Science of the Total Environment 944 (2024) 173720

Contents lists available at ScienceDirect

Science of the Total Environment

journal homepage: www.elsevier.com/locate/scitotenv

How can we quantify, explain, and apply the uncertainty of complex soil maps predicted with neural networks?

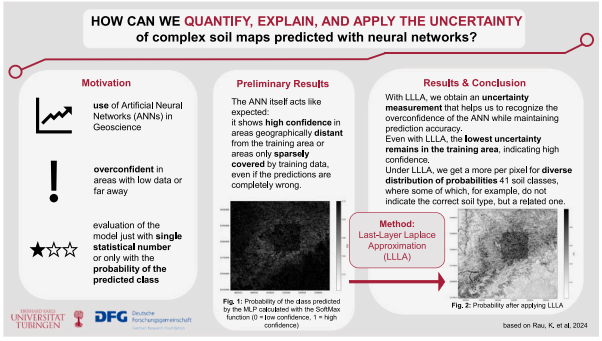
Kerstin Rau [a,c,d,*], Katharina Eggensperger [b,c], Frank Schneider [b], Philipp Hennig [b,c,d], Thomas Scholten [a,c]

[a] Department of Geoscience, University of Tübingen, Rümelinstraße 19-23, Tübingen 72070, Baden-Württemberg, Germany
[b] Department of Computer Science, University of Tübingen, Maria-von-Linden-Straße 6, Tübingen 72076, Baden-Württemberg, Germany
[c] Cluster of Excellence Machine Learning: New Perspectives for Sciene, University of Tübingen, Maria-von-Linden-Straße 6, Tübingen 72076, Baden-Württemberg, Germany
[d] Tübingen AI Center, Maria-von-Linden-Straße 6, Tübingen 72076, Baden-Württemberg, Germany

HIGHLIGHTS

- Solving the overconfidence problem of ANNs in extrapolation tasks
- Obtain a reliable uncertainty map using the Last-Layer Laplace Approximation method
- Provide insight into the predictability of soil types and identify knowledge gaps
- This research enhances the reliability and applicability of ANNs

GRAPHICAL ABSTRACT



ARTICLE INFO

Editor: Paulo Pereira

Keywords:
Spatial prediction
Model uncertainty
Digital soil mapping
Soil type
Last-layer Laplace approximation

ABSTRACT

Artificial neural networks (ANNs) have proven to be a useful tool for complex questions that involve large amounts of data. Our use case of predicting soil maps with ANNs is in high demand by government agencies, construction companies, or farmers, given cost and time intensive field work. However, there are two main challenges when applying ANNs. In their most common form, deep learning algorithms do not provide interpretable predictive uncertainty. This means that properties of an ANN such as the certainty and plausibility of the predicted variables, rely on the interpretation by experts rather than being quantified by evaluation metrics validating the ANNs. Further, these algorithms have shown a high confidence in their predictions in areas geographically distant from the training area or areas sparsely covered by training data. To tackle these challenges, we use the Bayesian deep learning approach "last-layer Laplace approximation", which is specifically designed to quantify uncertainty into deep networks, in our explorative study on soil classification. It corrects the overconfident areas without reducing the accuracy of the predictions, giving us a more realistic uncertainty expression of the model's prediction. In our study area in southern Germany, we subdivide the soils into soil

* Corresponding author at: Department of Geoscience, University of Tübingen, Rümelinstraße 19-23, Tübingen 72070, Baden-Württemberg, Germany.
E-mail addresses: kerstin.rau@uni-tuebingen.de (K. Rau), katharina.eggensperger@uni-tuebingen.de (K. Eggensperger), f.schneider@uni-tuebingen.de (F. Schneider), philipp.hennig@uni-tuebingen.de (P. Hennig), thomas.scholten@uni-tuebingen.de (T. Scholten).

https://doi.org/10.1016/j.scitotenv.2024.173720
Received 2 February 2024; Received in revised form 31 May 2024; Accepted 31 May 2024
Available online 10 June 2024

regions and as a test case we explicitly exclude two soil regions in the training area but include these regions in the prediction. Our results emphasize the need for uncertainty measurement to obtain more reliable and interpretable results of ANNs, especially for regions far away from the training area. Moreover, the knowledge gained from this research addresses the problem of overconfidence of ANNs and provides valuable information on the predictability of soil types and the identification of knowledge gaps. By analyzing regions where the model has limited data support and, consequently, high uncertainty, stakeholders can recognize the areas that require more data collection efforts.

