲۰۲۱ به سطح بالغ رسیدند. در سال‌های اخیر، حجم بالای انتشارات و بررسی‌های نظام‌مند نشان می‌دهد که هوش مصنوعی به یک ابزار کلیدی و گسترده در تحلیل و مدل‌سازی ویژگی‌های خاک تبدیل شده است. این روند تاریخی پایه‌ای برای درک اهمیت داده‌محوری و هوش مصنوعی در علوم خاک و توسعه مدل‌های نوین مدیریت خاک و آب فراهم می‌کند.

**جدول ۱- روند تاریخی کاربرد هوش مصنوعی و یادگیری ماشین در علوم خاک از ۱۹۹۰ تا ۲۰۲۵**

| دوره تقریبی | روند مشاهده‌شده در انتشارات | زمینه کلیدی (منبع) |
|---|---|---|
| پیش از ۱۹۹۰ | کاربرد اولیه و پراکنده شبکه‌های عصبی اولیه و سیستم‌های خبره | کاربردهای شبکه عصبی از حدود ۱۹۹۰ آغاز شد (Minasny and McBratney, 2025). اوج کاربرد سیستم‌های خبره در اواخر دهه ۱۹۸۰ تا اوایل ۱۹۹۰ بود (Wadoux, 2025; Minasny and McBratney, 2025) |
| اواسط دهه ۱۹۹۰ | افزایش جزئی در کاربردهای خاص شبکه‌های عصبی (مانند توابع پدوترانسفر) | کاربرد شبکه‌های عصبی به‌طور خاص از حدود سال ۱۹۹۶ مورد توجه قرار گرفت و بر پیش‌بینی ویژگی‌های هیدرولیکی خاک متمرکز شد (Minasny and McBratney, 2025). |
| پس از ۲۰۰۰ | شتاب در پذیرش یادگیری ماشین، هم‌زمان با ظهور نقشه‌برداری رقومی خاک | محبوبیت یادگیری ماشین به‌طور چشمگیری به دلیل افزایش قدرت محاسبات و در دسترس بودن داده‌ها افزایش یافت (Minasny and McBratney, 2025). |
| پس از ۲۰۱۰ | شتاب شدید، به‌ویژه در کاربردهای یادگیری عمیق | تعداد انتشارات ML از سال ۲۰۱۵ تقریباً هفت برابر افزایش یافته است (Wadoux, 2025). |

۳ - Large Language Models

| دوره تقریبی | روند مشاهده‌شده در انتشارات | زمینه کلیدی (منبع) |
|---|---|---|
| | | کاربرد شبکه‌های عصبی از سال ۲۰۱۰ افزایش یافته است (Minasny and McBratney, 2025). |
| ۲۰۱۹–۲۰۲۱ | پژوهش‌های پایش خاک با محوریت سنجش از دور به سطحی بالغ رسید | ۴۶ مطالعه متمرکز بر سنجش از دور/هوش مصنوعی بررسی شدند (Tziolas et al., 2021). |
| سال‌های اخیر (۲۰۲۵) | حجم بالای مستمر فعالیت‌های پژوهشی | بررسی هدفمند صورت گرفته ۱۲۱۲ مقاله مرتبط با هوش مصنوعی در علوم خاک را شناسایی کرد (Wadoux, 2025). |

این مقاله، مرور تحلیلی-اجمالی[۴] کاربردهای هوش مصنوعی و یادگیری ماشین در علوم خاک طی بازه ۱۹۹۰ تا ۲۰۲۵ است. نوآوری اصلی آن، تقسیم‌بندی تحلیلی بر اساس مقیاس (ریزوسفر، مزرعه، منطقه‌ای) و نوع داده (از اومیکس[۵] تا سنجش‌ازدور) و همچنین بررسی چالش‌های نوین نظیر شفافیت مدل‌ها (Explainable AI) و نقش هوش ترکیبی در آینده خاک‌شناسی است.

