

REVIEW ARTICLE OPEN ACCESS

Russell Review

Artificial intelligence in soil science

Alexandre M. J.-C. Wadoux 

LISAH, Univ Montpellier, AgroParisTech, INRAE, IRD, L'Institut Agro, Montpellier, France

Correspondence: Alexandre M. J.-C. Wadoux (alexandre.wadoux@inrae.fr)**Received:** 10 September 2024 | **Revised:** 9 February 2025 | **Accepted:** 10 February 2025**Funding:** This work was supported by European Union's Horizon 2020 Research and Innovation Programme under the Marie Skłodowska-Curie grant agreement No 101059012.**Keywords:** decision-making | deep learning | expert systems | information processing | machine intelligence | machine learning

ABSTRACT

Few would disagree that artificial intelligence (AI) holds potential for advancing knowledge and innovation. Over the past decades, substantial research has been devoted to the development and application of AI in soil science. While most of today's AI applications in soil science are related to machine learning (ML), AI also encompasses other fields such as digital image analysis, natural language processing (NLP), expert systems, and knowledge representation. This review aims to provide a comprehensive overview of AI in soil science. A definition of AI that equates intelligence with rationality is provided, followed by a typical classification of AI into the three main domains of sensing and interacting, reasoning and decision-making, and learning and predicting. From this framework, a taxonomy of AI in soil research is derived and serves as a basis for a literature review. The major findings are as follows: i) AI in soil science is diverse, with applications in decision support systems, image classification, prediction with ML and expert systems; ii) AI in soil science is currently almost exclusively characterized by ML; iii) applications of ML are predominantly found in the field of digital soil mapping and for the development of pedotransfer functions; and iv) most AI applications are used for prediction purposes. A few notable exceptions stand apart from mainstream applications, particularly in the realms of NLP, the development of soil cognitive models, and interpretable ML. Based on these findings, I discuss attention points, such as using AI almost exclusively for prediction at the expense of explanation and the lack of integration of soil knowledge in algorithmic AI solutions. I envision that future developments could include the use of AI for text recognition of legacy soil profile data, providing a new source of soil information. Another promising line of research is the language processing of soil texts to build meta-analyses that summarize the growing body of soil science literature. These new applications could foster substantial new contributions to soil science research.

1 | Introduction

Few would disagree that there is much to be learned and gained from artificial intelligence (AI). Soil science is no exception. Over the past decades, a considerable amount of research has been devoted to the development and application of AI to soil research. Consider the following cases: a) a machine learning (ML) algorithm is fitted between measured soil data and environmental factors of soil formation for predicting soil carbon stocks in space and time (Heuvelink et al. 2021), b) a decision

support system evaluates semi-quantitatively soil functions in agricultural lands (Debeljak et al. 2019), c) text-based metadata are automatically extracted from mining soil-related scientific publications (Blanchy et al. 2023), d) digital images of soil thin sections are analyzed by an algorithm detecting mineral and non-mineral constituents (Aydemir et al. 2004). In each of these cases, a system is built or trained to achieve a certain goal with some autonomy—whether this goal be spatial prediction, soil function assessment, the creation of a meta-analysis, or the automated detection of soil features.

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2025 The Author(s). *European Journal of Soil Science* published by John Wiley & Sons Ltd on behalf of British Society of Soil Science.

Summary

- Intelligence is defined in terms of rationality.
- A taxonomy for artificial intelligence in soil science is proposed.
- Applications exist in decision support systems, image classification, prediction with machine learning and expert systems.
- Applications to soil science are currently fueled by the field of digital soil mapping and pedotransfer functions.
- Promising future research lines are discussed.

The development of AI-based technologies has skyrocketed since the 2010s across all fields of science and industry. The AI Index, published annually by the AI100 panel of experts (Nestor et al. 2024), noted a sharp increase in AI research, with ML publications alone increasing nearly seven-fold since 2015. This surge is fueled by major private investments and, consequently, a significant rise in the number of AI patents. Soil science is no exception, with several reviews (e.g. Padarian et al. 2020; Grunwald 2022; Chen et al. 2022; Minasny et al. 2024) highlighting the excitement surrounding AI applications in soil research.

For many, however, the definition of AI remains unclear. While some focus on developing AI that models human thinking, others view AI as an algorithm designed to achieve optimal results. This distinction is discussed later in this review. AI is often equated with ML, and indeed, most of today's AI applications are related to ML. But AI encompasses more than just ML. According to Nestor et al. (2024), the study of AI also includes computer vision, pattern recognition, control theory, and process management, among others. Additionally, Russell and Norvig (2016) note that AI includes disciplines such as natural language processing (NLP), robotics, knowledge representation, and reasoning. Hereafter, I adopt a definition that considers AI beyond the mere application and development of ML algorithms.

The use and widespread application of AI in soil science is not new. Back in 1979, Holt (1978) foresaw a revolution in soil science driven by the three domains of expert systems, pattern recognition and natural language interfaces. From this early enthusiasm succeeded great expectations, several applications in fuzzy logic (e.g. Burrough 1992), decision support systems (e.g. Skidmore et al. 1991; Zhang and Berndtsson 1991) but the expected leap in research did not seem to materialize. From the 2010s onwards, AI applications to soil research experienced a renewed interest fueled by developments in ML, easy-to-use software resources and, in a number of ways, by the digital convergence. Digital convergence in soil science was defined previously by Wadoux and McBratney (2021) as “the convergence of digital data, electronic databases stored in computers, the internet and new tools to handle and process large amount of electronic data”. Perusal of the literature (e.g. Canillas and Salokhe 2002; Behrens et al. 2005; Fajardo et al. 2016; Alvarez et al. 2011) reveals that AI has been applied to address numerous

questions across different fields of soil science, yet there appears to be little inter-disciplinary dialogue, with each field operating in isolation. In that context, one could content that beyond mere progress in the development of AI in soil science, there is potential for synergies among the diverse and complementary applications.

This review aims to explore the integration of AI into soil science research and project potential future developments. The objectives are to: i) characterize the use of AI within soil science, ii) document its applications across different areas of the field, and iii) identify key areas of focus, attention points and discuss opportunities for future research.

2 | A definition and taxonomy of AI in soil research

2.1 | What is artificial intelligence?

The field of AI emerged in the 1950s, and while there might be earlier works (see, for example, the seminal contribution of Turing 1950), early explorations in this domain were synchronous to a series of scholarly gatherings between 1955 and 1956 on “Machine Learning”, “Artificial Intelligence” and “Thought Processes” (Nilsson 2009) and to the proposal by McCarthy et al. (1955) to explore the capacity of machines to “use language, form abstractions and concepts, solve kinds of problems now reserved for humans, and improve themselves”. The historical trajectory of AI can be divided into two relatively distinct periods. The first period, spanning the 1980s to the 1990s, was characterized by research and applications in expert systems, computer vision, fuzzy logic and speech recognition. The second period, from the 2010s onwards, while continuing to witness developments in the forementioned areas, is mostly characterized by data analytics, ML and extensive applications of deep learning in various AI sub-disciplines (e.g. robotics and speech recognition). The differences between the two periods are attributed to a change in computing power, the availability of massive volumes of digital data, the inter-connectivity of computer systems and the widespread deployment of sensors and actuators.

The Merriam-Webster Dictionary defines AI as “the capability of computer systems or algorithms to imitate intelligent human behaviour” (Merriam-Webster dictionary 2011). This definition is widely echoed in the scientific literature, where AI systems are portrayed as intelligent entities capable of emulating the cognitive functions of the human brain. This definition implies two key points: that the domain of AI is limited to the emulation of human behaviour, and that the concept of intelligence is clearly defined and characterized. Even if one could argue in favour of either point, both remain contentious. Regarding the first point, it is observed that the majority of AI research and applications address practical problems posed by the real world, rather than focusing on the replication of biologically observable behaviours in humans. This suggests that AI has a broader scope than merely studying and replicating human behaviour.

The second point presupposes that we have a clear understanding of what intelligence is. Yet, there is no universal definition

of intelligence that could be applied to humans and machines alike. As McCarthy (2007) writes, our definition of intelligence is invariably influenced by our conception of human intelligence, making it difficult to establish a standalone criterion for intelligence in computational systems. In some instances a computer system may display properties that we equate with human intelligence, such as memory and the capacity to perform a task (Legg and Hutter 2007). However, as McShea (2013) argues, no artificial intelligence system can answer to inquiries about personal preferences or desires, such as “What would you like to do this afternoon?”, because no want, preference or care is driven by a chain of pure logic as encoded in AI systems. It appears that significant distinctions exist between human intelligence that include consciousness, behaviour, and concede to emotions, and the functional capacities of computer systems. This difference presents a challenge to any direct equivalence one could draw between the two types of intelligence.

The concept of a universal form of intelligence shared by humans and machines is vague and debated, but the situation is not as bad as it may appear. Instead of focusing on intelligence as a fidelity to human performance and behaviour, some have defined intelligence as the production of *rational* thought and actions, that is, making appropriate decisions based on the context and objectives at hand. Clearly, considering intelligence as rationality and optimal behaviour in artificial systems diverges from equating intelligence strictly with human-like performance and behaviour. This difference is made clear in the seminal book of Russell and Norvig (2016), which posits that artificially intelligent systems may either mimic human thought and behaviour or operate based on rationality. The latter definition, which I advocate for, aligns with an early conceptualization of AI as “the science of making machines do things that would require intelligence if done by men” (Minsky 1969), as well as a more contemporary definition by the European Commission expert group on AI. This recent definition characterizes AI as “systems that display intelligent behaviour by analyzing their environment and taking actions—with some degree of autonomy—to achieve specific goals” (European Commission 2019). Here too, intelligence equates with rational actions.

Having explained that the notion of intelligence of artificial systems is linked to rationality, and having provided a definition of such an AI system, points us to outline sub-disciplines and techniques that constitute AI and facilitate the construction of AI systems. These techniques can be classified into three broad categories (European Commission 2019): reasoning and decision-making, learning, and robotics, though numerous other classifications exist. Techniques that aim at reasoning and decision-making include knowledge representation, inference with symbolic rules, and optimization algorithms. Learning is achieved with the popular techniques of machine and deep learning, while robotics pertains to the interaction of an AI system with the physical world, extending its capabilities beyond mere reasoning and learning by engaging with other systems and perceiving its environment. None of the practices subsumed under these three broad categories form a coherent whole, but they do help provide a separation between the hugely diverse sets of techniques that compose AI. These categories are widely accepted in the literature (e.g. Russell and Norvig 2016; Pournader et al. 2021; Doorn 2021).

2.2 | A taxonomy of AI in soil science

Having defined AI and highlighted the possible separation into the three broad categories of reasoning and decision-making, learning, and robotics, the question that arises is how this all applies to the field of soil science. There is, to date, no classification of what constitutes AI in soil science. I therefore propose a taxonomy designed to encompass all applications and methodologies used in AI soil science research. The process began with a review of academic literature across the broad field of AI and its applications to identify existing taxonomies. The previously described three categories defined by the European Commission (European Commission 2019) served as a starting point. The three categories were then refined by expertise and a comprehensive literature search using a wide range of keywords. A soil science-specific taxonomy was formulated. Input was solicited from three experts with diverse backgrounds in AI and soil science to refine and enhance the proposed taxonomy through discussion. The resulting taxonomy presented hereafter reflects a synthesis of these discussions, though it is not without limitations and areas of overlap. It shares similarities and overlaps with other established AI taxonomies (e.g. Kotu and Deshpande 2018; European Commission 2019) and taxonomies specific to applied AI in other fields of science (e.g. optical networks (Mata et al. 2018), supply chain management (Pournader et al. 2021) and water science (Doorn 2021)).

The resulting taxonomy of AI in soil science is shown in Figure 1. It illustrates three levels: i) the core domains of AI, ii) their applications, and iii) example techniques and methods. All levels have had at least one application in soil research, but techniques often overlap significantly. Explainable AI, for example, is composed of a broad range of methods that generate explanations from fitted ML algorithms. The first domain, headed *Sensing and interacting*, encompasses NLP, image analysis, computer vision, and robotics. The second domain, *Reasoning and decision-making*, involves knowledge representation, symbolic rule inference, expert systems, and search/optimization techniques. The third domain, *Learning and predicting*, includes a variety of data-driven techniques, such as neural networks and random forest. The next subsections provide more explanation of these three domains. This taxonomy is also the basis for the literature search conducted in Section 2.3.

2.2.1 | Reasoning and decision-making

The domain of **reasoning and decision-making** relates to algorithms and programs that emulate the reasoning (i.e. formalization and structuring) of experts to solve well-defined problems. It involves the steps of knowledge representation and inference with symbolic rules and numbers, and a search through various sets of possible solutions.

Expert systems or knowledge-based systems are programs that contain organized knowledge which can be used to emulate the behaviour of experts solving a problem. Expert systems were a major area of research in AI for soil science in the 1980s. McCracken and Cate (1986) characterized expert

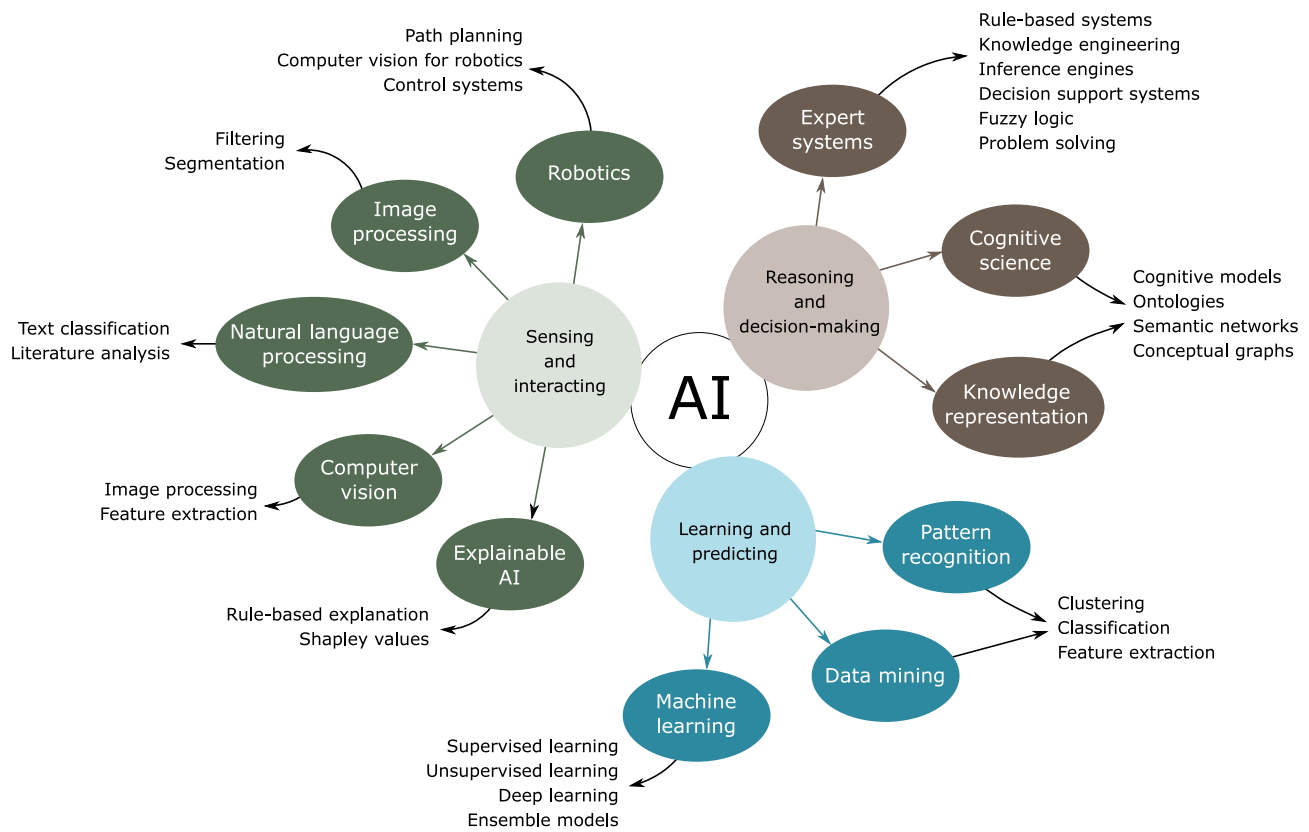


FIGURE 1 | A proposed taxonomy of AI in soil science. It has three layers composed of domains (three circles), applications (ellipses) and example techniques (outer layer). The techniques are presented for illustration purposes but usually overlap and complement each other. For example, Explainable AI typically requires prior learning, often using a machine learning algorithm.

systems as the computerized judgment and practical experience of experts. Expert systems and related techniques of knowledge engineering, inference engines, and decision support systems were used extensively for agronomic decision making (Smaling and Fresco 1993) and soil classification (Dale et al. 1989; Galbraith et al. 1998; Galbraith and Bryant 1998) in the 1980s, with recent, new applications and development for soil functions and soil quality assessment (e.g. Debeljak et al. 2019).