1. Introduction

The use of machine learning in science has become incredibly valuable and has significantly transformed many areas of research. The number of studies in which methods from the field of machine learning (ML) are used is constantly increasing (Zhang et al., 2022). Soil science is one of the pioneers here, where extensive applications in the field of soil mapping were already developed at the beginning of this century (Behrens et al., 2005; McBratney et al., 2003). Today, digital soil mapping is one of the largest areas in which the methods are widely used for all kinds of climatic and geomorphometric regions of the World and in different areas of soil science, which has been demonstrated by numerous papers (Minasny and McBratney, 2016; Rentschler et al., 2022; Scull et al., 2003; Taghizadeh-Mehrjardi et al., 2021b; Zhang et al., 2022). Methodologically, applications of ML in soil science range from linear regression to modelling soil properties and their relationships to complex deep learning methods (Moore et al., 1993; Veres et al., 2015). The increasing use of these methods is not only due to their suitability for soil scientific and geographical questions, but also because producing soil type maps in the traditional way with cartographers surveying the landscape is very costly and time-consuming. This effort can be reduced with machine learning, especially for larger or even difficult to access areas (Behrens et al., 2005; Grunwald et al., 2011; Hewitt, 1993). At the same time, machine learning methods and their source code are becoming more accessible due to open-source software and widely available computational resources (Dramsch, 2020), and with the publication of several large open source datasets containing digital elevation models, climate data and other remote sensing data, especially those describing the vegetation, it is getting more convenient to apply them (Gascon et al., 2017; McBratney et al., 2003).

Looking at the properties and functions of soils, for example carbon and water storage and plant nutrition, the soil type as a highly integrated prediction variable has the advantage that we can infer mechanical properties, dynamic processes and general characteristics from it with little effort (Albrecht et al., 2005; Hartemink and Bockheim, 2013). For example, Zhou et al. (2004) showed new spatial patterns in the predicted soil type map with their Bayesian predictive modelling approach. Grinand et al. (2008) uses classification tree analysis, which also supports decision-making in soil map extrapolation using machine learning methods. Adhikari et al. (2014) compared an existing soil map from a field survey with a predicted map calculated using a decision tree model. Artificial neural networks (ANNs) are currently one of the most popular machine learning methods (Taghizadeh-Mehrjardi et al., 2020, 2021a), as they are able to process large amounts of data and compute predictions comparably fast (Haykin, 1998; Schmidhuber, 2015; Silveira et al., 2013). Brungard et al. (2015) predicted soil taxonomy classes using eleven different models and found that the complex models containing neural networks were more accurate. Furthermore, Zhu (2000) found that ANNs can be used to obtain high-resolution soil maps. Although Heung et al. (2016) has also achieved good results with ANNs, they also have to admit that ANNs are difficult to interpret. Despite the results being rich in information, a major drawback of the predicted soil maps, and especially of the survey maps, is that they do not quantify the uncertainty of the individual soil types at a given geographical location (Hengl et al., 2017). Instead, mostly is only given an overall accuracy statement in the form of a single statistical number. This is usually calculated as a coefficient using cross-validation techniques, where a subset of the training dataset is used to quantify the uncertainty of the overall performance (Wadoux et al., 2020). However, this is not sufficient, especially for regional or global tasks using unbalanced data sets, and that further analysis on uncertainty statements is needed, which was highlighted by (Meyer and Pebesma, 2022). Also, studies considering the uncertainty of predicted classes, like soil or vegetation classes, only looking at the probability of the predicted class or its confidence interval, have been criticized as well (Wadoux et al., 2020). They reported that out of 175 papers, only 30 % included uncertainty quantification, most were focused on achieving high prediction accuracy and only a handful used machine learning methods for the uncertainty quantification. It is obvious that a better understanding and quantification of the uncertainty of soil maps modelled with ML is needed, especially when extrapolating from the training domain or when transferring the model to other more or less similar domains. In particular, working with ANNs as a black box requires such an assessment, as this model class is also known to be overconfident (Breiman, 2001; Nguyen et al., 2015; Hein et al., 2019). This means that ANNs can predict very reliable results, in our case soil classes, with a probability of up to 100 %, even if the input data is incorrect or uncertain. The lack of uncertainty measurement by the ANNs themselves makes it difficult to assess the reliability of the model predictions, which can lead to misinterpretations and incorrect decisions (Guo et al., 2017). With this study, we apply an ANN that predicts soil types inside and outside the known training domain in a trial study. We quantify the uncertainty of our model at every pixel in the area using last-layer Laplace Approximation (LLLA) (Kristiadi et al., 2020). Our aim is to add this uncertainty measurement to a soil classification problem to identify and correct the overconfidence of ANNs and to be able to spatially analyse and interpret in a following step the prediction of the ANN and its uncertainty derived from the LLLA. Further, we will discuss the transferability of the ANN to adjacent similar areas. Overall, our analyses will help to better understand and interpret results from ML models in soil science to provide new insights into soil processes and the spatial structure of the different domains.