## مواد و روش‌ها

این پژوهش شامل مروری تحلیلی است که با هدف شناسایی و تحلیل روندهای نوین در به‌کارگیری هوش مصنوعی و یادگیری ماشین در علوم خاک انجام شد. بازه زمانی ۱۹۹۰ تا ۲۰۲۵ انتخاب گردید تا هم تحولات آغازین چند دهه اخیر در نقشه‌برداری دیجیتال خاک و مدل‌سازی توابع انتقالی پوشش داده شود و هم آخرین دستاوردهای مربوط به یادگیری عمیق و مدل‌های زبانی بزرگ لحاظ گردد.

### راهبرد جستجو

منابع علمی از پایگاه‌های معتبر ازجمله Web of Science، Scopus، ScienceDirect، و Google Scholar استخراج شدند. کلیدواژه‌های اصلی شامل ترکیب‌هایی از واژه‌های Artificial Intelligence, Machine Learning, Deep Learning, Digital Soil Mapping, Remote Sensing, Explainable AI, Large Language Models, Soil Data, و معادل فارسی آن‌ها بود.

### معیارهای انتخاب منابع

انتخاب منابع بر اساس سه معیار اصلی شامل الف) مقالات مروری، که چارچوب‌های مفهومی یا تحولات تاریخی را تشریح کرده‌اند (Padarian et al., 2020; Wadoux et al., 2025)، ب) مقالات پژوهشی کلیدی، که الگوریتم‌ها یا فناوری‌های نوین را معرفی یا ارزیابی کرده‌اند (Tziolas et al., 2021; Wu et al., 2025)، ج) گزارش‌ها و اسناد مرجع شامل پروژه‌های جهانی پایش خاک و بانک‌های داده (Arrouays et al., 2020; Heuvelink et al., 2021) بود.

### شیوه تحلیل

برای تجمیع محتوای منابع و استخراج محورهای مشترک، علاوه بر تحلیل دستی، از ابزار هوش مصنوعی NoteBookLM استفاده شد که امکان سازمان‌دهی، خلاصه‌سازی و مقایسه منظم یافته‌ها را فراهم می‌کند. بدین‌ترتیب، خروجی نهایی ترکیبی از بررسی تحلیلی پژوهشگر و توان پردازشی AI در مدیریت داده‌های متنی بوده است. منابع انتخاب‌شده بر اساس مقیاس کاربرد (ریزوسفر، مزرعه، منطقه‌ای)، نوع داده (omics، DSM، PSS، EO) و الگوریتم غالب (RF، SVM، NN، DL، LLMs) طبقه‌بندی شدند. سپس برای هر گروه، روندهای تاریخی، نقاط قوت، چالش‌ها و چشم‌انداز آینده تحلیل گردید.

## نتایج و بحث

### برایند و دستاورد پژوهش‌ها در مقیاس ریزوسفر

۴ - Analytical-Synthetic

5 - omics: In biology the word omics refers to the sum of constituents within a cell

در مقیاس میکرو، ظهور فناوری‌های Multi-omics (متاژنومیک، متابولومیک، پروتئومیک) امکان تحلیل جامع‌تر تعاملات میکروبی خاک را فراهم کرده است (Jansson & Hofmockel, 2018). این داده‌ها حجم بسیار بالایی داشته و اغلب دارای ویژگی‌های پراکنده، و غیرخطی هستند. به همین دلیل، الگوریتم‌های یادگیری ماشین مانند شبکه‌های عصبی عمیق و مدل‌های گرافی برای کشف الگوهای پیچیده به‌کار گرفته شده‌اند (Yu et al., 2021). چالش اصلی در این مقیاس، کمبود داده‌های استاندارد و هزینه‌بر بودن آزمایش‌ها است. در سال‌های اخیر، ترکیب داده‌های اومیکس با مدل‌های خاک‌شناسی سنتی، راهکاری برای افزایش قابلیت تفسیر و کارایی معرفی شده است (Levy-Booth et al., 2020).