Cognitive science and knowledge representation are closely linked in their applications in soil science. The knowledge base for expert systems is composed of heuristic and formal knowledge. The knowledge representation is the interface between the expert's system of reasoning and the symbolic rules contained in expert systems. It makes use of an ontology and conceptual graphs. The role of cognitive science, in parallel, is to *discover and describe the mental processes by which experts encode and organize their observations* (McCracken and Cate 1986). Cognitive science and knowledge representation support the development of expert systems in cases where the formalization of knowledge is feasible.

2.2.2 | Learning and predicting

The domain of **learning and predicting** represents algorithms and groups of techniques that can solve problems that cannot be easily formalized with rules,

by learning patterns from data and solving problems with examples of input–output relationships to generalize and predict.

Machine learning is a branch of artificial intelligence that focuses on the development of algorithms and statistical models that enable computers to perform specific tasks using patterns found in training data. It is currently the most popular domain of AI in soil science (Padarian et al. 2020). ML for soil science can be separated into supervised and unsupervised algorithms. In supervised ML, the algorithms are fitted on labeled data and predict outcomes for new, unseen data based on the calibrated relationships. Common supervised algorithms are neural networks and random forest. For example, de Oliveira et al. (2021) fitted four supervised ML algorithms to predict the spatio-temporal variation of soil moisture in tropical forests in Brazil, while Minasny et al. (2021) developed a neural networks model to estimate soil water retention and hydraulic conductivity curves. Popular fields in supervised ML are deep learning and ensemble modelling. Deep learning uses neural networks with many layers to analyze various levels of abstraction in the original data, whereas ensemble modelling combines multiple ML models to improve predictions, using techniques such as boosting, bagging, model stacking, or averaging. Brungard et al. (2021), for instance, used an ensemble of random forest models fitted with various regional subsets of the data for soil depth class prediction in the upper Colorado River Basin in the Western United States. The second set of algorithms involve unsupervised ML which, conversely,

identifies patterns and structures directly from data, such as with clustering and association algorithms. An example is De Bruin and Stein (1998) where fuzzy c-means is employed for clustering soil-landscape attributes in a small drainage basin in southern Spain.

Pattern recognition and data mining have methodologies that are similar to those used in ML. Pattern recognition involves the identification and classification of patterns in the data. This field overlaps with supervised ML, where models are trained on labeled data to recognize these patterns, but pattern recognition specifically focuses on identification and the definition of new classes. For example, Saadatpour et al. (2011) used pattern recognition to classify and rank the soil erosion potential in several catchments of Ethiopia.

Data mining, conversely, is concerned with the extraction of insights from large datasets by discovering patterns and correlations, using techniques from ML (including unsupervised learning) and outside of AI with statistical and database system management. It is therefore not limited to predictive modelling and includes descriptive and exploratory analysis to understand the properties and relationships within the data. Examples in soil science are Behrens and Scholten (2006) and Bui et al. (2006).

2.2.3 | Sensing and interacting

The domain of **sensing and interacting** contains a set of applications that link AI with the physical world, including human–AI interfaces and vision/sensing based methods that enable an algorithm to process and interpret images and signals.

Robotics has advanced within the fields of agronomy and soil analyses. It performs tasks such as soil sampling (Sikora et al. 2011), monitoring and automated soil analyses (Iqbal et al. 2020; Kitić et al. 2022), dealing with the dynamics of the physical world and interacting with it using methods of path planning, computer vision and sensing, and control systems. Robots can be considered AI systems, but they interact with many disciplines other than soil science and AI (e.g. mechanical engineering, control theory, optics).

Sensing in soil science is also characterized by proximal and remote soil sensors that relate a signal to soil properties and characteristics using a model. This overlaps with the field of ML because most models are highly empirical, relying on algorithmic tools to analyze the sensor data, to identify patterns and make predictions. Example studies are Ng et al. (2020) and Baumann et al. (2022).

Computer vision and image processing are two closely related applications that make use of image filtering and enhancement, segmentation, and feature extraction. Although the difference is not always clearly made, image processing focuses on the identification and extraction of features from digital images, whereas computer vision uses image analysis to perform a task. An example of image processing is Bogrekci and Godwin (2007), in

which a framework was developed to understand the size distribution of clod and aggregate sizes from digital images. In Abeje et al. (2024), a computer vision pipeline is developed to classify soil types, with the aim of searching future targets for further image processing.

NLP aims to make computers understand, interpret, and generate human language. It involves text generation, word embedding, and text mining from a corpus of documents to answer user queries. NLP has been used to generate a language model specific to soil science with word embedding (Padarian and Fuentes 2019) and to perform a semi-automated extraction of drivers corresponding to different soil management practices or amendments from a large number of published studies (Blanchy et al. 2023).

Explainable AI are methods to understand the predictions made by complex algorithms. They are used in various applications such as to understand the important contributors to soil organic carbon (SOC) stocks (Wadoux et al. 2023) or soil fertility (Chandra et al. 2023) prediction. These methods are classified into the domain of sensing and interacting because they are a human–AI interface that reports the explanation in a way humans can understand. They are most often based in the domain of ML. For example, Beucher et al. (2022) used an explainable AI method called Shapley values to understand how a complex deep learning model performed an acid sulfate soil classification.

2.3 | Literature review and classification

Using the taxonomy illustrated in Figure 1, I applied a systematic review methodology to make an inventory of AI applications in soil science. The search was performed on the Web of Science (WoS) Core Collection database on April 28th, 2024 using the strings “soil*” or “pedolog*” in the title and with at least one of the following keywords: “Artificial intelligence”, “Expert System”, “Knowledge Representation”, “Decision support system”, “Machine learning”, “Data mining”, “Pattern recognition”, “Evolutionary computation”, “Robotics”, “Image processing”, “Machine vision”, “Computer vision”, “Human-AI interaction”, “Natural language processing”, “NLP”, “Explainable AI”. I adopted a trial and error approach in which several lists of keywords were tested. This final list of keywords was obtained to provide sufficient coverage of the field without introducing too many irrelevant articles. While I acknowledge that the inventory obtained by this list of keywords is not exhaustive, it was deemed sufficient to capture the diversity of AI research in soil science. The search yielded 8607 records which were refined for articles only, published in English and in the categories Soil Science and Environmental Sciences and “Topic Meso” for soil science. The term “Topic Meso” refers to a specific category within the WoS classification system, which organizes over 300 available categories group of papers that are related to each other through citations.

This reduced the total number to 1392 articles, which were further screened to remove studies that had no direct link to soil science. After applying all the exclusion criteria, there were 1212 remaining articles used as a basis for the abstract analysis and literature review.

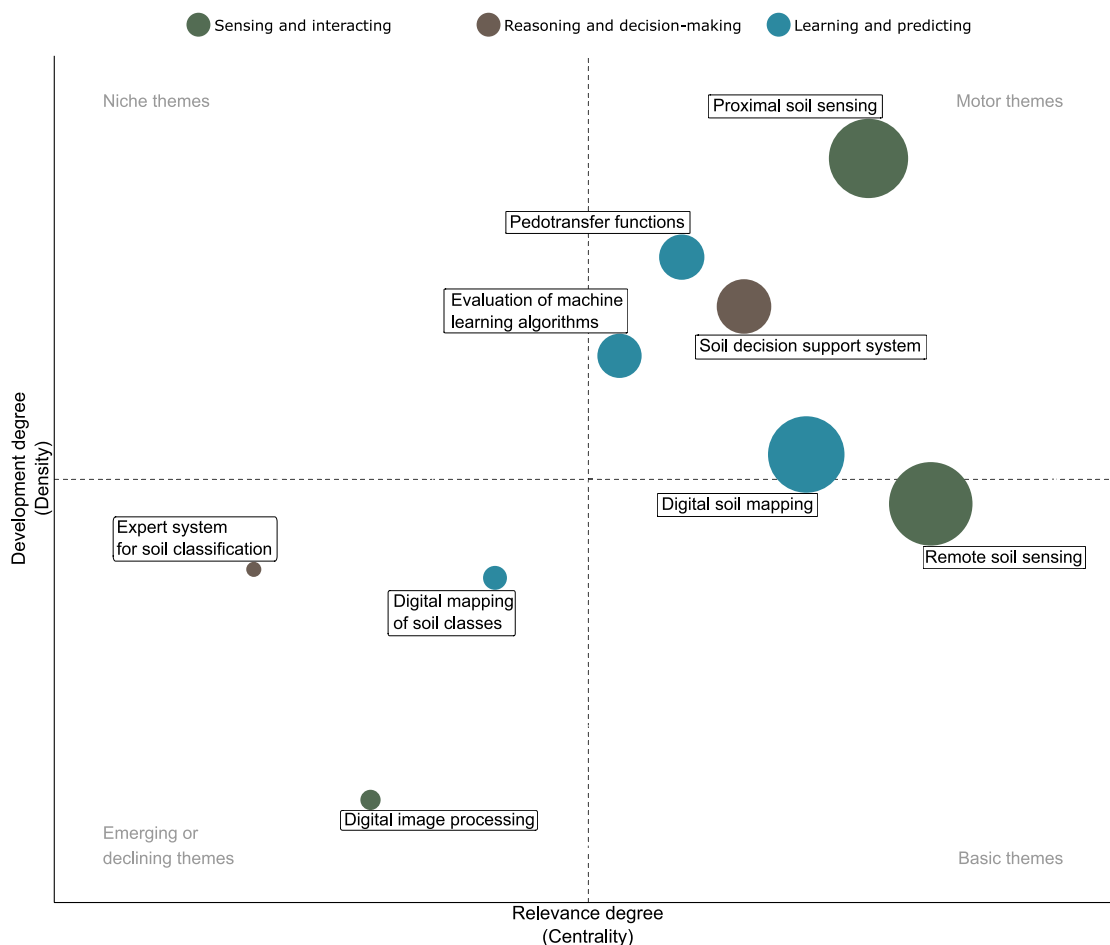


FIGURE 2 | Research themes related to AI in soil science. The circles indicate the theme while the size of the circle signifies the relative number of articles within this theme. The location of the theme relates to its type (i.e. niche, motor, emerging or declining and basic). The colour indicates the domain of the taxonomy to which they were assigned.

To classify the 1212 articles into research themes, I performed a thematic mapping using a co-occurrence keyword network and plotted them in a two-dimensional plot. The approach consisted of several steps (Cobo et al. 2011). First, the full record of 1212 articles was exported from the WoS database (i.e. title, abstract, keywords, authors, etc.). The second step consisted of extracting the co-occurrence frequency of groups of keywords. The co-occurrence was obtained by counting the number of documents in which groups of keywords appeared together. The similarity between documents was then calculated from the information extracted in the previous step. Hereafter, I used the equivalence index. It equals unity when the keywords always appear together and is zero when they are never associated. Finally, subgroups of keywords were clustered using the simple centers algorithm and referred to as themes hereafter. Two statistics, namely the Callon centrality and density (Callon et al. 1991), were calculated from the themes. The themes were plotted on a two-dimensional diagram with centrality on the x-axis and density on the y-axis. The location of a theme in the diagram informs on the type of themes, classified in the quadrant into niche themes (i.e. well developed but marginal themes, upper left), motor themes (well developed and central research fields, upper right), emerging or declining themes (i.e. not well developed and currently marginal in the field, bottom left) and basic themes that are transversal and general (i.e. themes that are important for a research field but not currently well developed,

bottom right). Further information is found in Cahlik (2000) and Cobo et al. (2011).

The resulting themes plotted in the diagram with four quadrants are shown in Figure 2. I found nine main themes across the 1212 articles. The colour represents the domain to which the theme relates most. The domain of reasoning and decision-making contains the decision support systems and expert systems. The themes digital mapping of soil classes, evaluation of ML algorithms, pedotransfer functions, and digital soil mapping are part of the domain of learning and predicting. Finally, three themes relate to sensing and interacting (i.e. digital image processing, proximal soil sensing and remote soil sensing). There is a dominance of motor themes (e.g. proximal soil sensing) and three emerging or declining themes, and only one (i.e. remote soil sensing) classified as a basic theme. No theme appeared to be a niche theme.

3 | Reasoning and decision-making

3.1 | Theme 1: Soil decision support systems

The dominating applications in this theme appear to be the development and application of decision support systems applied to the evaluation of soil management practices, erosion potential of soils and to quantify soil functions and soil

quality. A decision support system (DSS) is a computer-based information system supporting decision-making activities of unstructured and semi-structured problems through data, models or analytical tools and user-friendly interfaces (Booty et al. 2001; Power 2007).

In one early work included in this theme, Van Alphen and Stoorvogel (2000) developed two components of a soil DSS to include soil variation in precision agriculture for Dutch arable lands: the building of a farm-scale soil database and the delineation of field-scale functional units which serve as basic management units for fertilizer or irrigation recommendations. Here, the DSS is based on unit-specific mechanistic simulations. This was different in Kerebel et al. (2013) where an agro-meteorological water balance model was used in a DSS to estimate the risk of organic fertilizer pollution in relation to the soil water status. In Kaur et al. (2007), a user-friendly mechanistic field-scale DSS was proposed to suggest crop rotation and water management plans in salt-affected soils. Schwenke et al. (2019) found in a survey that simpler DSSs (e.g. spreadsheet from trial data) were preferred by farmers over complex models including process knowledge in a case study on nitrogen fertilizer recommendations by crop advisors in Australia. In this theme, studies on soil management recommendations with DSS span from 1997 to the present day.

Another active line of research in this theme is the development of DSS for soil conservation and erosion mitigation. In Montas and Madramootoo (1992), a DSS is developed for an agricultural watershed in Quebec from the combination of a geographic information system, a distributed erosion model, and an expert system. The system takes individual cells of the erosion model and returns through an interface soil conservation strategies for specific land use, soil type, and existing management practices. A wide range of similar DSSs have been developed for soil erosion mitigation with soil management, making use of geographic databases and expert systems. Most of these were developed for a specific geographic context, such as the Mediterranean region (e.g. De la Rosa et al. 2004; Anaya-Romero et al. 2015) or the semi-arid and humid regions of India (e.g. Agnihotri et al. 2021).

While there are many example applications of DSS reported in this theme, for example for soil protection (Sarangi et al. 2004) and precision agricultural management (Young et al. 2021), many studies have recently applied DSS for the evaluation of soil functions at field scale. The most common approach was described in Debeljak et al. (2019) where the DSS is composed of data, a knowledge base relying on expertise, a model, and a user interface. The core of the DSS is the model which returns a prediction given decision choices and input variables. Several studies used a model that takes the form of a hierarchical decision tree from the field of multi-criteria decision analysis. An example structure for the climate regulation function is given in Figure 3. Sandén et al. (2019), Van Leeuwen et al. (2019), Schröder et al. (2016), Trajanov et al. (2019), Van de Broek et al. (2019) and Wall et al. (2020) are relevant applications for the primary productivity, habitat for biodiversity, nutrient cycling, climate regulation and water regulation functions, respectively, and Bohanec et al. (2007) for evaluating the effect of cropping systems on soil quality.