2. Material and methods

2.1. Study area

Our study area is located in the centre of Baden-Württemberg in Germany and covers an area of about 35 km^2. The average altitude in this region is 504 m, but if we look at the Swabian Jura (SJ) and the rest of the region separately, we have an average of 731 m and 444 m respectively, which can also be seen in Fig. 1A. The Neckar valley with its tributaries dominates this area, which extends from the southwest to the northeast over the area of the study site. In between are predominantly agriculturally used landscapes with settlements and towns, and the Schönbuch with its extensive, characteristic forests.

In the southeast is the SJ, which is characterized by its unique maritime geologic formation with calcaric sedimentary parent material and the resulting different terrain, climate, geological substrates and thus soil types. Because of these features, it stands out from the region and in our case can be considered as an almost distinct area. In the northwest we have the Black Forest (BF), which also differs from the rest of the area in terms by the features mentioned, but at the same time has

similarities due to the likewise terrestrial geologic formation including sandstone. This great difference in such a small area naturally influences the vegetation and the processes in the soil, including soil formation. In total, the area comprises five major soil landscapes with different characterization, shown in Fig. 1B. These are areas in which, under similar geological, morphological and climatic conditions and under the influence of human, a landscape-typical association of soils has developed.

2.2. Data

Fig. 2A shows the soil types in the study area, with each number representing a soil type and its characterization. The soil types are determined according to the German soil classification system, which is based on the processes taking place in the soil and their properties (Eckelmann et al., 2005). In our area, there are 40 different soil types and the urban area, which is represented by the number 0. A detailed description can be seen in the Table 1, including the translation from the German into the World Reference Base (WRB) soil systematics (WRB, 2022).

In order to preserve the diversity that is lost in this translation, we will stick to the German classification. The soil type map used for our prediction variable was initially provided by Landesamt für Geologie, Rohstoffe und Bergbau (LGRB) Baden-Württemberg as a polygon map (Fig. 2A). We converted this polygon map to a raster file using a rasterization function based on the digital elevation grid. While the original scale of the map is 1:50,000, its rasterization allowed to produce a raster with pixels of 10 × 10 m. As covariates for the neural network, exemplified in Fig. 2B, we looked for spatially dense data over the whole region to get as detailed data as possible, which is also important for the performance of the neural network. For this purpose, we use a digital elevation model (Fig. 2B(a)), which was also provided by the LGRB with a resolution of 10 m, based on which topographic indices were calculated, also with a resolution of 10 m. The decision on which of the variables we use as covariates is based on expert geographical knowledge of the region, commonly used variables in the geosciences and by using the SCORPAN model introduced by McBratney et al. (2003), which is based on Jenny (1983). To cover most of the covariates presented in the SCORPAN model, we also included satellite data. Copernicus provides the Sentinel-2 data, available from 2017 in 13 spectral bands with a 5-day repetition frequency. For us, the most important variables are the visible (R, G, B) and near-infrared bands, which have a resolution of 10 m. We use these spectral bands to calculate important indices such as the Normalized Difference Vegetation Index (Fig. 2B(d)) to describe vegetation cover. Finally, we calculate the median value for each index over the time series from March to May 2019. In our analysis, we used the median as the mean over years to mitigate the influence of outliers and to ensure a more robust representation of the data. To capture the influence of geology, we add a geological map with the scale of 1:50,000, provided by the LGRB and rasterized in the same way as the soil type map. We provide an overview of all the covariates used for the ANN and the corresponding references in Table 2.