## برآیند و دستاورد پژوهش‌ها در مقیاس مزرعه

یکی از پرتکرارترین حوزه‌های کاربرد AI/ML در علوم خاک، نقشه‌برداری دیجیتال خاک (DSM) است که از سال ۱۹۹۰ به بعد رشد چشمگیری یافته (Arrouays et al., 2014; Wadoux et al., 2025). در این حوزه، الگوریتم‌های Random Forest (RF) و Support Vector Machines (SVM) عملکرد بالایی در پیش‌بینی خصوصیات خاک داشته‌اند (Heung et al., 2016). در سال‌های اخیر، شبکه‌های عصبی عمیق (NN, DL) نیز برای استخراج ویژگی‌های پیچیده از داده‌های مکانی و سنجش‌ازدور به‌کار رفته‌اند (Padarian et al., 2020). البته در کنار DSM، توسعه توابع انتقالی خاک (Pedotransfer Functions, PTFs) هم یکی از مهم‌ترین کاربردها بوده است. توابع انتقالی با استفاده از یادگیری ماشین دقت بیشتری در برآورد پارامترهای هیدرولیکی خاک نسبت به مدل‌های رگرسیونی کلاسیک نشان داده‌اند (Tziolas et al., 2021). چالش اصلی این حوزه، انتقال‌پذیری محدود مدل‌ها بین مناطق مختلف است که باعث نیاز به رویکردهای transfer learning شده است (Wu et al., 2025).

## فناوری‌های نوین پایش

حسگرهای نزدیک‌برد (PSS[6]) نقش کلیدی در تولید داده‌هایی با جزئیات دقیق برای مدل‌سازی خاک ایفا می‌کنند. فناوری‌هایی مانند طیف‌سنجی مادون‌قرمز (NIR/MIR)، تصویربرداری CT و پردازش تصویر به کمک الگوریتم‌های یادگیری ماشین توانسته‌اند دقت برآورد توزیع اندازه منافذ، کربن آلی و رطوبت خاک را افزایش دهند (Rossel & Behrens, 2010; Dangal et al., 2019).

در مقیاس وسیع‌تر، سنجش‌ازدور با داده‌های ماهواره‌ای (Landsat، Sentinel) و اخیراً تصاویر ابرطیفی و چندزمانه ابزار اصلی پایش تغییرات خاک و ویژگی‌های سطحی بوده است. یادگیری ماشین در این زمینه کمک کرده است تا روابط غیرخطی پیچیده بین بازتاب طیفی و خصوصیات خاک استخراج شود (Mulder et al., 2011; Zhang et al., 2022).

## چالش‌های نوین

با وجود دستاوردهای چشمگیر، استفاده گسترده از AI/ML در علوم خاک با چالش‌هایی همراه است. نخستین چالش، قابلیت تفسیرپذیری مدل‌ها است. بسیاری از الگوریتم‌های دقیق مانند شبکه‌های عمیق، به دلیل فقدان داده لازم برای ارزیابی نتایج "جعبه سیاه" باقی می‌مانند و در تصمیم‌گیری‌های مدیریتی اعتمادپذیری محدودی دارند (Arrieta et al., 2020). بنابراین، یکی از راهکارها برای تکیه بر نتایج حاصل از روش‌های هوشمند، تلفیق دانش خاک‌شناسی کلاسیک با مدل‌های داده‌محور است. مطالعات اخیر تأکید دارند که صرف اتکا به داده کافی نیست و باید مفاهیم فیزیکی و فرآیندی خاک در کنار مدل‌های یادگیری ماشین به‌کار گرفته شوند تا پایداری و قابلیت انتقال مدل افزایش یابد (Padarian et al., 2020; Wadoux et al., 2025). در نهایت، ظهور هوش معنایی[7] و مدل‌های زبانی بزرگ چشم‌انداز جدیدی برای تحلیل ویژگی‌های خاکی گشوده است. این فناوری‌ها می‌توانند در دسته‌بندی دستاوردها، استخراج دانش از متون، و حتی تولید خودکار PTFهای مبتنی بر داده‌های متنی به کار روند، هرچند همچنان در مراحل آغازین توسعه هستند (Biswas et al., 2023).