3.2 | Theme 2: Expert system for soil classification

Expert systems were classified as a niche research theme. Expert systems were considered in the 1990s as one of the most important realizations of AI (Fisher 1989). An expert system is an inference engine coupled with a knowledge base. It provides an outcome from predetermined condition-action rules. Expert systems and the cognitive theory to build a knowledge base in soil science were extensively discussed in McCracken and Cate (1986) and Dale et al. (1989) with application in soil classification. Dale et al. (1989), for example, argued that the elements of expert systems correspond to the class definition and class assignment phases of soil classification. Most applications to soil classification in this theme relate to the assignment of an observation to a pre-defined class. Examples are MacMillan et al. (2005), in which soils are assigned to landform position based on a soil landscape database, likelihood tables and a rule-base, or Galbraith et al. (1998), where an object-oriented expert system was developed based on decision rules to identify the Histosol, Spodosol, Andisol, and Oxisol orders. Despite few developments, notably with fuzzy logic (McBratney and Odeh 1997), expert systems for soil classification have not had any major progress in the last two decades.

4 | Learning and predicting

4.1 | Theme 1: Digital mapping of soil classes

The main themes in this theme are the development and application of ML algorithms for mapping soil classes and the disaggregation of polygon-based soil type maps. Early studies in this theme are Cialella et al. (1997) and Bui et al. (1999), in which classification trees were used to predict soil drainage classes and soil map units, respectively, using the relationship between covariates and a sample taken within the class. A typical example of a study in this theme is Behrens et al. (2005), where soil units are predicted in a regional area of Western Germany using artificial neural networks and a large set of terrain, geologic and land use attributes. In recent years, the field has matured with methodological developments (e.g. probabilistic prediction or the accounting of taxonomic distance between classes, Odgers et al. 2015; Minasny and McBratney 2007) and the application of various ML algorithms to case studies (e.g. Heung et al. 2016; Brungard et al. 2015).

Another stream of research in this theme is the updating of conventional soil maps using ML and data-driven AI methods. This theme contains studies on the disaggregation of existing soil map units to increase their resolution, the updating of existing maps with new data and the harmonization of several conventional soil maps using ML methods. A popular algorithm for soil map disaggregation with decision trees was developed in Odgers et al. (2014). The algorithm works by selecting a set of sampling points randomly within soil map units, and by fitting a model between the points and a set of environmental covariates. A decision tree is used to generate realizations of the soil type from which a probability of occurrence is obtained. Chaney et al. (2016) used this method to disaggregate the soil map of the contiguous United States of America, whereas Vincent et al. (2018) combined this approach with expert-based

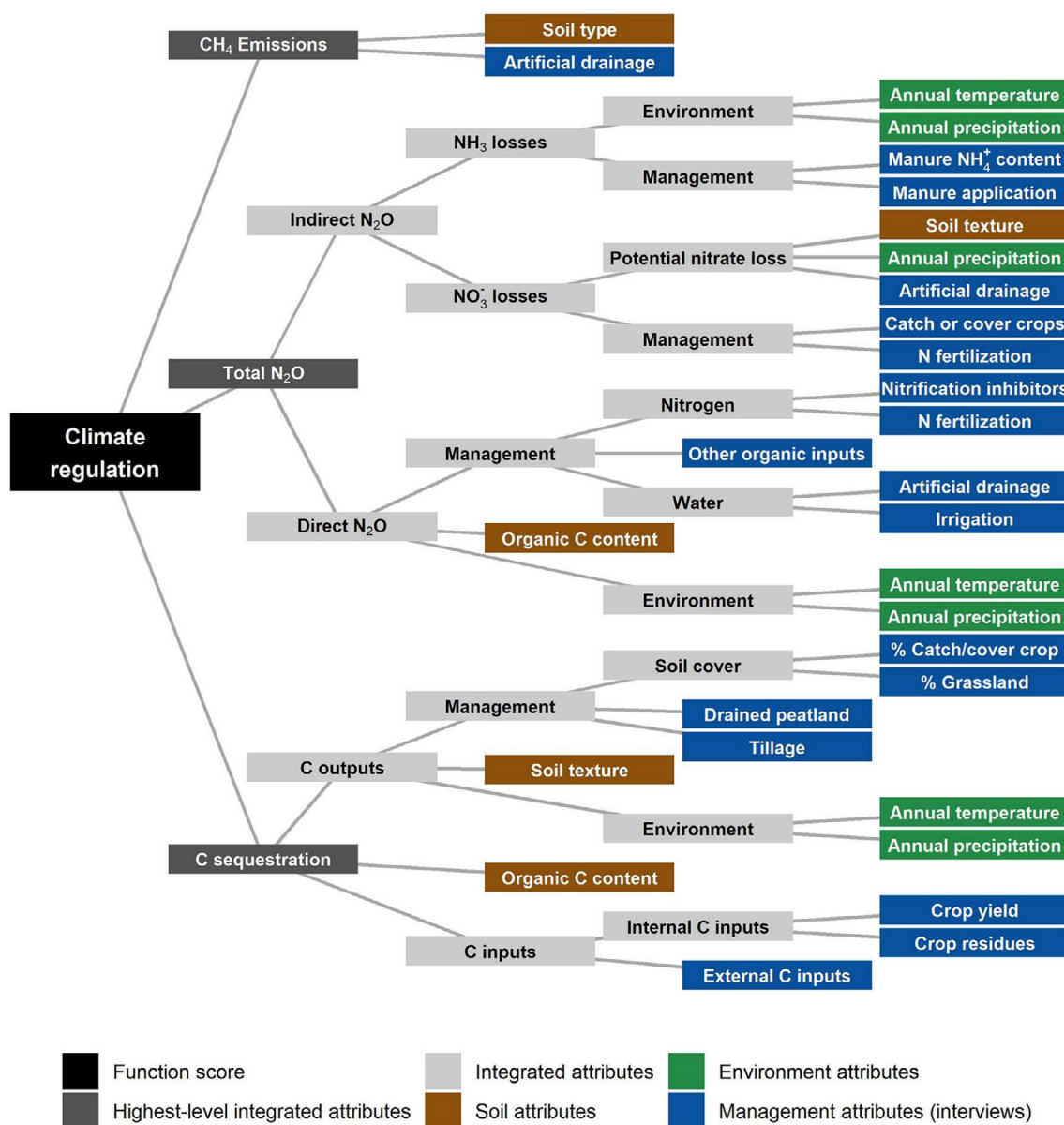


FIGURE 3 | Example structure of a hierarchical multi-criteria decision model used for soil function assessment in a decision support system. The model takes as input soil, environment and management variables and returns a qualitative evaluation of the climate regulation function through multiple levels of qualitative aggregation. With permission from Zwetsloot et al. (2021).

soil-landscape rules to disaggregate a soil map in a regional area of Brittany, France. Various methodological developments are also reported in this theme to support the application of ML algorithms to various study cases, one of which is the allocation of sampling points correcting for the unbalanced class size (e.g. Sharififar et al. 2019; Liu et al. 2020).

4.2 | Theme 2: Evaluation of machine learning algorithms

Within this theme, most research papers referred to the comparison and evaluation of ML algorithms applied to a specific purpose. The articles in this theme partly overlap with other themes but appeared as a separate theme because of their specific focus on algorithm comparison. Many articles in this theme focus on the predictive accuracy of algorithms for soil mapping

and sensing purposes. For example, Xu et al. (2022) compared four ML algorithms for soil iron content mapping on intact soil profiles with hyperspectral imaging. Comparison was made with two validation statistics, and the results were compared to that of a conventional linear model. The study concluded that a support vector machine was suitable for soil iron content prediction at profile scale, both for its high predictive accuracy and its low computational cost. Another example study is Zhang et al. (2020), which compared five ML algorithms for mapping classes or continuous values of soil texture in a regional case study in China. They concluded that the random forest algorithm was better than other algorithms for both classification and prediction purposes, but that XGboost was more efficient computationally. This theme contains many studies that performed a similar comparison between algorithms for a large range of applications, such as soil erosion susceptibility mapping (e.g. Chakraborty and Pal 2023), prediction of soil properties

from visible and near-infrared spectra (e.g. Gruszczyński and Gruszczyński 2022) or soil salinity estimation with field spectroscopy (e.g. Wang et al. 2019). The studies show that there appears to be no clear conclusion on the relative merit of the various algorithms in different situations. This conclusion was supported in Khaledian and Miller (2020), in which various ML algorithms were compared on the basis of five factors for soil mapping purposes. It was concluded that the appropriate ML algorithm depends on the nature of the data and purpose of the mapping activity.

4.3 | Theme 3: Pedotransfer functions

The main theme in this theme is the development of pedotransfer functions (PTF) with ML algorithms. In most studies, the use of AI was typically described as a means to obtain accurate predictions. Pachepsky and Rawls (2004) describe two methods for obtaining a PTF, either with linear or non-linear statistical regression or with data mining and algorithmic tools. Within the latter set of methods, artificial neural networks have been popular since the 1980s for their ability to find complex non-linear relationships between the PTF input and output (e.g. Haghverdi et al. 2012). Several studies found that neural networks performed well for prediction purposes (e.g. Minasny and McBratney 2002) and better than linear statistical regression models (e.g. Tamari et al. 1996; Schaap et al. 1998), although Pachepsky and Schaap (2004) found that ML performed similarly in terms of accuracy when comparing neural networks with regression trees. A typical study in this theme is Koekkoek and Booltink (1999) where three types of neural networks were fitted with different sets of input properties to predict soil water retention characteristics for Dutch and Scottish soil horizons. The authors obtained satisfactory results but concluded that the validation statistics were not significantly higher than those of previous studies that used linear statistical regression models. Another, more recent study, is that of Li et al. (2024) where an ensemble modelling approach composed of several ML algorithms is used to predict soil hydraulic properties. Historically, a lot has been done on developing PTFs for estimating soil hydraulic properties; see, for example, the reviews by Vereecken et al. (2010) and the comparisons amongst PTFs in Wang et al. (2022). Nemes et al. (2005) and Bagnall et al. (2022) presented PTF studies that went beyond the development of PTFs and attempted to explore relationships between selected soil inputs (e.g. organic carbon content) and the estimated soil hydraulic properties of interest. The complexity and multi-dimensionality of such relationships warrant that AI tools may further assist similar studies in the future.

The studies show that many different AI methods were used for fitting PTFs, such as nearest neighbour, support vector machines, decision trees, random forest, and boosted regression trees (Nemes et al. 2006; Jorda et al. 2015; Van Looy et al. 2017; da Silva et al. 2023). Twarakavi et al. (2009) introduced support vector machines for predicting soil hydraulic properties using texture, bulk density, and retention points, and found improvements compared to PTF developed with neural networks. A similar exercise was made in Akpa et al. (2016) with random forest, a tree-based ensemble ML algorithm, to predict bulk density and cation exchange capacity with a set of basic properties and environmental covariates at various depth intervals. Overall,

the study found that a PTF made with random forest outperformed the ones made with linear models. A review by Van Looy et al. (2017) provides a broad overview of PTFs-estimated soil properties along with useful ML-based models (e.g. support vector machines) and evaluation criteria of the PTFs.

4.4 | Theme 4: Digital soil mapping

The theme here is the application of ML to the mapping of continuous soil properties. This theme is the largest of this review and overlaps with the theme on the evaluation of ML algorithms discussed above, but the studies discussed in this section have as a main objective the creation of a digital soil map with data-driven methods, which is not the case for the other category that has as its main objective the comparison of algorithms. A typical example study of this theme is Baltensweiler et al. (2021), where three ML algorithms, geostatistics and model averaging were used to predict seven chemical and physical soil properties at various depth intervals in Swiss forested areas. The predictions were validated with usual statistics and indicated that random forest had better prediction performance than other models for most soil properties and depth intervals.

Most studies of this theme follow the digital soil mapping (DSM) conceptual framework proposed in McBratney et al. (2003), in which room for using data mining and ML algorithms in DSM is made in the empirical quantitative function that links the soil property to the factors of soil formation (for an example of DSM, see Figure 4). Other studies in this theme using this approach are Malone et al. (2020), Lacoste et al. (2014) and Poggio et al. (2021), for example. The large number of studies has logically led to several review papers documenting the application of ML in DSM of continuous properties. Minasny and McBratney (2016) reviewed the development of DSM and noted that advances in DSM were concomitant with the development of data mining and ML algorithms. A review of ML for DSM was done specifically in Wadoux et al. (2020a) whereas more specific reviews on DSM were done in Minasny et al. (2019) for peatlands, Minasny et al. (2013) for soil carbon mapping and Piikki et al. (2021) for the statistical validation of quantitative soil maps. In each of the cases, AI methods for prediction purposes were described extensively.

A review by Wadoux et al. (2020a) highlighted several challenges related to the use of ML in DSM. These challenges are about sampling designs, accounting of spatial information, multivariate mapping, quantification of uncertainty, statistical validation of models, integration of pedological knowledge and, finally, model interpretation. Many studies in recent years focus on one or several of these challenges (e.g. van der Westhuizen et al. 2023; Dasgupta et al. 2023; Fantappiè et al. 2023).

5 | Sensing and interacting

5.1 | Theme 2: Digital image processing

Within the set of articles found in the domain of sensing and interacting, a small number referred explicitly or implicitly to the use of ML in combination with imaging methods for the

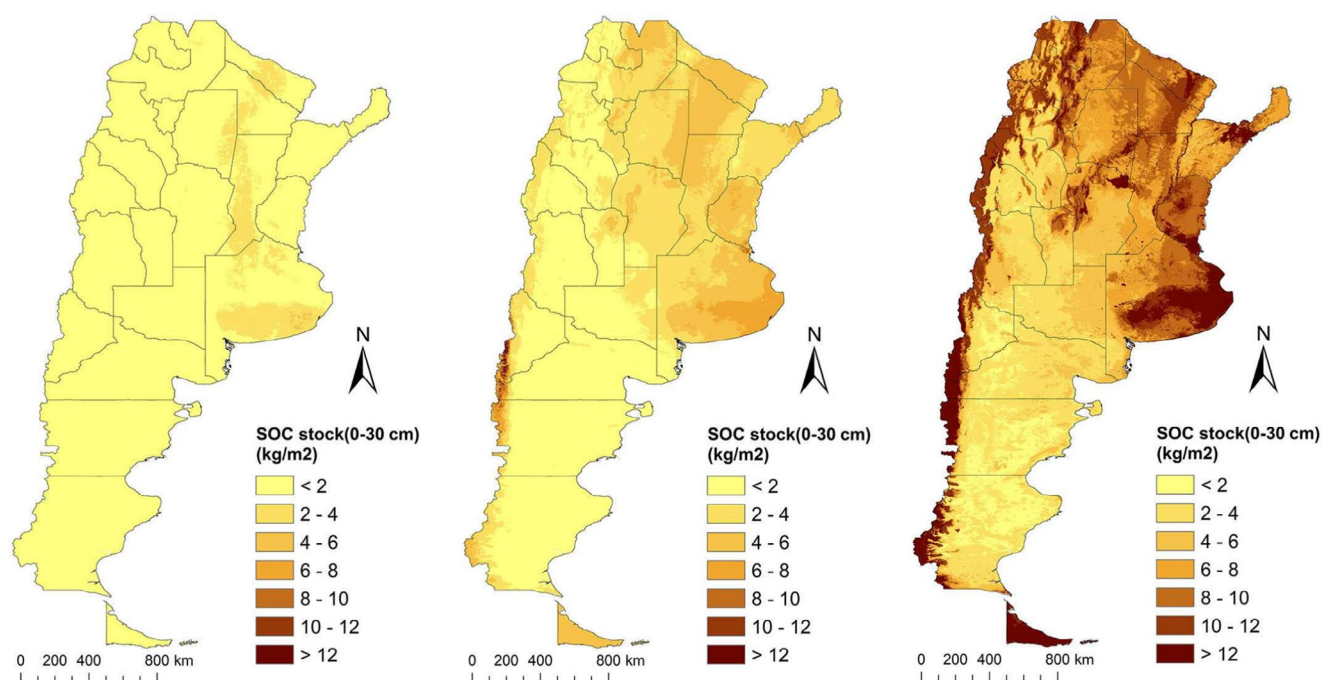


FIGURE 4 | Example maps of SOC stocks predicted with quantile regression forest, a tree-based ML algorithm. The map on the left shows the 0.05 quantiles, the map in the center shows the 0.50 quantiles and the map on the right shows the 0.95 quantiles of the predicted SOC stocks for the 0–30 cm depth interval in Argentina for the year 2017. With permission from Heuvelink et al. (2021).

non-destructive characterization of the architecture of soil samples. Most articles in this category focus on pattern analysis and classification of images obtained with imaging equipment such as nanoscale secondary ion mass spectrometry (NanoSIMS), X-ray computed tomography (CT) or nuclear magnetic resonance (NMR). A CT image of a soil sample, for example, is composed of a series of grey level values associated with the individual voxels. The grey scale is transformed into binary black and white using a threshold to segment the image into pores and solids. Techniques from ML and deep learning have been adopted for this purpose; for example, in Han et al. (2019), Lavrukhin et al. (2021), Han et al. (2021) and Ferreira et al. (2022). A representative example of such analysis is Lavrukhin et al. (2021), where 3D samples are sliced into individual images, split between calibration and validation folds, and fed into neural network models that classify each pixel/voxel into a pore phase (see Figure 5).