2.3. Model architecture

The origin of Artificial Neural Networks (ANNs) lies in the field of image recognition, especially in the area of classification (Goodfellow et al., 2016). These models are known for their ability to model multiple outcomes quickly and efficiently with a large amount of data, even with absence of prior knowledge about the data. Inspired by the neuronal structure, they look for dependencies and patterns in the given data that include input variables and a responding output variable. ANNS are organized in layers consisting of neurons using a (non-)linear activation function to transform and forward their inputs to the next layer, allowing the ANN to learn complex patterns. The input layer receives the input data and consists of one neuron per input feature, in our case, one neuron per covariate. The neurons in the hidden layers pass the weighted sum of the outputs from the previous layer to their activation function. The final layer outputs the prediction and consists of one neuron per output variable, in our case, one neuron per soil type. During training the weights of the connections between the layers are learned via stochastic gradient descent to minimize a loss function measuring the error of the predictions. There is a wide variability of different constructs for an ANN for computation or information processing in terms of the architecture of the neural network, the number, types and dimensions of layers, or the activation function chosen. Since the focus of our study is on uncertainty of machine learning models in a soil context rather than on model performance, the simplicity of the model was very important to us. We choose a fully connected multilayer perceptron as described in Table 3. As the activation function for the hidden layer, the rectified linear unit function was chosen, first used by Hahnloser et al. (2000) and defined as

$$\mathrm{ReLU}(x) = max(0, x)$$

with x as input to a neuron.

2.4. Confidence and uncertainty measurement of ANNs

We use the softmax function for the output layer to transform the previous layer's outputs into a vector of probabilities, essentially a probability distribution over the input classes. Mathematically, the Softmax function is defined as follows (Bridle, 1990; Goodfellow et al.,

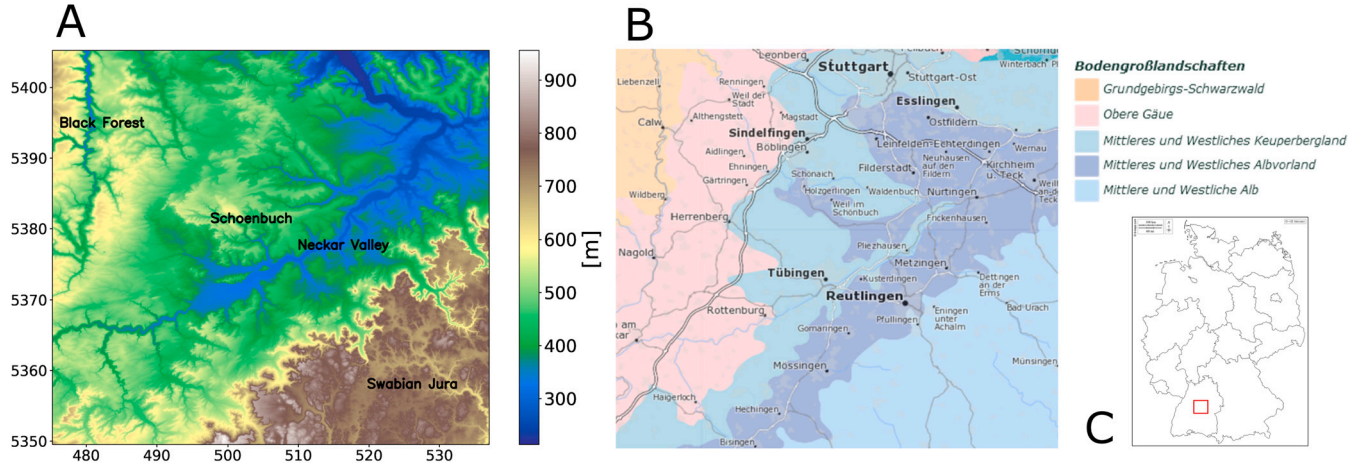


Fig. 1. (A) Digital elevation model of the study area with its important landscapes, (B) distribution of the soil families, (C) location of the study area in Germany.