جدول (۲) خلاصه‌ای از کاربردهای اصلی هوش مصنوعی و یادگیری ماشین در علوم خاک را ارائه می‌دهد. هر ردیف حوزه کاربرد مشخص، مقیاس مورد استفاده، نوع داده ورودی، الگوریتم‌های کلیدی ML/AI، و منابع مرجع مرتبط را نشان می‌دهد. مشاهده

6 - Proximal Soil Sensing
7 - Semantic Intelligence

می‌شود که الگوریتم‌هایی مانند جنگل تصادفی (RF) و شبکه‌های عصبی کانولوشنی یا پیچشی[8] (CNN) در پردازش داده‌های پیچیده و چندبعدی خاک، به‌ویژه در مقیاس مزرعه و منطقه‌ای، کاربرد گسترده‌ای یافته‌اند. همچنین روند استفاده از مدل‌های زبانی بزرگ و تحلیل معنایی داده‌های متنی برای تجمیع دانش در حوزه کیفیت خاک، نشان‌دهنده توسعه رویکردهای داده‌محور فراتر از داده‌های کمی و سنجش از دور است. جدول (۲) همچنین ارتباط میان نوع داده، مقیاس پژوهش و الگوریتم‌های انتخاب شده را نشان می‌دهد و به عنوان نقشه راهی برای پژوهشگران جدید در زمینه هوش مصنوعی خاک می‌تواند مورد استفاده قرار گیرد.

**جدول ۲- کاربردهای اصلی هوش مصنوعی و یادگیری ماشین در علوم خاک**

| منابع | الگوریتم کلیدی ML/AI | نوع داده | مقیاس | حوزه کاربرد |
|---|---|---|---|---|
| Heuvelink et al., 2021; Sothe et al., 2022 | جنگل تصادفی / (RF) روش‌های Ensemble | سنجش از دور (چندطیفی، ابرطیفی) + متغیرهای محیطی | منطقه‌ای تا جهانی | نقشه‌برداری رقومی خاک (DSM) برای کربن آلی خاک |
| Schaap et al., 1998; Twarakavi et al., 2009 | شبکه‌های عصبی مصنوعی / (ANN) ماشین بردار پشتیبان(SVM) | ویژگی‌های پایه خاک (با بافت، وزن مخصوص ظاهری) | محلی / افق خاک | توابع پدوترانسفر (PTF) |
| Lavrukhin et al., 2021; Ferreira et al., 2022 | یادگیری عمیق / (CNN) بینایی کامپیوتر | تصاویر پرتونگاری کامپیوتری (CT) | مقیاس حفره‌ای (نمونه آزمایشگاهی) | مشخصه‌یابی غیرمخرب ساختار خاک |
| Viscarra Rossel & Behrens, 2010; Ng et al., 2020 | یادگیری عمیق / (CNN) الگوریتم‌های متنوع ML | طیف‌سنجی (مرئی-مادون قرمز نزدیک، میانی) | مزرعه / نمونه آزمایشگاهی | پردازش داده‌های سنجش از نزدیک |
| Bui et al., 1999; Behrens et al., 2005 | درختان رده‌بندی / جنگل تصادفی / (RF) سیستم‌های خبره | متغیرهای محیطی + نقشه‌های خاک legacy | منطقه‌ای | رده‌بندی / پیش‌بینی رده‌های خاک |
| White et al., 2017 | روش‌های بیوانفورماتیک / محاسباتی | داده‌های چنداُمیکس (متاژنومیکس) | ریزوسفر (مقیاس میکرو) | تحلیل میکروبیوم ریزوسفر |
| Wu, 2025; Wadoux, 2025 | پردازش زبان طبیعی / (NLP) مدل‌های زبانی بزرگ (LLM) | داده‌های متنی غیرساختاریافته (مقالات، گزارش‌ها، اسناد سیاستی) | متنی) معنایی scalable) | سنتز دانش سلامت خاک |