The use of ML and deep learning has occasionally been used in X-ray CT image processing for purposes other than soil pore space detection. Many of the articles published in this theme were reviewed in Hou et al. (2022) and Ghosh et al. (2023). Some of the soil features estimated with deep learning and X-ray CT are root distribution (Douarre et al. 2018; Smith et al. 2020). In a review, Hou et al. (2022) described deep learning algorithms as one of the four main root segmentation algorithms for soils, which could be applied to examine a range of soil features, such as soil–root interaction, root system architecture, soil micro-organism behaviour, earthworm burrows, and stones (e.g. Wieland et al. 2021). Specific techniques employed are surrogate learning (Wieland et al. 2021), convolutional deep network with transfer learning (Douarre et al. 2018) and multi-resolution encoder-decoder networks (Soltaninejad et al. 2020). These are all variants of artificial

neural networks. As an alternative, Schweizer et al. (2018) studied the microspatial arrangement of SOM coatings on soil microaggregate using an image processing procedure based on a multichannel machine-learning segmentation and nanoscale secondary ion mass spectrometry (NanoSIMS) data. Similarly, Steffens et al. (2017) used NanoSIMS to investigate secondary ion characteristics in intact soil aggregates with a supervised classification.

5.2 | Theme 2: Proximal soil sensing

The main theme here is the application of ML and deep learning to the processing of proximal soil sensing data. Proximal soil sensing is defined as the measurement of soil features with a sensor at a distance no greater than 2 m or within the soil body. Development in this area was fueled by the digital convergence (Wadoux and McBratney 2021) coupled with the availability of new sensors providing abundant and cheap data, which served as a basis for the deployment of ML and deep learning models. Viscarra Rossel et al. (2011) reviewed existing proximal soil sensors (e.g. γ -rays, X-rays, micro- and radio-waves, seismic, radar), but progress in these areas was largely driven by the application of ML to the analysis and interpretation of visible, near and mid-infrared digital soil spectra. A typical study in this theme is Viscarra Rossel and Behrens (2010), where several ML algorithms were compared for the prediction of organic carbon, clay content and pH from diffuse reflectance spectra of soils. They showed how a small number of wavelengths selected by the model related to molecular vibration of organic and mineral groups. Many similar studies used AI techniques to process infrared spectra and obtain information from the high-dimensional and correlated spectral dataset. Nawar and Mouazen (2022) used four different ML algorithms

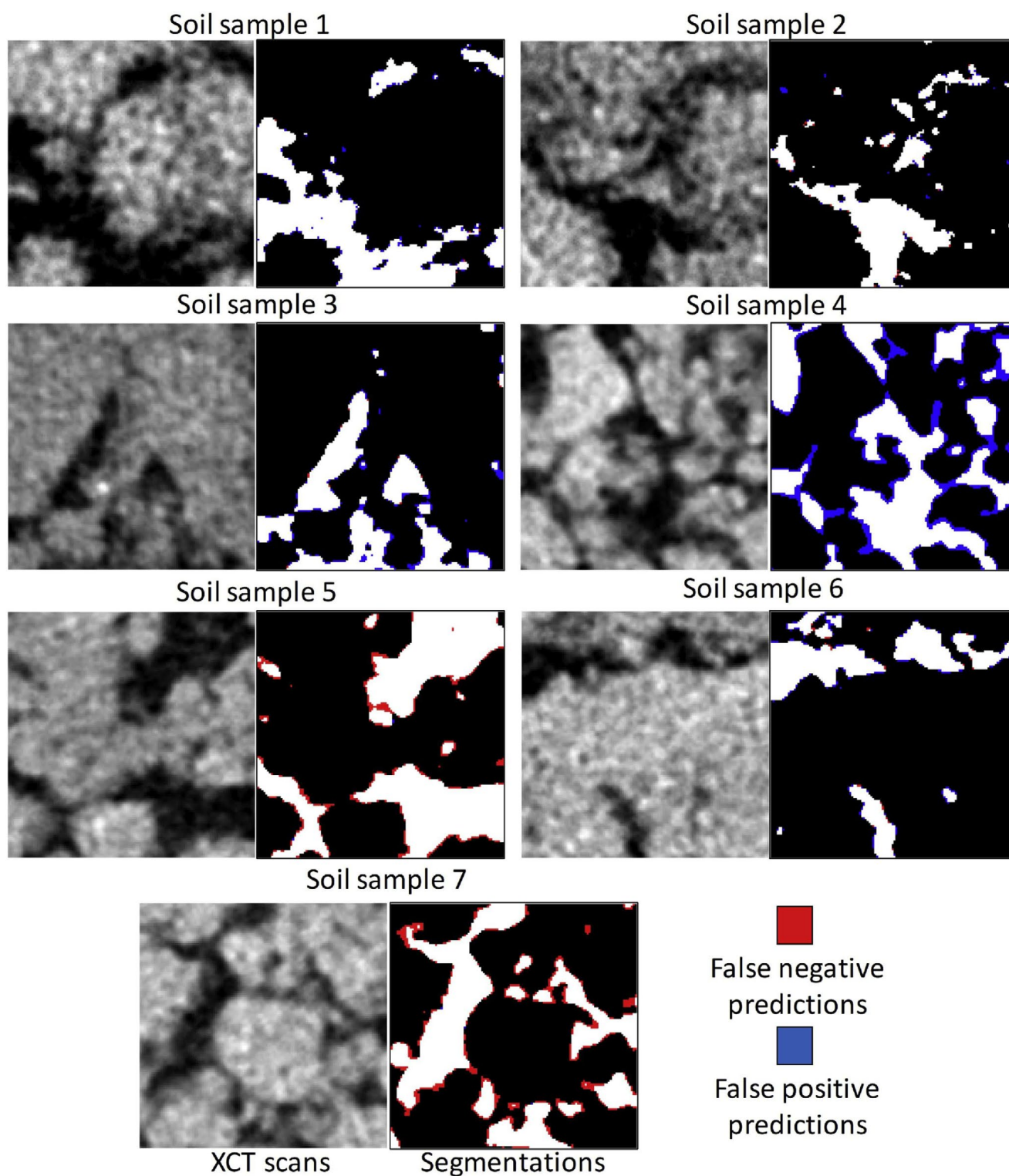


FIGURE 5 | Example greyscale images of soil samples obtained with X-ray CT (left hand-side images) and binary classification into black and white areas with convolutional neural networks (right hand-side images) along with areas of false negative and false positive predictions of each pixel into a pore phase. With permission from Lavrukhin et al. (2021).

combined into a single meta-model for predicting several soil properties. The results were compared to those obtained with an artificial neural network model. This theme of studies suggests that possibly all ML algorithms were tested for similar purposes (e.g. Ng et al. 2020; Xu et al. 2022; Gruszczyński and Gruszczyński 2022), although recent studies show an

interest in using deep learning (e.g. Padarian et al. 2019b; Shen and Viscarra Rossel 2021), and a focus on solving particular issues pertaining to the use of large soil datasets (e.g. the localization of spectroscopic models through transfer learning (Shen et al. 2022), or the interpretation of complex models (Wadoux 2023; Viscarra Rossel et al. 2022)).

In the same fashion, several studies in this theme applied ML and deep learning algorithms to the analysis of other proximal soil sensor data. Priori et al. (2014), for example, tested two ML algorithms for the prediction of soil texture with γ -radiometric data in vineyard fields in Italy. Biney et al. (2022) tested portable X-ray fluorescence for the quantification of arsenic level in a field in Czech Republic. They tested various ML algorithms and an ensemble model and found that the combination of an ensemble model, portable XRF data obtained in the lab coupled with field spectroscopy data predicted arsenic level accurately. Similar research lines are found in Ji et al. (2019) for the fusion of γ -rays, electrical conductivity, visible and near infrared spectra, whereas in Romero-Ruiz et al. (2024) electromagnetic induction data were combined with supervised clustering to delineate compaction fields.

5.3 | Theme 3: Remote soil sensing

Articles categorized in this theme typically combine remotely sensed data with ML or deep learning algorithms to assess soil properties or use remotely sensed data as a source of secondary information in DSM. With no specific focus on the use of AI, Mulder et al. (2011) reviewed these two themes and listed soil properties that can be sensed by airborne and spaceborne spectroscopy. The publications in this theme mirror the list of properties from Mulder et al. (2011) and the sensors available. A representative study of this theme is Pasolli et al. (2014), where soil moisture was obtained from synthetic aperture radar (SAR) images in a mountain area in Northern Italy. The soil moisture retrieval algorithm comprised a support vector regression model and the use of covariates (e.g. a digital elevation model) calibrated with field measurements. The fitted model was used in combination with pre-processed SAR images in an operational estimation phase for soil moisture monitoring over time. The list of publications in this theme reveals a large list of properties that were estimated with remote sensing imagery of bare soil coupled with ML, such as soil texture, CEC, and SOC (e.g. Chang and Islam 2000), salinity (e.g. Wang et al. 2020) and soil nutrients (e.g. Zhang et al. 2023). Common ML algorithms used for this purpose are neural networks (e.g. Chang and Islam 2000), random forest (e.g. Rezaei et al. 2023) and deep learning (e.g. Odebiri et al. 2021). A review of the use of ML to estimate soil properties with remote sensing is provided in Diaz-Gonzalez et al. (2022).

The review showed a large proportion of studies in which remote sensing is a source of information for DSM. This category partly overlaps with Theme 4 of learning and predicting, but because of its specific focus on remote sensing, it serves a different purpose. The use of remote sensing for direct estimation of soil properties is usually hampered by vegetation cover. This renders an incomplete coverage of the property over an area of interest, and only for the surface soil layer. To solve this, remote sensing is combined with soil sampled data and environmental covariates through a soil inference system or transfer function. AI is used here to build such a system or function. Examples are Lacoste et al. (2011), de Sousa Mendes et al. (2022) and Bao et al. (2024). In these examples, remote sensing is used directly as a covariate or indirectly through the calculation of indices (e.g. the normalized difference vegetation index).

6 | Discussion

6.1 | The use of AI in soil science

The literature review showed that AI found its early application in soil science through expert systems and decision support systems (e.g. McCracken and Cate 1986; Van Alphen and Stoorvogel 2000), which involved the programming and emulation of expert reasoning. Yet, the field appeared to stagnate post-1990s, a period that coincided with the so-called “AI winter”: a period of reduced funding and interest in AI research, typically due to unmet expectations and the perceived limitations of AI technologies. The literature review also showed an acceleration in the number of AI applications in soil science since the 2000s with a marked acceleration in the 2010s. This resurgence is largely attributed to the advancements and widespread adoption of ML and deep learning techniques. There are numerous instances where these algorithms have been employed for predictive purposes, such as the spatial prediction of soil classes (Heung et al. 2016), analyzing the architecture of soil samples (e.g. Lavrukhin et al. 2021), making accurate pedotransfer functions (e.g. Li et al. 2024), and determining soil properties from both proximal and remote sensing data (e.g. Ng et al. 2020). While differing in scope, they share the common objective to use algorithms for prediction purposes. A notable distinction between contemporary applications and those of the 1980s is the sheer volume of digital data now available. This abundance, driven by the digital convergence, offers a vast and diverse dataset that is ideally suited for training ML and deep learning models (Wadoux and McBratney 2021). A few notable exceptions stand apart from the mainstream applications, particularly in the realms of NLP, the development of soil cognitive models, and the field of interpretable ML. While these areas have seen some scholarly attention, the volume of research has not been sufficient to form a theme within the broader body of studies previously discussed. These themes present promising avenues for future soil research and will be discussed hereafter.

6.2 | Attention points

While the literature review showed that AI has been applied to a wide range of study cases, I identified five main risks, limitations, or attention points that pertain to the current use of AI in soil research.

- *Predicting and explaining.* Recent studies focus on AI almost exclusively for prediction purposes, such as for generating accurate digital soil maps, pedotransfer functions, or spectroscopic model predictions. This is very different from the “first wave” of AI research in soil science, where prediction accuracy was not the primary objective. The main goals in data analysis are usually twofold: to make predictions and to extract information (Breiman 2001). The current trend in AI applications within soil science, however, indicates a rise in modelling complexity, which pairs with the increasing size, abundance, and complexity of soil datasets. While highly sophisticated AI algorithms are advantageous for making predictions, the increased complexity in modelling poses challenges for soil scientists because it becomes increasingly difficult to explain the

rationale behind the predictions and to formulate an explanation from the link between predictor and response soil variables (Wadoux et al. 2021b, Section 5). Several studies (e.g. Wadoux et al. 2020b; McBride 2022) have stressed that with complex ML models it is possible to obtain accurate predictions, with the possibility that the correlations found amongst the data do not link to any causal relationship.