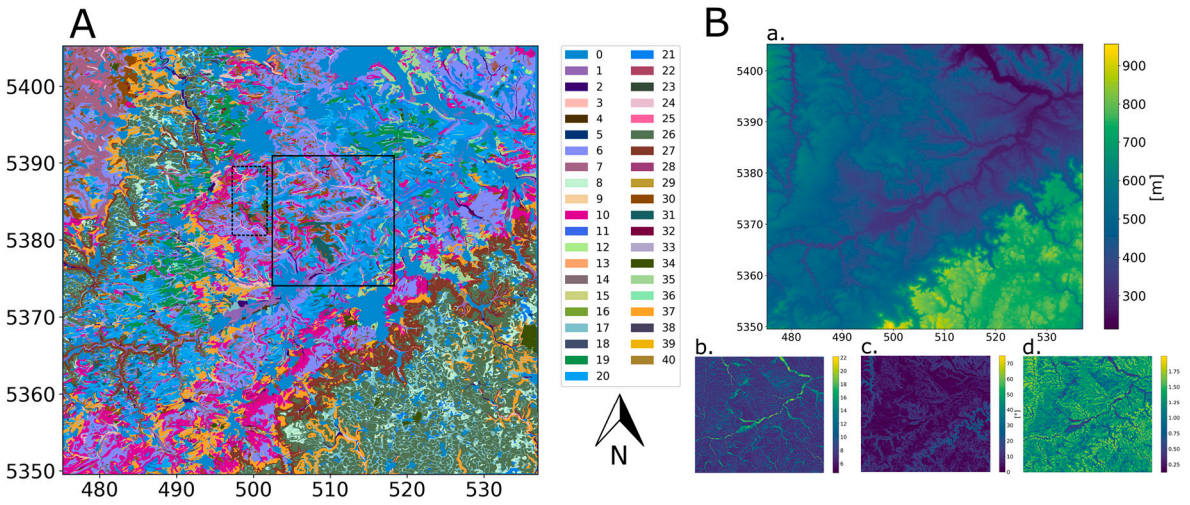


Fig. 2. (A) Soil type map over the study area, created by the LGRB with the train and validation area (solid rectangle) and the test area (dashed rectangle) (B) Examples of four used covariates over the study area: a. digital elevation model, b. soil moisture, c. slope and d. normalized difference vegetation index.

2016):

$$\text{Softmax}(x_i) = \frac{exp(x_i)}{\sum_j exp(x_i)}$$

where x is the vector of raw values for all classes. The class with the highest value is often used as the ANNs' prediction. However, we are not only interested in determining the soil classes for a given location and creating a soil type map, but also in evaluating the probability distribution of these classes over individual pixels. These probability values can be interpreted as a measure of confidence in the classification result. A higher maximum probability indicates that the predicted class is more likely to accurately represent the soil type at the given pixel position. In other words, the ANN has predicted this class with a low uncertainty and is therefore very confident about the prediction. A value of 1 demonstrates that there is a very high level of confidence in the predicted class for that particular pixel, suggesting that the model is almost certain about its prediction. On the other hand, a value of 0 reflects a very low level of confidence, indicating that the model shows a high uncertainty about the predicted class for that pixel.

2.5. Last-Layer Laplace Approximation

To analyse the uncertainty of ANNs, most studies only consider the confidence measurement from the previous section. However, this does not yet take into account the uncertainty of the model itself. As mentioned earlier, ANNs are known to be miscalibrated in terms of their uncertainty (Guo et al., 2017) or even tend to be overconfident in areas that are not well covered by training points or that are far away from them (Hein et al., 2019). To improve this overconfidence and at the same time obtain a spatially broad uncertainty expression, we use the Last-Layer Laplace Approximation (LLLA) by Kristiadi et al. (2020). This method is based on a probabilistic and Bayesian method by MacKay (1995), which calculate the a posteriori uncertainty for the neural network weights. More precisely, we approximate the posterior with the Laplace approximation by estimating the posterior with

$$p(\Theta|D) \approx \mathcal{N}(\Theta; , \Sigma) \qquad with \qquad \Sigma : \Theta_{MAP} = \left(\nabla^2_{\Theta} \mathscr{L}(\mathscr{D}; \Theta)\,|_{\Theta_{MAP}}\right)^{-1}$$

where Θ_{MAP} is the maximum a posteriori estimate of the last-layer parameters, obtained by minimizing the negative log posterior $\mathscr{L}(\mathscr{D}; \Theta)$, e. g. standard deep learning with a cross-entropy loss and an isotropic Gaussian prior. In other words, this amounts to training the network as usual to find Θ_{MAP}, then computing the Hessian of the training loss at this point. This approach is significantly cheaper than alternative methods based on sampling weights (both in terms of compute and memory cost). It also has the benefit that the point estimate (Θ_{MAP}) is unaffected by the uncertainty estimation, which simplifies development and tuning. Nevertheless, the computation of the Hessian adds a computational overhead. Computing the full Hessian is not feasible for large networks. But prior work by other authors (Kristiadi et al., 2020) has shown that limiting the Hessian to just the *last layer*, which is much less costly, already produces structured and useful uncertainty. Other approximations to curvature are also possible (Daxberger et al., 2021), but we limit ourselves to the Last-Layer Laplace approximation in this work.