## نتیجه‌گیری

این مرور تحلیلی-اجمالی نشان داد که هوش مصنوعی و یادگیری ماشین از دهه ۱۹۹۰ به بعد به‌طور چشمگیری علوم خاک را متحول کرده‌اند. در مقیاس ریزوسفر و میکرو، تحلیل داده‌های multi-omics و مدل‌سازی پیچیده زیستی امکان شناسایی الگوهای ریزساختاری خاک و میکروبیوم ریزوسفر را فراهم کرده است، در حالی که در مقیاس مزرعه و منطقه‌ای، نقشه‌برداری رقومی خاک و توابع انتقالی با استفاده از الگوریتم‌هایی مانند جنگل تصادفی، شبکه‌های عصبی و یادگیری عمیق، دقت پیش‌بینی ویژگی‌های خاک را بهبود بخشیده‌اند. فناوری‌های نوین پایش، شامل سنجش از نزدیک و سنجش از دور چندطیفی و ابرطیفی، همراه با پردازش تصویر و تحلیل داده‌های چندبعدی، امکان برداشت داده‌های گسترده و دقیق‌تر را فراهم کرده‌اند. با این حال، چالش‌هایی باقی مانده است، از جمله ماهیت جعبه‌سیاه مدل‌های یادگیری عمیق، کمبود داده‌های استاندارد، و تلفیق ناکافی دانش فیزیکی و خاک‌شناسی با مدل‌های داده‌محور. چشم‌انداز آینده شامل استفاده گسترده از مدل‌های زبانی بزرگ و هوش ترکیبی برای تجمیع و دسته‌بندی دانش، استخراج داده‌های متنی و توسعه مدل‌های فیزیکی–داده‌ای است که می‌تواند پیش‌بینی‌های دقیق‌تر و مدیریت جامع‌نگر خاک و آب را ممکن سازد.

8 - convolutional neural network

## منابع


Arrieta, A. B., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., Herrera, F. (2020). Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion*, 58, 82–115.

Arrouays, D., McBratney, A. B., Minasny, B., Hempel, J. W., Heuvelink, G. B., MacMillan, R. A., Zhang, G. (2014). GlobalSoilMap: Toward a fine-resolution global grid of soil properties. *Advances in Agronomy*, 125, 93–134.

Arrouays, D., Poggio, L., Mulder, V. L., Lacoste, M., Chen, S., Heuvelink, G. B. M., Thompson, J. A. (2020). GlobalSoilMap and the World Soil Information Service: Shareable soil data. *Geoderma*, 370, 114260.

Behrens, T., Förster, H., Scholten, T., & Steinrücken, U. (2005). Digital soil mapping using artificial neural networks. *Journal of Plant Nutrition and Soil Science*, 168(1), 21–33.

Biswas, A., et al. (2023). Semantic artificial intelligence in soil science: Opportunities and challenges. *Soil Systems*, 7(1), 12.

Bui, E. N., Henderson, B. L., & Viergever, K. (1999). Knowledge discovery from models of soil properties. *Australian Journal of Soil Research*, 37(6), 1037–1050.

Dangal, S. R. S., 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.

Ferreira, A. D., Oliveira, M. S., & Tarquis, A. M. (2022). Deep learning for 3D soil structure characterization using X-ray CT. *Geoderma*, 405, 115377.

Heuvelink, G. B. M., Brus, D. J., & de Gruijter, J. J. (2021). Digital soil mapping: Bridging the gap between research and practice. *Geoderma*, 401, 115155.

Heung, B., Ho, H. C., Zhang, J., Knudby, A., Bulmer, C. E., & Schmidt, M. G. (2016). An overview and comparison of machine-learning techniques for classification purposes in digital soil mapping. *Geoderma*, 265, 62–77.

Jansson, J. K., & Hofmockel, K. S. (2018). The soil microbiome—from metagenomics to metaphenomics. *Current Opinion in Microbiology*, 43, 162–168.

Lavrukhin, A., Munkholm, L. J., & Lamandé, M. (2021). Predicting soil structural properties using convolutional neural networks applied to X-ray CT images. *Soil & Tillage Research*, 205, 104750.

Levy-Booth, D. J., Prescott, C. E., & Grayston, S. J. (2020). Microbial functional genes in forest soils: Recent advances in metagenomic and metatranscriptomic analyses. *Soil Biology and Biochemistry*, 149, 107898.

Lobsey, C. R., & Viscarra Rossel, R. A. (2016). Towards accurate proximal sensing of soil organic C with vis–NIR spectroscopy. *Geoderma*, 265, 110–122.

Minasny, B., & McBratney, A. B. (2016). Digital soil mapping: A brief history and some lessons. *Geoderma*, 264, 301–311.

Minasny, B., & McBratney, A. B. (2025). Machine learning and soil science: Thirty years on. *Geoderma*, in press.

Mulder, V. L., de Bruin, S., Schaepman, M. E., & Mayr, T. R. (2011). The use of remote sensing in soil and terrain mapping—A review. *Geoderma*, 162(1-2), 1–19.