- *Accounting for soil knowledge.* There is a need to better integrate existing soil knowledge into AI soil models. Current AI models, driven by ML and deep learning algorithms, aim to establish a function that links inputs to outputs. The form of this function, however, is not constrained by any physical principle or existing process knowledge. It is instead purely driven by criteria that assess the quality of the fit. Questions can be raised on the validity of such output. For instance, is the AI model accurately predicting carbon change over time for the right reasons, that is, because certain dynamic covariates can inform on carbon addition over time. This is likely due to the fact that we rarely have enough soil samples to fully capture the high variability of the soil object being studied. In other words, truly large datasets are missing, preventing AI from accounting for this variability. AI models often lack a foundation on existing knowledge, but incorporating constraints and soil knowledge into these models is a challenging task. This issue has been highlighted in numerous studies (e.g. Angelini et al. 2016; Nussbaum et al. 2018) and identified as a significant challenge in pedometrics research (Wadoux et al. 2021a, Challenge 6). As the authors note, there is currently limited research on integrating various types of soil knowledge (e.g. tacit, conceptual) into the highly empirical ML models for soil. Although some work is being done in this direction (e.g. Koop et al. 2023), this review underscores the necessity of guiding ML models towards directions where their internal functions are mechanistically sound. Without this, the validity of conclusions drawn by AI models in soil science may be called into question.
- *Proprietary-based AI software.* The advantages of AI are the scalability, repeatability (i.e. reproducibility of outcomes and bit-reproducibility), and the ability to handle large amounts of data. Models can also be algorithmically transparent, provided that certain caveats, such as the need for interpretability of ML algorithms, are addressed. However, these benefits are at risk when using proprietary, closed-source AI software. For instance, in research applications, cartographic software now includes deep learning tools that are readily applicable for the spatial prediction of soil properties. Similarly, black-box image analysis software is routinely used for the supervised classification of images obtained by X-ray CT. The use of AI is widespread in commercial soil applications, with on-the-spot soil testing instruments capable of predicting a range of soil properties with a single scan. In these examples, the proprietary nature of some software limits the ability to fully understand and verify these predictions. Another problem with proprietary software is the strong dependency on the software producer. To ensure the reliability of AI models, it is crucial to incorporate uncertainty measures and develop standards for assessing the quality of AI models, which is in line with the recent proposal from the field of PTFs (see Weber et al. 2024). Additionally, there should be a requirement for software back-up, including details of internal functionalities, and the ability to keep a version control of the software to evaluate and modify the AI models in the long term. Finally, explicit information on the reproducibility and replicability of outcomes would ensure that the results and conclusions of a study can be duplicated in the long term (Konkol et al. 2019; Kedron et al. 2021).
- *Ethics on data sharing and harvesting.* The literature review suggests emerging ethical considerations regarding data sharing and harvesting. AI models typically require large amounts of data to function effectively, which raises issues concerning the ethical use and sharing of data. One of the primary concerns is the harmonization of data, ensuring that data from various sources can be integrated and used cohesively while maintaining data provenance (Todd-Brown et al. 2022). There is a risk of building models on data that are not sufficiently curated or fit for the intended purpose. Another risk involves developing models without acknowledging the data providers, especially when the data are not publicly available and agreements are not in place. Finally, there is a risk that the metadata may not be adequately propagated to all downstream uses of soil data, which could result in the loss of substantial information and a loss of meaning. Recent developments in soil science have called for a decentralized, data-preserving policy for AI. The blockchain can provide a secure and transparent way to store and share data used by AI systems (for an example, see Padarian and McBratney 2020).
- *Running out of quality data.* AI models are typically trained on extensive datasets, with advancements in AI modelling often achieved by increasing the dataset size. For instance, Hengl et al. (2021) observed significant improvements in mapping African soil nutrients by expanding the training dataset from 30,000 to 150,000 soil samples, depending on the target property. This dataset size is substantially larger than previous mapping efforts for the same region. The reliance of AI models on large datasets raises the question of whether we will eventually run out of quality data to further enhance these models. There are valid concerns that the existing stock of high-quality data is nearly depleted. Although new data sources may emerge (see Section 6.3), the rate of new basic soil data collection does not appear to be increasing in most countries worldwide. This challenge of limited new data in the coming years could be addressed through new data collection efforts, thorough standardization of new measurements and harmonization of existing data. Another solution is to rely more on synthetic soil data. Examples of this approach already exist in soil science literature. For instance, pseudo-points (e.g. Koch et al. 2019) or pseudo-covariates (e.g. Behrens et al. 2018) are commonly generated to inform about the value of a soil property or the spatial context, or data augmentation techniques can be employed. An example of data augmentation in soil science is provided by Padarian et al. (2019a), where the stack of images used to train convolutional neural networks for soil mapping was rotated to increase the size of the training dataset.

6.3 | Opportunities and promising future lines of research

Innovative advances in the field of AI offer new perspectives and lines of research that could advance soil science. I list five of them in the following paragraphs with examples of preliminary studies.

- *Open up data sources.* Soil science possesses a rich repository of historical qualitative descriptive data that is currently being converted into digital text format. Although this type of information is often overlooked in contemporary soil research, AI methods present a promising approach for harvesting qualitative data, and open up data sources that were until now hardly accessible, as demonstrated in other fields (e.g. Colavizza et al. 2021). For example, descriptive soil data can be used to complement numerical data, as illustrated by Furey et al. (2019). Additionally, legacy soil profile information can be digitized using automated text recognition techniques (e.g. Chemchem et al. 2021). Digitizing archived soil profile descriptions can help counteract the global decline in soil profile descriptions by introducing a new set of historical data currently stored in archives. This process can be facilitated through computer vision and text mining techniques. Recent studies have highlighted the added value of harvesting legacy data, despite its higher uncertainty (Lagacherie et al. 2024).
- *Interpretable AI.* To move beyond mere prediction with AI and explain how predictions are made, the field of explainable AI is gaining traction, showing promising results in soil research. In the taxonomy of AI described in Section 2.2, explainable AI falls into the category of sensing and interacting, as it aims to provide insights into model functioning in a user-friendly manner. This line of research is expanding within soil science, with notable examples including Beucher et al. (2022) and Albinet et al. (2022). A representative study by Wadoux et al. (2023) demonstrates the application of Shapley values, a method from game theory, to interpret a ML algorithm predicting SOC stocks in France, thereby elucidating how the model adjusts predictions locally. Beyond interpretable AI, the emerging field of “self-explaining” AI (Elton 2020) will certainly achieve progress for soil research. Self-explaining AI techniques provide human-understandable explanations of model predictions, along with confidence levels for both the decisions and explanations. This capability is particularly relevant in soil research, where understanding the extrapolation and reliability of predictions is crucial.
- *Conscious AI.* While some research has been done to extract soil information from the fitted ML and deep learning algorithms, the applications have not kept pace with techniques that include domain knowledge in AI algorithms. Hybrid physical/dynamic modelling at the interface between data-driven AI and physical modelling is one avenue of future research. Some early examples are Ghorbani et al. (2021), which derived soil water flow from soil moisture observations with ML constrained on physically consistent solutions from the Richardson–Richards equation. In Zhang et al. (2024), the ML model used prediction of process-model

RothC as an additional source of data for space–time soil carbon modelling (Figure 6). There are also recent developments in the statistical literature that have not made it to soil research, for example, causal discovery and inference (Reichstein et al. 2019). Despite some successful implementations in earth system sciences (e.g. Tesch et al. 2023), it is not clear whether the stringent assumptions made during modelling would hold for soil research with observational data.

- *Language processing of soil texts.* The field of NLP in AI offers an opportunity for summarizing the information of a large body of scientific literature. The number of publications in soil science has been steadily increasing (Hartemink and McBratney 2008), and these publications are now readily available in digital format. As a result, meta-analyses and review articles have attempted to summarize this growing body of literature. Large Language Models (LLMs), such as ChatGPT, can be employed for this purpose. LLMs are ML models that can understand, generate, and manipulate human language by leveraging vast amounts of text data. Despite some attempts in the broader scientific literature (e.g. Gougherty and Clipp 2024), the field still needs to advance to enable LLMs to extract useful and accurate qualitative and quantitative information from scientific literature. An entry point in soil science is provided by Blanchy et al. (2023), Figure 7, where a set of 36 meta-analyses were used to extract relationships between soil management and target soil properties using NLP techniques. The authors further extracted the type of relationship, whether positive, negative, or absent. The results support the use of NLP with human supervision to synthesize scientific literature. Several systems not specific to soil science exist (e.g. ChatGPT, Google Gemini, Claude), and while they allow long texts as input, they are not sufficiently trained on soil science corpora. A future line of research will involve building a LLM specifically designed to synthesize soil science literature, which

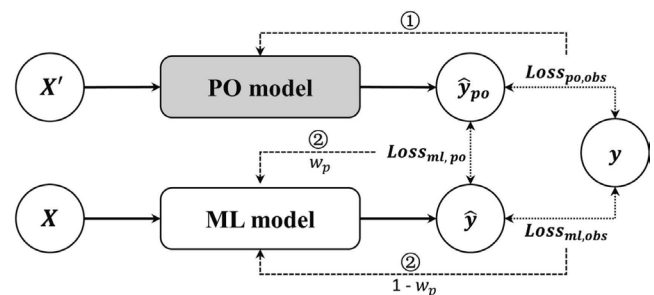


FIGURE 6 | Framework of a hybrid model that incorporates the output of the process-oriented (PO) model RothC into an ML algorithm. The numbers 1 and 2 above the arrows indicate the order of model calibration. The process-oriented model is first fitted with environmental covariates (i.e. X) and observed SOC data (i.e. y). The simulation of SOC stocks for all sampling locations (i.e. $\hat{y}_{po}$) and all years are then used as additional source of calibration data to feed the ML algorithm in addition to the measured SOC stocks. The final prediction is obtained from a weighted (w) function of the loss functions (**Loss**) of the two models. With permission from Zhang et al. (2024).

In the short term, tillage operations significantly increased K [...].

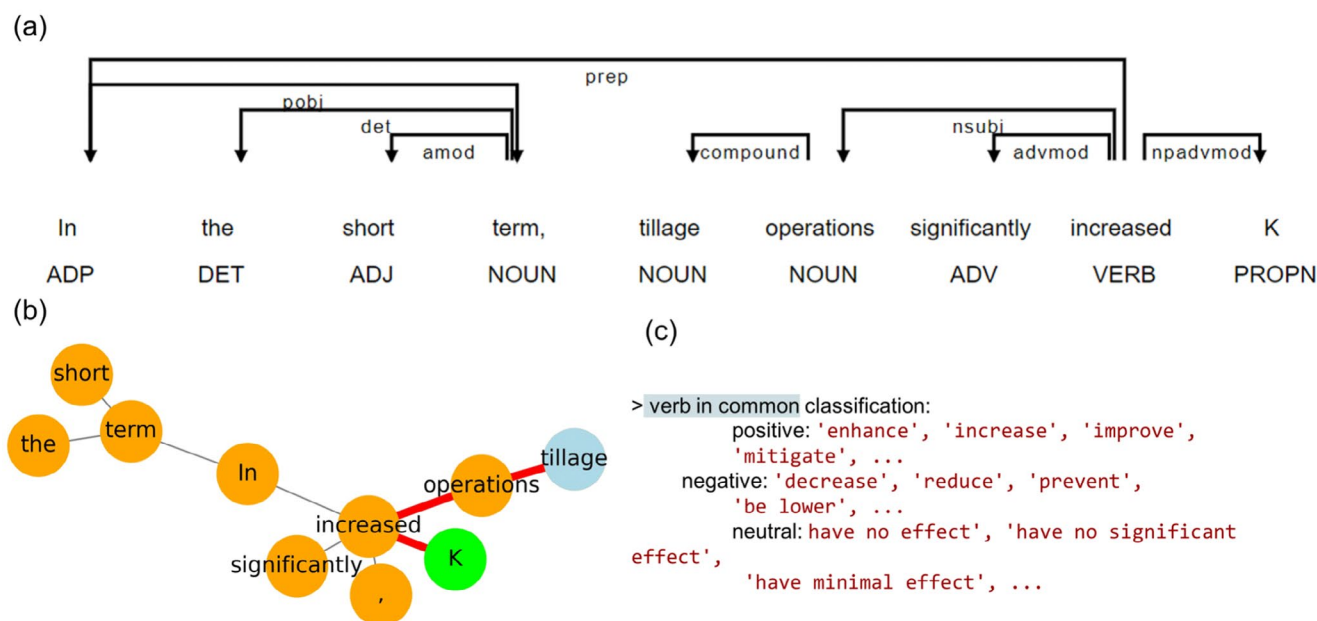


FIGURE 7 | Example of a extraction from a sentence of a soil text. The panel (a) shows the split of the sentence into words and assigned to a category (e.g. verb, noun, adjective) in dependence (i.e. with the arrows) to the other words. Panel (b) shows a graph with one link per word, differentiating the driver (blue), the variable (green) whereas the shortest path is shown with red links. In the panel (c), the verb is classified into a positive, neutral or negative according to a pre-defined list. With permission from Blanchy et al. (2023).

should be sufficiently large and adaptable to achieve the cognitive capacity to answer user queries effectively. Such LLM would certainly remove the language barrier and enable users to reach a large body of literature currently not available in the English language.

- *AI beyond ML.* The literature review has highlighted that our current use of AI in soil science is focused almost exclusively on ML for prediction purposes. However, there are other aspects of AI beyond ML that warrant investigation. One notable example is generative AI (Goodfellow et al. 2020), which uses generative models (e.g. Generative Adversarial Networks) to produce data (e.g. images, text or videos). Generative AI likely holds future applications in soil science, to create a soil digital twin (e.g. generating fully consistent pedons given certain boundary conditions). Another example is genetic algorithms and evolutionary computation, inspired by the principles of natural selection and genetics. Although these methods have been used in soil science, their application for prediction has been limited (e.g. Tsakiridis et al. 2019). Genetic algorithms present an intriguing research avenue to reconcile prediction with the understanding of soil dynamics and evolution. Beyond ML, there is also an opportunity to rediscover other aspects of AI that are currently overlooked in soil science, such as expert systems and decision support systems, which have seen little progress in the past two decades. Despite some recent studies (e.g. Debeljak et al. 2019), there is significant potential for new developments driven by the availability of computational resources and spatial data. Advancements in this area would certainly make a valuable contribution to the

soil science literature at the interface of domain knowledge and data science.

6.4 | Towards AI-driven soil scientific discovery?

The review discussed several applications of AI in soil research. It explained how these applications advanced the field by automating or accelerating parts of the scientific process that were previously handled by humans, including summarizing the literature, reading text, and making predictions. AI analyzes patterns in data to make predictions, while humans rely on experience, knowledge, and cognitive processes to form educated guesses about future events. Similarly, AI performs automated text recognition through algorithms and data, just as humans use vision and cognition to read and interpret written language. In the same way, critical thinking, knowledge synthesis, and a systematic approach to reading scientific papers are the human equivalent of how AI-NLP processes and synthesizes scientific literature in a meta-analysis (Hastings 2023). While AI aids part of the scientific process in soil research, it yet appears that it performs only minor parts of it and mainly under human supervision and refinement.

Thinking about the process of performing research in soil science can focus attention on all steps of the scientific process and the means to perform them. In soil science, a human researcher usually gathers relevant background information, formulates a series of testable hypotheses and designs a method for evaluation. Both fieldwork and laboratory experiments can be carried out to collect data supporting or refuting the hypotheses.

Finally, the researcher analyses the evidence and can make statistical analyses to draw conclusions which can, eventually, be communicated to the scientific community and end users. The near future of soil science will undoubtedly see more of the scientific process become automated and managed by AI. The question that arises is whether in the more distant future parts of soil science research could become automated?

The scientific literature reports on numerous efforts to automate the entire scientific process. For instance, the Automatic Statistician (Steinruecken et al. 2019) was designed to generate automated statistical analyses, complete with validation and reporting. More recently, the first attempt at creating an “AI scientist” has been introduced, aiming to automate the entire research pipeline—generating ideas, conducting literature reviews, planning and refining experiments, performing experimental iterations, drafting manuscripts, and even undergoing the peer review process (Lu et al. 2024). While one can rightly argue that scientific inquiry extends beyond merely producing papers, this end-to-end system offers a preview of the potential for a fully automated scientific process in the future (Castelvecchi 2024). This vision could also include traditionally human-led tasks, such as laboratory analysis and fieldwork, where AI-assisted robots might play a significant role thanks to recent developments (Gibney 2024).

While it may not be appealing to envision a future where soil science is fully automated, no AI to date—particularly in soil science—has demonstrated the ability to formulate truly creative or novel scientific directions that extend beyond the existing body of knowledge (Krenn et al. 2022). In discussing the future of soil science, Hartemink et al. (2006) and McBratney (2024) highlight that breakthroughs often come from exploring *unknown unknowns*, requiring speculative ideas that push the boundaries of creativity. Here, imagination plays a crucial role, something that AI is not yet equipped to replicate. But for the rest? Surely AI can help for normal research: much of our work involves testing hypotheses or addressing *known unknowns*, as McBratney (2024) describes—areas where we spend most of our time and where AI can offer substantial assistance. In the future, it is likely that soil science will be driven by AI-augmented researchers, where AI assist and accelerate nearly all the tasks involved in the scientific process. Ultimately, the vision remains with the human, while the execution increasingly falls to AI.

7 | Conclusion

AI initially found its application in soil science through expert systems and decision support systems that emulated expert reasoning. However, the field experienced stagnation post-1990s during the “AI winter”, a period marked by reduced funding and interest in AI research in all fields of science. Since the 2000s, there has been a notable resurgence in AI applications, particularly in the 2010s, driven by advancements and the widespread adoption of machine learning and deep learning techniques. Based on my analysis and a review of the literature, I draw the following conclusions:

- The field of AI can be defined in various ways. A practical definition equates intelligence with rationality, meaning

that artificial intelligent systems take actions with some degree of autonomy to achieve specific goals.

- AI encompasses more than just ML. In the broader literature, AI techniques are classified into three broad categories: reasoning and decision-making, learning, and robotics. For soil research, these categories can be adapted into three domains: reasoning and decision-making, learning and predicting, and sensing and interacting. Each domain has seen a significant number and diversity of applications in soil science, with the domain of learning and predicting currently experiencing the most research activity due to advancements and widespread adoption of ML and deep learning techniques.
- The literature review indicates that AI applications in soil science are diverse, including examples in image processing, decision support systems, expert systems, pedotransfer functions, and a wide range of sensing technologies coupled with ML algorithms.
- While AI applications in soil science are diverse, the by far dominant application is the use of ML for prediction.
- Grouping of research items reveals that the field is currently mainly driven by developments in digital soil mapping and for the development of pedotransfer functions, with numerous applications in soil remote and proximal sensing, situated at the interface between sensing technologies and ML algorithm development.
- Five risks and points of attention regarding the current use of AI in soil science are discussed: the distinction between prediction and explanation, the difficulty of incorporating soil knowledge into AI algorithms, the risks associated with using proprietary AI software, ethical considerations in data sharing and harvesting, and the potential risk of running out of quality data to train AI models in the coming years.
- Opportunities for future research include enabling more observations to train AI models, developing interpretable and conscious AI algorithms that better extract or incorporate soil knowledge into modelling, utilizing language processing of soil texts to summarize existing literature, exploring new techniques such as evolutionary computation, and rediscovering expert systems and decision support systems to create interfaces between pedological knowledge and data science.

Based on these findings, it is clear that AI has a significant role to play in soil science, extending beyond the mere application of ML for prediction purposes. Substantial advancements in soil science can be achieved with AI if more exploration is directed towards mining existing legacy data in text format, both to open up data sources that were until now hardly accessible and to summarize information. With the digital convergence, more information is becoming available; however, this information has not yet been fully integrated into AI-driven developments in soil science, at least not at the same pace as is currently being done in other scientific fields.

Author Contributions

Alexandre M. J.-C. Wadoux: conceptualization, writing – original draft, writing – review and editing, formal analysis.

Acknowledgements

I am grateful to Gerard Heuvelink and Jenni Dungait for inviting me to write a Russell Review for the European Journal of Soil Science 75th Anniversary year. I thank Gerard Heuvelink and the three anonymous referees who provided constructive criticisms and suggestions to my first draft. This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 101059012. For the purpose of Open Access, a CC-BY public copyright licence has been applied by the authors to the present document and will be applied to all subsequent versions up to the Author Accepted Manuscript arising from this submission.

Data Availability Statement

Data sharing is not applicable to this article as no new data were created or analysed in this study.

References

- Abeje, B. T., A. O. Salau, B. M. Gela, and A. D. Mengistu. 2024. "Soil Type Identification Model Using a Hybrid Computer Vision and Machine Learning Approach". *Multimedia Tools and Applications* 83: 575–589.
- Agnihotri, D., T. Kumar, and D. Jhariya. 2021. "Intelligent Vulnerability Prediction of Soil Erosion Hazard in Semi-Arid and Humid Region". *Environment, Development and Sustainability* 23: 2524–2551.
- Akpa, S. I. C., S. U. Ugbaje, T. F. A. Bishop, and I. O. A. Odeh. 2016. "Enhancing Pedotransfer Functions With Environmental Data for Estimating Bulk Density and Effective Cation Exchange Capacity in a Data-Sparse Situation". *Soil Use and Management* 32: 644–658.
- Albinet, F., Y. Peng, T. Eguchi, E. Smolders, and G. Dercon. 2022. "Prediction of Exchangeable Potassium in Soil Through Mid-Infrared Spectroscopy and Deep Learning: From Prediction to Explainability". *Artificial Intelligence in Agriculture* 6: 230–241.
- Alvarez, R., H. S. Steinbach, and A. Bono. 2011. "An Artificial Neural Network Approach for Predicting Soil Carbon Budget in Agroecosystems". *Soil Science Society of America Journal* 75: 965–975.
- Anaya-Romero, M., S. K. Abd-Elmabod, M. Muñoz-Rojas, et al. 2015. "Evaluating Soil Threats Under Climate Change Scenarios in the Andalusia Region, Southern Spain". *Land Degradation & Development* 26: 441–449.
- Angelini, M. E., G. B. M. Heuvelink, B. Kempen, and H. J. M. Morrás. 2016. "Mapping the Soils of an Argentine Pampas Region Using Structural Equation Modelling". *Geoderma* 281: 102–118.
- Aydemir, S., S. Keskin, and L. R. Drees. 2004. "Quantification of Soil Features Using Digital Image Processing (DIP) Techniques". *Geoderma* 119: 1–8.
- Bagnall, D. K., C. L. S. Morgan, M. Cope, et al. 2022. "Carbon-Sensitive Pedotransfer Functions for Plant Available Water". *Soil Science Society of America Journal* 86: 612–629.
- Baltensweiler, A., L. Walthert, M. Hanewinkel, S. Zimmermann, and M. Nussbaum. 2021. "Machine Learning Based Soil Maps for a Wide Range of Soil Properties for the Forested Area of Switzerland". *Geoderma Regional* 27: e00437.
- Bao, Y., F. Yao, X. Meng, et al. 2024. "A Fine Digital Soil Mapping by Integrating Remote Sensing-Based Process Model and Deep Learning Method in Northeast China". *Soil and Tillage Research* 238: 106010.
- Baumann, P., J. Lee, T. Behrens, et al. 2022. "Modelling Soil Water Retention and Water-Holding Capacity With Visible-Near-Infrared Spectra and Machine Learning". *European Journal of Soil Science* 73: e13220.
- Behrens, T., H. Förster, T. Scholten, U. Steinrücken, E.-D. Spies, and M. Goldschmitt. 2005. "Digital Soil Mapping Using Artificial Neural Networks". *Journal of Plant Nutrition and Soil Science* 168: 21–33.
- Behrens, T., K. Schmidt, R. A. Viscarra Rossel, P. Gries, T. Scholten, and R. A. MacMillan. 2018. "Spatial Modelling With Euclidean Distance Fields and Machine Learning". *European Journal of Soil Science* 69: 757–770.
- Behrens, T., and T. Scholten. 2006. "A Comparison of Data-Mining Techniques in Predictive Soil Mapping". *Developments in Soil Science* 31: 353–617.
- Beucher, A., C. B. Rasmussen, T. B. Moeslund, and M. H. Greve. 2022. "Interpretation of Convolutional Neural Networks for Acid Sulfate Soil Classification". *Frontiers in Environmental Science* 9: 809995.
- Biney, J. K. M., R. Vašát, J. R. Blöcher, L. Boruvka, and K. Němeček. 2022. "Using an Ensemble Model Coupled With Portable X-Ray Fluorescence and Visible Near-Infrared Spectroscopy to Explore the Viability of Mapping and Estimating Arsenic in an Agricultural Soil". *Science of the Total Environment* 818: 151805.
- Blanchy, G., L. Albrecht, J. Koestel, and S. Garré. 2023. "Potential of Natural Language Processing for Metadata Extraction From Environmental Scientific Publications". *Soil* 9: 155–168.
- Bogrekcı, I., and R. J. Godwin. 2007. "Development of an Image-Processing Technique for Soil Tilth Sensing". *Biosystems Engineering* 97: 323–331.
- Bohanec, M., J. Cortet, B. Griffiths, et al. 2007. "A Qualitative Multi-Attribute Model for Assessing the Impact of Cropping Systems on Soil Quality". *Pedobiologia* 51: 239–250.
- Booty, W. G., D. C.-L. Lam, I. W. S. Wong, and P. Siconolfi. 2001. "Design and Implementation of an Environmental Decision Support System". *Environmental Modelling & Software* 16: 453–458.
- Breiman, L. 2001. "Statistical Modeling: The Two Cultures (With Comments and a Rejoinder by the Author)". *Statistical Science* 16: 199–231.
- Brungard, C., T. Nauman, M. Duniway, et al. 2021. "Regional Ensemble Modeling Reduces Uncertainty for Digital Soil Mapping". *Geoderma* 397: 114998.
- Brungard, C. W., J. L. Boettinger, M. C. Duniway, S. A. Wills, and T. C. Edwards Jr. 2015. "Machine Learning for Predicting Soil Classes in Three Semi-Arid Landscapes". *Geoderma* 239: 68–83.
- Bui, E. N., B. L. Henderson, and K. Viergever. 2006. "Knowledge Discovery From Models of Soil Properties Developed Through Data Mining". *Ecological Modelling* 191: 431–446.
- Bui, E. N., A. Loughhead, and R. Corner. 1999. "Extracting Soil-Landscape Rules From Previous Soil Surveys". *Soil Research* 37: 495–508.
- Burrough, P. A. 1992. "Development of Intelligent Geographical Information Systems". *International Journal of Geographical Information Systems* 6: 1–11.
- Cahlik, T. 2000. "Comparison of the Maps of Science". *Scientometrics* 49: 373–387.
- Callon, M., J. P. Courtial, and F. Laville. 1991. "Co-Word Analysis as a Tool for Describing the Network of Interactions Between Basic and Technological Research: The Case of Polymer Chemistry". *Scientometrics* 22: 155–205.
- Canillas, E. C., and V. M. Salokhe. 2002. "A Decision Support System for Compaction Assessment in Agricultural Soils". *Soil and Tillage Research* 65: 221–230.
- Castelvecchi, D. 2024. "Researchers Built an 'AI Scientist' – What Can It Do?" Nature News. Accessed September 6, 2024. <https://www.nature.com/articles/d41586-024-02842-3>.

- Chakraborty, R., and S. C. Pal. 2023. "Modeling Soil Erosion Susceptibility Using GIS-Based Different Machine Learning Algorithms in Monsoon Dominated Diversified Landscape in India". *Modeling Earth Systems and Environment* 9: 2927–2942.
- Chandra, H., P. M. Pawar, R. Elakkiya, P. S. Tamizharasan, R. Muthalagu, and A. Panthakkan. 2023. "Explainable AI for Soil Fertility Prediction". *IEEE Access* 11: 97866–97878.
- Chaney, N. W., E. F. Wood, A. B. McBratney, et al. 2016. "POLARIS: A 30-Meter Probabilistic Soil Series Map of the Contiguous United States". *Geoderma* 274: 54–67.
- Chang, D.-H., and S. Islam. 2000. "Estimation of Soil Physical Properties Using Remote Sensing and Artificial Neural Network". *Remote Sensing of Environment* 74: 534–544.
- Chemchem, A., Q. Styc, L. A. Sow, et al. 2021. Terra Occitania: From Text Recognition to Soil Map Generation.
- Chen, Q., L. Li, C. Chong, and X. Wang. 2022. "Ai-Enhanced Soil Management and Smart Farming". *Soil Use and Management* 38: 7–13.
- Cialella, A. T., W. Lawrence, and E. Levine. 1997. "Predicting Soil Drainage Class Using Remotely Sensed and Digital Elevation Data". *Photogrammetric Engineering and Remote Sensing* 63: 171–177.
- Cobo, M. J., A. G. López-Herrera, E. Herrera-Viedma, and F. Herrera. 2011. "An Approach for Detecting, Quantifying, and Visualizing the Evolution of a Research Field: A Practical Application to the Fuzzy Sets Theory Field". *Journal of Informetrics* 5: 146–166.
- Colavizza, G., T. Blanke, C. Jeurgens, and J. Noordegraaf. 2021. "Archives and Ai: An Overview of Current Debates and Future Perspectives". *ACM Journal on Computing and Cultural Heritage (JOCCH)* 15: 1–15.
- da Silva, L. d. C. M., R. S. S. Amorim, E. I. Fernandes-Filho, E. D. Bocuti, and D. D. da Silva. 2023. "Pedotransfer Functions and Machine Learning: Advancements and Challenges in Tropical Soils". *Geoderma Regional* 35: e00720.
- Dale, M. B., A. B. McBratney, and J. S. Russell. 1989. "On the Role of Expert Systems and Numerical Taxonomy in Soil Classification". *Journal of Soil Science* 40: 223–234.
- Dasgupta, S., S. Debnath, A. Das, et al. 2023. "Developing Regional Soil Micronutrient Management Strategies Through Ensemble Learning Based Digital Soil Mapping". *Geoderma* 433: 116457.
- De Bruin, S., and A. Stein. 1998. "Soil-Landscape Modelling Using Fuzzy c-Means Clustering of Attribute Data Derived From a Digital Elevation Model (DEM)". *Geoderma* 83: 17–33.
- De la Rosa, D., F. Mayol, E. Daz-Pereira, M. Fernandez, and D. de la Rosa Jr. 2004. "A Land Evaluation Decision Support System (MicroLEIS DSS) for Agricultural Soil Protection: With Special Reference to the Mediterranean Region". *Environmental Modelling & Software* 19: 929–942.
- de Oliveira, V. A., A. F. Rodrigues, M. A. V. Morais, M. d. C. N. S. Terra, L. Guo, and C. R. de Mello. 2021. "Spatiotemporal Modelling of Soil Moisture in an Atlantic Forest Through Machine Learning Algorithms". *European Journal of Soil Science* 72: 1969–1987.
- de Sousa Mendes, W., J. A. M. Demattê, B. Minasny, et al. 2022. "Free Iron Oxide Content in Tropical Soils Predicted by Integrative Digital Mapping". *Soil and Tillage Research* 219: 105346.
- Debeljak, M., A. Trajanov, V. Kuzmanovski, et al. 2019. "A Field-Scale Decision Support System for Assessment and Management of Soil Functions". *Frontiers in Environmental Science* 7: 115.
- Diaz-Gonzalez, F. A., J. Vuelas, C. A. Correa, V. E. Vallejo, and D. Patino. 2022. "Machine Learning and Remote Sensing Techniques Applied to Estimate Soil Indicators—Review". *Ecological Indicators* 135: 108517.
- Doorn, N. 2021. "Artificial Intelligence in the Water Domain: Opportunities for Responsible Use". *Science of the Total Environment* 755: 142561.
- Douarre, C., R. Schielein, C. Frindel, S. Gerth, and D. Rousseau. 2018. "Transfer Learning From Synthetic Data Applied to Soil-Root Segmentation in X-Ray Tomography Images". *Journal of Imaging* 4: 65.
- Elton, D. C. 2020. "Self-Explaining AI as an Alternative to Interpretable AI". In *Artificial General Intelligence: 13th International Conference, AGI 2020, St. Petersburg, Russia, September 16–19, 2020, Proceedings 13*, edited by B. Goertzel, A. I. Panov, A. Potapov, and R. Yampolskiy, 95–106. Springer.
- European Commission. 2019. "High-Level Expert Group on Artificial Intelligence". *Ethics Guidelines for Trustworthy AI*: 6.
- Fajardo, M., A. B. McBratney, D. J. Field, and B. Minasny. 2016. "Soil Slaking Assessment Using Image Recognition". *Soil and Tillage Research* 163: 119–129.
- Fantappiè, M., G. L'Abate, C. Schillaci, and E. A. C. Costantini. 2023. "Digital Soil Mapping of Italy to Map Derived Soil Profiles With Neural Networks". *Geoderma Regional* 32: e00619.
- Ferreira, T. R., F. A. M. Cássaro, H. Zhou, and L. F. Pires. 2022. "X-Ray Computed Tomography Image Processing & Segmentation: A Case Study Applying Machine Learning and Deep Learning-Based Strategies". In *X-Ray Imaging of the Soil Porous Architecture*, edited by S. J. Mooney, I. M. Young, R. J. Heck, and S. Peth, 57–80. Springer Cham.
- Fisher, P. F. 1989. "Expert System Applications in Geography". *Area* 21: 279–287.
- Furey, J., A. Davis, and J. Seiter-Moser. 2019. "Natural Language Indexing for Pedoinformatics". *Geoderma* 334: 49–54.
- Galbraith, J. M., and R. B. Bryant. 1998. "A Functional Analysis of Soil Taxonomy in Relation to Expert System Techniques". *Soil Science* 163: 739–747.
- Galbraith, J. M., R. B. Bryant, and R. J. Ahrens. 1998. "An Expert System for Soil Taxonomy". *Soil Science* 163: 748–758.
- Ghorbani, A., M. Sadeghi, and S. B. Jones. 2021. "Towards New Soil Water Flow Equations Using Physics-Constrained Machine Learning". *Vadose Zone Journal* 20: e20136.
- Ghosh, T., P. P. Maity, S. M. F. Rabbi, T. K. Das, and R. Bhattacharyya. 2023. "Application of X-Ray Computed Tomography in Soil and Plant – A Review". *Frontiers in Environmental Science* 11: 1216630.
- Gibney, E. 2024. "The Ai Revolution Is Coming to Robots: How Will It Change Them?" *Nature* 630, no. 8015: 22–24. <https://doi.org/10.1038/d41586-024-01442-5>.
- Goodfellow, I., J. Pouget-Abadie, M. Mirza, et al. 2020. "Generative Adversarial Networks". *Communications of the ACM* 63: 139–144.
- Gougherty, A. V., and H. L. Clipp. 2024. "Testing the Reliability of an AI-Based Large Language Model to Extract Ecological Information From the Scientific Literature". *npj Biodiversity* 3: 13.
- Grunwald, S. 2022. "Artificial Intelligence and Soil Carbon Modeling Demystified: Power, Potentials, and Perils". *Carbon Footprints* 1: 1–5.
- Gruszczyński, S., and W. Gruszczyński. 2022. "Supporting Soil and Land Assessment With Machine Learning Models Using the vis-NIR Spectral Response". *Geoderma* 405: 115451.
- Haghverdi, A., W. Cornelis, and B. Ghahraman. 2012. "A Pseudo-Continuous Neural Network Approach for Developing Water Retention Pedotransfer Functions With Limited Data". *Journal of Hydrology* 442: 46–54.
- Han, Q., H. Bai, L. Liu, Y. Zhao, and Y. Zhao. 2021. "Model Representation and Quantitative Analysis of Pore Three-Dimensional Morphological