Ng, W., Minasny, B., Malone, B. P., & Padarian, J. (2020). Convolutional neural networks for soil property prediction using proximal and remote sensing. *Geoderma*, 366, 114227.

Padarian, J., Minasny, B., & McBratney, A. B. (2020). Machine learning and soil science: Progress and prospects. *Geoderma*, 375, 114540.

Rossel, R. A. V., & Behrens, T. (2010). Using data mining to model and interpret soil diffuse reflectance spectra. *Geoderma*, 158(1-2), 46–54.

Schaap, M. G., Leij, F. J., & van Genuchten, M. T. (1998). Neural network analysis for hierarchical prediction of soil hydraulic properties. *Soil Science Society of America Journal*, 62(4), 847–855.

Sothe, C., de Souza Filho, C. R., & Liesenberg, V. (2022). Mapping soil organic carbon with remote sensing and machine learning: A review. *Geoderma*, 427, 116100.

Taghizadeh-Mehrjardi, R., Schmidt, K., & Behrens, T. (2020). Improving digital soil mapping with remote sensing time series. *Geoderma*, 379, 114623.

Tziolas, N., Tsakiridis, N., Theocharis, J. B., & Zalidis, G. (2021). Machine learning algorithms for mapping soil properties: A systematic review. *Geoderma*, 405, 115237.

Twarakavi, N. K. C., Šimůnek, J., & Schaap, M. G. (2009). Development of pedotransfer functions for estimation of soil hydraulic parameters using support vector machines. *Soil Science Society of America Journal*, 73(5), 1443–1452.

Viscarra Rossel, R. A., Behrens, T., Ben-Dor, E., Brown, D. J., Demattê, J. A., Shepherd, K. D., ... & Ji, W. (2019). A global spectral library to characterize the world's soil. *Earth-Science Reviews*, 193, 441–458.

White, R. A., et al. (2017). Analysis of soil microbial communities with metagenomics and metatranscriptomics. *Soil Biology and Biochemistry*, 112, 91–99.

Wu, W., et al. (2025). Artificial intelligence in soil science: From machine learning to large language models. *Geoderma*, in press.

Wadoux, A. M. J. C., et al. (2025). Artificial intelligence for soil mapping and prediction: A systematic review. *Geoderma*, in press.

Yu, K., Pieterse, C. M., & Bakker, P. A. (2021). Deep learning for multi-omics data integration in soil-plant-microbe interactions. *Trends in Plant Science*, 26(11), 1113–1127.

Zhang, X., et al. (2022). Prediction of soil properties from hyperspectral data using machine learning: A review. *Remote Sensing of Environment*, 269, 112805.

**A Review on Multi-Scale Analysis of Soil Data Using Artificial Intelligence and Machine Learning**

Ehsan Ghezllbash

Assistant Professor, Department of Soil Science and Technology, Faculty of Agriculture, Tarbiat Modares University, Tehran, Iran, ghezelbash@modares.ac.ir

**Abstract**

In recent decades, data-driven approaches have become a cornerstone of soil science. Concurrent with the increase in the volume and diversity of soil data (from biological data at the rhizosphere scale to satellite imagery at the regional scale), the need for novel analysis methods has become increasingly evident. Artificial Intelligence (AI) and Machine Learning (ML) have experienced remarkable growth in this field since the early 1990s and are now considered key tools for digital soil mapping, developing pedotransfer functions, proximal and remote sensing, and even analyzing pedological scientific texts. This review aimed to systematically examine the main applications of AI/ML in soil science from 1990 to 2025. The results indicated that algorithms such as Random Forest, Support Vector Machines, and Deep Neural Networks have been most frequently applied at the field and regional scales, while modern monitoring technologies (spectroscopy, CT imaging, and multi-temporal remote sensing) have provided new opportunities for soil data analysis. Challenges such as the need for model transparency (Explainable AI), the lack of standardized data, and the necessity of integrating pedological knowledge with data-driven approaches persist. By introducing a conceptual framework linking scalability and novel technologies, this article also outlines future directions, including the utilization of Large Language Models and hybrid intelligence.

**Keywords (Times New Roman, 10pt, Bold):** Data-driven approaches, Explainable Artificial Intelligence, Large Language Models, Multi-scale