- Structure Based on Soil Computed Tomography Images". *European Journal of Soil Science* 72: 1530–1542.
- Han, Q., Y. Zhao, L. Liu, Y. Chen, and Y. Zhao. 2019. "A Simplified Convolutional Network for Soil Pore Identification Based on Computed Tomography Imagery". *Soil Science Society of America Journal* 83: 1309–1318.
- Hartemink, A. E., and A. B. McBratney. 2008. "A soil science renaissance". *Geoderma* 148: 123–129.
- Hartemink, A. E. 2006. *The Future of Soil Science*. IUSS International Union of Soil Sciences.
- Hastings, J. 2023. *AI for Scientific Discovery*. CRC Press.
- Hengl, T., M. A. Miller, J. Križan, et al. 2021. "African Soil Properties and Nutrients Mapped at 30 m Spatial Resolution Using Two-Scale Ensemble Machine Learning". *Scientific Reports* 11: 6130.
- Heung, B., H. C. Ho, J. Zhang, A. Knudby, C. E. Bulmer, and M. G. Schmidt. 2016. "An Overview and Comparison of Machine-Learning Techniques for Classification Purposes in Digital Soil Mapping". *Geoderma* 265: 62–77.
- Heuvelink, G. B. M., M. E. Angelini, L. Poggio, et al. 2021. "Machine Learning in Space and Time for Modelling Soil Organic Carbon Change". *European Journal of Soil Science* 72: 1607–1623.
- Holt, D. 1978. "Potentials for Artificial Intelligence and Supercomputers in Soil Science". In *Future Developments in Soil Science Research*, edited by L. L. Boersma, 459–468. American Society of Agronomy.
- Hou, L. H., W. Gao, Z. H. Weng, et al. 2022. "Use of X-Ray Tomography for Examining Root Architecture in Soils". *Geoderma* 405: 115405.
- Iqbal, J., R. Xu, H. Halloran, and C. Li. 2020. "Development of a Multi-Purpose Autonomous Differential Drive Mobile Robot for Plant Phenotyping and Soil Sensing". *Electronics* 9: 1550.
- Ji, W., V. I. Adamchuk, S. Chen, et al. 2019. "Simultaneous Measurement of Multiple Soil Properties Through Proximal Sensor Data Fusion: A Case Study". *Geoderma* 341: 111–128.
- Jorda, H., M. Bechtold, N. Jarvis, and J. Koestel. 2015. "Using Boosted Regression Trees to Explore Key Factors Controlling Saturated and Near-Saturated Hydraulic Conductivity". *European Journal of Soil Science* 66: 744–756.
- Kaur, R., R. Malik, and M. Paul. 2007. "Long-Term Effects of Various Crop Rotations for Managing Salt-Affected Soils Through a Field Scale Decision Support System—a Case Study". *Soil Use and Management* 23: 52–62.
- Kedron, P., A. E. Frazier, A. B. Trgovac, T. Nelson, and A. S. Fotheringham. 2021. "Reproducibility and Replicability in Geographical Analysis". *Geographical Analysis* 53: 135–147.
- Kerebel, A., R. Cassidy, P. Jordan, and N. M. Holden. 2013. "Soil Moisture Deficit as a Predictor of the Trend in Soil Water Status of Grass Fields". *Soil Use and Management* 29: 419–431.
- Khaledian, Y., and B. A. Miller. 2020. "Selecting Appropriate Machine Learning Methods for Digital Soil Mapping". *Applied Mathematical Modelling* 81: 401–418.
- Kitić, G., D. Krklješ, M. Panić, C. Petes, S. Birgermajer, and V. Crnojević. 2022. "Agrobot Lala—An Autonomous Robotic System for Real-Time, In-Field Soil Sampling, and Analysis of Nitrates". *Sensors* 22: 4207.
- Koch, J., S. Stisen, J. C. Refsgaard, V. Ernsten, P. R. Jakobsen, and A. L. Højberg. 2019. "Modeling Depth of the Redox Interface at High Resolution at National Scale Using Random Forest and Residual Gaussian Simulation". *Water Resources Research* 55: 1451–1469.
- Koekkoek, E. J. W., and H. Bultink. 1999. "Neural Network Models to Predict Soil Water Retention". *European Journal of Soil Science* 50: 489–495.
- Konkol, M., C. Kray, and M. Pfeiffer. 2019. "Computational Reproducibility in Geoscientific Papers: Insights From a Series of Studies With Geoscientists and a Reproduction Study". *International Journal of Geographical Information Science* 33: 408–429.
- Koop, A. N., D. R. Hirmas, S. A. Billings, et al. 2023. "Is Macroporosity Controlled by Complexed Clay and Soil Organic Carbon?" *Geoderma* 437: 116565.
- Kotu, V., and B. Deshpande. 2018. *Data Science: Concepts and Practice*. Morgan Kaufmann.
- Krenn, M., R. Pollice, S. Y. Guo, et al. 2022. "On Scientific Understanding With Artificial Intelligence". *Nature Reviews Physics* 4: 761–769.
- Lacoste, M., B. Lemerrier, and C. Walter. 2011. "Regional Mapping of Soil Parent Material by Machine Learning Based on Point Data". *Geomorphology* 133: 90–99.
- Lacoste, M., B. Minasny, A. B. McBratney, D. Michot, V. Viaud, and C. Walter. 2014. "High Resolution 3D Mapping of Soil Organic Carbon in a Heterogeneous Agricultural Landscape". *Geoderma* 213: 296–311.
- Lagacherie, P., M. Arregui, and D. Fages. 2024. "Evaluating the Quality of Soil Legacy Data Used as Input of Digital Soil Mapping Models". *European Journal of Soil Science* 75: e13463.
- Lavrukhin, E. V., K. M. Gerke, K. A. Romanenko, K. N. Abrosimov, and M. V. Karsanina. 2021. "Assessing the Fidelity of Neural Network-Based Segmentation of Soil XCT Images Based on Pore-Scale Modelling of Saturated Flow Properties". *Soil and Tillage Research* 209: 104942.
- Legg, S., and M. Hutter. 2007. "Universal Intelligence: A Definition of Machine Intelligence". *Minds and Machines* 17: 391–444.
- Li, X., H. Wang, S. Qin, L. Lin, X. Wang, and W. Cornelis. 2024. "Evaluating Ensemble Learning in Developing Pedotransfer Functions to Predict Soil Hydraulic Properties". *Journal of Hydrology* 640: 131658.
- Liu, X., A.-X. Zhu, L. Yang, et al. 2020. "A Graded Proportion Method of Training Sample Selection for Updating Conventional Soil Maps". *Geoderma* 357: 113939.
- Lu, C., C. Lu, R. T. Lange, J. Foerster, J. Clune, and D. Ha. 2024. "The AI Scientist: Towards Fully Automated Open-Ended Scientific Discovery". Accessed September 26, 2024. <https://www.arxiv.org/abs/2408.06292>.
- MacMillan, R. A., W. W. Pettapiece, and J. A. Brierley. 2005. "An Expert System for Allocating Soils to Landforms Through the Application of Soil Survey Tacit Knowledge". *Canadian Journal of Soil Science* 85: 103–112.
- Malone, B. P., Z. Luo, D. He, R. A. V. R. Rossel, and E. Wang. 2020. "Bioclimatic Variables as Important Spatial Predictors of Soil Hydraulic Properties Across Australia's Agricultural Region". *Geoderma Regional* 23: e00344.
- Mata, J., I. De Miguel, R. J. Duran, et al. 2018. "Artificial Intelligence (AI) Methods in Optical Networks: A Comprehensive Survey". *Optical Switching and Networking* 28: 43–57.
- McBratney, A. B. 2024. "Developing the New Soil Science—Advice for Early-Career Soil Scientists". *Pedosphere* 34: 3–4.
- McBratney, A. B., and I. O. A. Odeh. 1997. "Application of Fuzzy Sets in Soil Science: Fuzzy Logic, Fuzzy Measurements and Fuzzy Decisions". *Geoderma* 77: 85–113.
- McBratney, A. B., M. L. M. Santos, and B. Minasny. 2003. "On Digital Soil Mapping". *Geoderma* 117: 3–52.
- McBride, M. B. 2022. "Estimating Soil Chemical Properties by Diffuse Reflectance Spectroscopy: Promise Versus Reality". *European Journal of Soil Science* 73: e13192.
- McCarthy, J. 2007. What Is Artificial Intelligence.

- McCarthy, J., M. L. Minsky, N. Rochester, and C. E. Shannon. 1955. A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence. August 31, 1955.
- McCracken, R. J., and R. B. Cate. 1986. "Artificial Intelligence, Cognitive Science, and Measurement Theory Applied in Soil Classification". *Soil Science Society of America Journal* 50: 557–561.
- McShea, D. W. 2013. "Machine Wanting". *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences* 44: 679–687.
- Merriam-Webster Dictionary. 2011. "Merriam-Webster". Accessed November 14, 2024.
- Minasny, B., T. Bandai, T. A. Ghezzehei, et al. 2024. "Soil Science-Informed Machine Learning". *Geoderma* 452: 117094.
- Minasny, B., Ö. Berglund, J. Connolly, et al. 2019. "Digital Mapping of Peatlands—a Critical Review". *Earth-Science Reviews* 196: 102870.
- Minasny, B., N. W. Chaney, F. Maggi, et al. 2021. "Pedotransfer Functions for Estimating Soil Hydraulic Properties From Saturation to Dryness". *Geoderma* 403: 115194.
- Minasny, B., and A. B. McBratney. 2002. "The Neuro-m Method for Fitting Neural Network Parametric Pedotransfer Functions". *Soil Science Society of America Journal* 66: 352–361.
- Minasny, B., and A. B. McBratney. 2007. "Incorporating Taxonomic Distance Into Spatial Prediction and Digital Mapping of Soil Classes". *Geoderma* 142: 285–293.
- Minasny, B., and A. B. McBratney. 2016. "Digital Soil Mapping: A Brief History and Some Lessons". *Geoderma* 264: 301–311.
- Minasny, B., A. B. McBratney, B. P. Malone, and I. Wheeler. 2013. "Digital Mapping of Soil Carbon". *Advances in Agronomy* 118: 1–47.
- Minsky, M. L. 1969. *Semantic Information Processing*. MIT Press.
- Montas, H., and C. A. Madramootoo. 1992. "A Decision Support System for Soil Conservation Planning". *Computers and Electronics in Agriculture* 7: 187–202.
- Mulder, V. L., S. De Bruin, M. E. Schaepman, and T. R. Mayr. 2011. "The Use of Remote Sensing in Soil and Terrain Mapping—A Review". *Geoderma* 162: 1–19.
- Nawar, S., and A. M. Mouazen. 2022. "Combining Mid Infrared Spectroscopy With Stacked Generalisation Machine Learning for Prediction of Key Soil Properties". *European Journal of Soil Science* 73: e13323.
- Nemes, A., W. J. Rawls, and Y. A. Pachepsky. 2005. "Influence of Organic Matter on the Estimation of Saturated Hydraulic Conductivity". *Soil Science Society of America Journal* 69: 1330–1337.
- Nemes, A., W. J. Rawls, Y. A. Pachepsky, and M. T. van Genuchten. 2006. "Sensitivity Analysis of the Nonparametric Nearest Neighbor Technique to Estimate Soil Water Retention". *Vadose Zone Journal* 5: 1222–1235.
- Nestor, M., F. Loredana, P. Raymond, et al. 2024. *The AI Index 2024 Annual Report*. AI Index Steering Committee, Institute for Human-Centered AI, Stanford University.
- Ng, W., B. Minasny, W. d. S. Mendes, and J. A. M. Demattê. 2020. "The Influence of Training Sample Size on the Accuracy of Deep Learning Models for the Prediction of Soil Properties With Near-Infrared Spectroscopy Data". *Soil* 6: 565–578.
- Nilsson, N. J. 2009. *The Quest for Artificial Intelligence*. Cambridge University Press.
- Nussbaum, M., K. Spiess, A. Baltensweiler, et al. 2018. "Evaluation of Digital Soil Mapping Approaches With Large Sets of Environmental Covariates". *Soil* 4: 1–22.
- Odebiri, O., J. Odindi, and O. Mutanga. 2021. "Basic and Deep Learning Models in Remote Sensing of Soil Organic Carbon Estimation: A Brief Review". *International Journal of Applied Earth Observation and Geoinformation* 102: 102389.
- Odgers, N. P., A. B. McBratney, and B. Minasny. 2015. "Digital Soil Property Mapping and Uncertainty Estimation Using Soil Class Probability Rasters". *Geoderma* 237: 190–198.
- Odgers, N. P., A. B. McBratney, B. Minasny, W. Sun, and D. Clifford. 2014. "DSMART: An Algorithm to Spatially Disaggregate Soil Map Units". In *GlobalSoilMap: Basis of the Global Spatial Soil Information System*, edited by D. Arrouays, N. McKenzie, J. Hempel, A. C. Richer-de Forges, and A. B. McBratney, 261–266. CRC Press.
- Pachepsky, Y., and W. J. Rawls. 2004. *Development of Pedotransfer Functions in Soil Hydrology*. Vol. 30. Elsevier.
- Pachepsky, Y. A., and M. G. Schaap. 2004. "Data Mining and Exploration Techniques". *Developments in Soil Science* 30: 21–32.
- Padarian, J., and I. Fuentes. 2019. "Word Embeddings for Application in Geosciences: Development, Evaluation, and Examples of Soil-Related Concepts". *Soil* 5: 177–187.
- Padarian, J., and A. B. McBratney. 2020. "A New Model for Intra- and Inter-Institutional Soil Data Sharing". *Soil* 6: 89–94.
- Padarian, J., B. Minasny, and A. B. McBratney. 2019a. "Using Deep Learning for Digital Soil Mapping". *Soil* 5: 79–89.
- Padarian, J., B. Minasny, and A. B. McBratney. 2019b. "Using Deep Learning to Predict Soil Properties From Regional Spectral Data". *Geoderma Regional* 16: e00198.
- Padarian, J., B. Minasny, and A. B. McBratney. 2020. "Machine Learning and Soil Sciences: A Review Aided by Machine Learning Tools". *Soil* 6: 35–52.
- Pasolli, L., C. Notarnicola, G. Bertoldi, et al. 2014. "Soil Moisture Monitoring in Mountain Areas by Using High-Resolution SAR Images: Results From a Feasibility Study". *European Journal of Soil Science* 65: 852–864.
- Piikki, K., J. Wetterlind, M. Söderström, and B. Stenberg. 2021. "Perspectives on Validation in Digital Soil Mapping of Continuous Attributes—A Review". *Soil Use and Management* 37: 7–21.
- Poggio, L., L. M. De Sousa, N. H. Batjes, et al. 2021. "SoilGrids 2.0: Producing Soil Information for the Globe With Quantified Spatial Uncertainty". *Soil* 7: 217–240.
- Pournader, M., H. Ghaderi, A. Hassanzadegan, and B. Fahimnia. 2021. "Artificial Intelligence Applications in Supply Chain Management". *International Journal of Production Economics* 241: 108250.
- Power, D. J. 2007. "A Brief History of Decision Support Systems". Accessed July 17, 2024. <https://dssresources.com/history/dsshistory.html>.
- Priori, S., N. Bianconi, and E. A. C. Costantini. 2014. "Can/-Radiometrics Predict Soil Textural Data and Stoniness in Different Parent Materials? A Comparison of Two Machine-Learning Methods". *Geoderma* 226: 354–364.
- Reichstein, M., G. Camps-Valls, B. Stevens, M. Jung, J. Denzler, and N. Carvalhais. 2019. "Deep Learning and Process Understanding for Data-Driven Earth System Science". *Nature* 566: 195–204.
- Rezaei, M., S. R. Mousavi, A. Rahmani, et al. 2023. "Incorporating Machine Learning Models and Remote Sensing to Assess the Spatial Distribution of Saturated Hydraulic Conductivity in a Light-Textured Soil". *Computers and Electronics in Agriculture* 209: 107821.
- Romero-Ruiz, A., D. O'Leary, E. Daly, et al. 2024. "An Agrogeophysical Modelling Framework for the Detection of Soil Compaction Spatial Variability due to Grazing Using Field-Scale Electromagnetic Induction Data". *Soil Use and Management* 40: e13039.
- Russell, S. J., and P. Norvig. 2016. *Artificial Intelligence: A Modern Approach*. Pearson.

- Saadatpour, M., A. Afshar, and M. H. Afshar. 2011. "Fuzzy Pattern Recognition Method for Assessing Soil Erosion". *Environmental Monitoring and Assessment* 180: 385–397.
- Sandén, T., A. Trajanov, H. Spiegel, et al. 2019. "Development of an Agricultural Primary Productivity Decision Support Model: A Case Study in France". *Frontiers in Environmental Science* 7: 58.
- Sarangi, A., C. A. Madramootoo, and C. Cox. 2004. "A Decision Support System for Soil and Water Conservation Measures on Agricultural Watersheds". *Land Degradation & Development* 15: 49–63.
- Schaap, M. G., F. J. Leij, and M. T. Van Genuchten. 1998. "Neural Network Analysis for Hierarchical Prediction of Soil Hydraulic Properties". *Soil Science Society of America Journal* 62: 847–855.
- Schröder, J. J., R. P. O. Schulte, R. E. Creamer, et al. 2016. "The Elusive Role of Soil Quality in Nutrient Cycling: A Review". *Soil Use and Management* 32: 476–486.
- Schweizer, S. A., C. Hoeschen, S. Schlüter, I. Kögel-Knabner, and C. W. Mueller. 2018. "Rapid Soil Formation After Glacial Retreat Shaped by Spatial Patterns of Organic Matter Accrual in Microaggregates". *Global Change Biology* 24: 1637–1650.
- Schwenke, G., L. Beange, J. Cameron, M. Bell, and S. Harden. 2019. "What Soil Information Do Crop Advisors Use to Develop Nitrogen Fertilizer Recommendations for Grain Growers in New South Wales, Australia?" *Soil Use and Management* 35: 85–93.
- Sharififar, A., F. Sarmadian, B. P. Malone, and B. Minasny. 2019. "Addressing the Issue of Digital Mapping of Soil Classes With Imbalanced Class Observations". *Geoderma* 350: 84–92.
- Shen, Z., L. Ramirez-Lopez, T. Behrens, et al. 2022. "Deep Transfer Learning of Global Spectra for Local Soil Carbon Monitoring". *ISPRS Journal of Photogrammetry and Remote Sensing* 188: 190–200.
- Shen, Z., and R. A. Viscarra Rossel. 2021. "Automated Spectroscopic Modelling With Optimised Convolutional Neural Networks". *Scientific Reports* 11: 208.
- Sikora, F. J., P. Howe, D. Reid, D. Morgan, and E. Zimmer. 2011. "Adopting a Robotic Ph Instrument for Soil and Soil-Buffer pH Measurements in a Soil Test Laboratory". *Communications in Soil Science and Plant Analysis* 42: 617–632.
- Skidmore, A. K., P. J. Ryan, W. Dawes, D. Short, and E. O'LOUGHLIN. 1991. "Use of an Expert System to Map Forest Soils From a Geographical Information System". *International Journal of Geographical Information Systems* 5: 431–445.
- Smaling, E. M. A., and L. O. Fresco. 1993. "A Decision-Support Model for Monitoring Nutrient Balances Under Agricultural Land Use (NUTMON)". *Geoderma* 60: 235–256.
- Smith, A. G., J. Petersen, R. Selvan, and C. R. Rasmussen. 2020. "Segmentation of Roots in Soil With U-Net". *Plant Methods* 16: 13.
- Soltaninejad, M., C. J. Sturrock, M. Griffiths, T. P. Pridmore, and M. P. Pound. 2020. "Three Dimensional Root CT Segmentation Using Multi-Resolution Encoder-Decoder Networks". *IEEE Transactions on Image Processing* 29: 6667–6679.
- Steffens, M., D. M. Rogge, C. W. Mueller, et al. 2017. "Identification of Distinct Functional Microstructural Domains Controlling C Storage in Soil". *Environmental Science & Technology* 51: 12182–12189.
- Steinruecken, C., E. Smith, D. Janz, J. Lloyd, and Z. Ghahramani. 2019. "The Automatic Statistician". In *Automated Machine Learning: Methods, Systems, Challenges*, edited by F. Hutter, L. Kotthoff, and J. Vanschoren, 161–173. Springer International Publishing.
- Tamari, S., J. H. M. Wösten, and J. C. Ruiz-Suarez. 1996. "Testing an Artificial Neural Network for Predicting Soil Hydraulic Conductivity". *Soil Science Society of America Journal* 60: 1732–1741.
- Tesch, T., S. Kollet, and J. Garcke. 2023. "Causal Deep Learning Models for Studying the Earth System". *Geoscientific Model Development* 16: 2149–2166.
- Todd-Brown, K. E. O., R. Z. Abramoff, J. Beem-Miller, et al. 2022. "Reviews and Syntheses: The Promise of Big Diverse Soil Data, Moving Current Practices Towards Future Potential". *Biogeosciences* 19: 3505–3522.
- Trajanov, A., J. Schröder, D. Wall, A. Delgado, R. Schulte, and M. Debeljak. 2019. "Assessing the Nutrient Cycling Potential in Agricultural Soils Using Decision Modelling". In *15th International Symposium on Operational Research, SOR 2019*, edited by L. Z. Stirn, M. K. Borstnar, J. Zerovnik, S. Drobne, and J. Povh, 23–27. Slovenian Society Informatika.
- Tsakiridis, N. L., N. V. Tziolas, J. B. Theocharis, and G. C. Zalidis. 2019. "A Genetic Algorithm-Based Stacking Algorithm for Predicting Soil Organic Matter From Vis-Nir Spectral Data". *European Journal of Soil Science* 70: 578–590.
- Turing, A. M. 1950. "1-Computing Machinery and Intelligence". *Mind* 59: 433.
- Twarakavi, N. K. C., J. Simunek, and M. G. Schaap. 2009. "Development of Pedotransfer Functions for Estimation of Soil Hydraulic Parameters Using Support Vector Machines". *Soil Science Society of America Journal* 73: 1443–1452.
- Van Alphen, B. J., and J. J. Stoorvogel. 2000. "A Functional Approach to Soil Characterization in Support of Precision Agriculture". *Soil Science Society of America Journal* 64: 1706–1713.
- Van de Broek, M., C. B. Henriksen, B. B. Ghaley, et al. 2019. "Assessing the Climate Regulation Potential of Agricultural Soils Using a Decision Support Tool Adapted to Stakeholders' Needs and Possibilities". *Frontiers in Environmental Science* 7: 131.
- van der Westhuizen, S., G. B. M. Heuvelink, and D. P. Hofmeyr. 2023. "Multivariate Random Forest for Digital Soil Mapping". *Geoderma* 431: 116365.
- Van Leeuwen, J. P., R. E. Creamer, D. Cluzeau, et al. 2019. "Modeling of Soil Functions for Assessing Soil Quality: Soil Biodiversity and Habitat Provisioning". *Frontiers in Environmental Science* 7: 113.
- Van Looy, K., J. Bouma, M. Herbst, et al. 2017. "Pedotransfer Functions in Earth System Science: Challenges and Perspectives". *Reviews of Geophysics* 55: 1199–1256.
- Vereecken, H., M. Weynants, M. Javaux, Y. Pachepsky, M. G. Schaap, and M. T. Van Genuchten. 2010. "Using Pedotransfer Functions to Estimate the van Genuchten–Mualem Soil Hydraulic Properties: A Review". *Vadose Zone Journal* 9: 795–820.
- Vincent, S., B. Lemerrier, L. Berthier, and C. Walter. 2018. "Spatial Disaggregation of Complex Soil Map Units at the Regional Scale Based on Soil-Landscape Relationships". *Geoderma* 311: 130–142.
- Viscarra Rossel, R. A., V. I. Adamchuk, K. A. Sudduth, N. J. McKenzie, and C. Lobsey. 2011. "Proximal Soil Sensing: An Effective Approach for Soil Measurements in Space and Time". *Advances in Agronomy* 113: 243–291.
- Viscarra Rossel, R. A., and T. Behrens. 2010. "Using Data Mining to Model and Interpret Soil Diffuse Reflectance Spectra". *Geoderma* 158: 46–54.
- Viscarra Rossel, R. A., T. Behrens, E. Ben-Dor, et al. 2022. "Diffuse Reflectance Spectroscopy for Estimating Soil Properties: A Technology for the 21st Century". *European Journal of Soil Science* 73: e13271.
- Wadoux, A. M. J.-C. 2023. "Interpretable Spectroscopic Modelling of Soil With Machine Learning". *European Journal of Soil Science* 74: e13370.
- Wadoux, A. M. J.-C., G. B. M. Heuvelink, R. M. Lark, et al. 2021a. "Ten Challenges for the Future of Pedometrics". *Geoderma* 401: 115155.

- Wadoux, A. M. J.-C., and A. B. McBratney. 2021. "Digital Soil Science and Beyond". *Soil Science Society of America Journal* 85: 1313–1331.
- Wadoux, A. M. J.-C., B. Minasny, and A. B. McBratney. 2020a. "Machine Learning for Digital Soil Mapping: Applications, Challenges and Suggested Solutions". *Earth-Science Reviews* 210: 103359.
- Wadoux, A. M. J.-C., M. Román-Dobarco, and A. B. McBratney. 2021b. "Perspectives on Data-Driven Soil Research". *European Journal of Soil Science* 72: 1675–1689.
- Wadoux, A. M. J.-C., N. Saby, and M. P. Martin. 2023. "Shapley Values Reveal the Drivers of Soil Organic Carbon Stock Prediction". *Soil* 9: 21–38.
- Wadoux, A. M. J.-C., A. Samuel-Rosa, L. Poggio, and V. L. Mulder. 2020b. "A Note on Knowledge Discovery and Machine Learning in Digital Soil Mapping". *European Journal of Soil Science* 71: 133–136.
- Wall, D. P., A. Delgado, L. O'Sullivan, et al. 2020. "A Decision Support Model for Assessing the Water Regulation and Purification Potential of Agricultural Soils Across Europe". *Frontiers in Sustainable Food Systems* 4: 115.
- Wang, J., J. Ding, D. Yu, et al. 2020. "Machine Learning-Based Detection of Soil Salinity in an Arid Desert Region, Northwest China: A Comparison Between Landsat-8 OLI and Sentinel-2 MSI". *Science of the Total Environment* 707: 136092.
- Wang, S., Y. Chen, M. Wang, and J. Li. 2019. "Performance Comparison of Machine Learning Algorithms for Estimating the Soil Salinity of Salt-Affected Soil Using Field Spectral Data". *Remote Sensing* 11: 2605.
- Wang, Y., J. Zhou, R. Ma, G. Zhu, and Y. Zhang. 2022. "Development of a New Pedotransfer Function Addressing Limitations in Soil Hydraulic Models and Observations". *Water Resources Research* 58: e2021WR031406.
- Weber, T. K. D., L. Weihermüller, A. Nemes, et al. 2024. "Hydro-Pedotransfer Functions: A Roadmap for Future Development". *Hydrology and Earth System Sciences* 28: 3391–3433.
- Wieland, R., C. Ukawa, M. Joschko, et al. 2021. "Use of Deep Learning for Structural Analysis of Computer Tomography Images of Soil Samples". *Royal Society Open Science* 8: 201275.
- Xu, S., Y. Zhao, M. Wang, and X. Shi. 2022. "A Comparison of Machine Learning Algorithms for Mapping Soil Iron Parameters Indicative of Pedogenic Processes by Hyperspectral Imaging of Intact Soil Profiles". *European Journal of Soil Science* 73: e13204.
- Young, M. D., G. H. Ros, and W. de Vries. 2021. "A Decision Support Framework Assessing Management Impacts on Crop Yield, Soil Carbon Changes and Nitrogen Losses to the Environment". *European Journal of Soil Science* 72: 1590–1606.
- Zhang, L., G. B. M. Heuvelink, V. L. Mulder, S. Chen, X. Deng, and L. Yang. 2024. "Using Process-Oriented Model Output to Enhance Machine Learning-Based Soil Organic Carbon Prediction in Space and Time". *Science of the Total Environment* 922: 170778.
- Zhang, M., W. Shi, and Z. Xu. 2020. "Systematic Comparison of Five Machine-Learning Models in Classification and Interpolation of Soil Particle Size Fractions Using Different Transformed Data". *Hydrology and Earth System Sciences* 24: 2505–2526.
- Zhang, T., and R. Berndtsson. 1991. "Analysis of Soil Water Dynamics in Time and Space by Use of Pattern Recognition". *Water Resources Research* 27: 1623–1636.
- Zhang, W., L. Zhu, Q. Zhuang, D. Chen, and T. Sun. 2023. "Mapping Cropland Soil Nutrients Contents Based on Multi-Spectral Remote Sensing and Machine Learning". *Agriculture* 13: 1592.
- Zwetsloot, M. J., J. van Leeuwen, L. Hemerik, et al. 2021. "Soil Multifunctionality: Synergies and Trade-Offs Across European Climatic Zones and Land Uses". *European Journal of Soil Science* 72: 1640–1654.