2.6. Training and optimization process of the model

In contrast to most studies, our train and validation area is not arbitrarily chosen, but we deliberately select a specific area, that contains all soil types typical for the four soil landscapes, developed under terrestrial conditions, shown as a solid rectangle in Fig. 2A. The reason for the intentional selection is that we want to ensure that the soil types of the maritime soil landscape, i.e., those from the SJ, are not included in order to look at how the model relates to regions about which it does not receive information on soil types. In addition, we choose the centre of our map because we want to simulate a situation that often occurs, that some parts of a study area are well sampled due to previous individual projects.

The datasets contain the 30 covariates from Table 2 and the soil type labels. The SJ and its typical soil types are intentionally not present in the training and validation dataset to avoid providing information to the model. The test dataset (dashed rectangle in Fig. 2A) is similar in respect of the occurring soil types to the training and validation area to be able to determine the overall accuracy of the model. The detailed breakdown of soil types in the different data sets is shown in Table 4 and in Fig. 3. The training and validation set comprise 1,922,946 datapoints, while the test set contains 824,120 datapoints.

We tuned architectural and training hyperparameters using sequential model-based optimization on 1 % of the full dataset, to obtain results in a feasible time. Specifically, we used Bayesian optimization (Garnett, 2022) combined with Successive Halving to allocate resources to promising settings, as implemented in SMAC (Lindauer et al., 2022). We tuned a total of six hyperparameters comprising the number of neurons for each of the three layers (between 32 and 512 units), the initial learning rate (between $1e-4$ and 1.0), the learning rate scheduler (cosine annealing or exponentially decaying), and weight decay

Table 1
Detailed description of the soil types.

Classes number	Label	German soil classification	WRB-classification	Detailed information
0	None	None	None	Ablation, order, settlement
1	A1	Brauner Auenboden, Auenbraunerde	Fluvisol, Cambisol	Partly with gleying in the near subsoil, of alluvial sand and alluvial loam
2	A2	Auengley-Brauner Auenboden, Auengley-Auenbraunerde	Cambisol	From alluvial sand and alluvial clay
3	A3	Auengley, Auenpseudogley-Auengley, Brauner Auenboden-Auengley	Fluvisol	From alluvial sand and alluvial clay
4	A7	Auenbraunerde, Auenparabraunerde	Cambisol	From older alluvial sediment
5	B1	Terra fusca-Braunerde, Terra fusca-Parabraunerde, Reliktbraunerde	Leptosol, Cambisol	From solifluction soils over limestone and dolomite stone
6	B2	Braunerde, Pelosol-Braunerde, Pseudogley-Braunerde	Cambisol	From solifluction soils, partly alluvial and flood loam
7	B4	Braunerde, Podsol-Braunerde	Arenosol	Mostly podzolic, from sandstone, debris-rich fluvial soils and slope debris
8	CF1	Braunerde-Terra fusca, Terra fusca	Cambisol	From limestone and dolomite
9	CF2	Terra fusca, Braunerde-Terra fusca	Cambisol	From relocated river gravels
10	D1	Pelosol, Braunerde-Pelosol, Pseudogley-Pelosol	Luvisol	From solifluction soils, subordinate from alluvial debris
11	D2	Pelosol, Pseudogley-Pelosol	Luvisol	From flood and terrace sediments
12	D3	Pelosol, Braunerde, Parabraunerde, Nass- and Quellengley	Cambisol	From sliding masses
13	G1	Gley, Quellengley, Kolluvium-Gley	Gleysol	From solifluction soils and sedimentary formations, mostly alluvial deposits
14	G2	Pseudogley-Gley, Braunerde-Gley, Gley	Planosol	From flood loam, old water and alluvial sediment
15	G3	Anmoorgley, Nassgley, Humus- and Moorgley	Gleysol	From alluvial deposits, floodplain and flood sediments, and glacial deposits
16	H1	Niedermoor, Gley-Niedermoor, Hochmoor	Gleysol	From peat
17	K1	Kolluvium	Anthrosol	Partly over Braunerde and Parabraunerde, from alluvial deposits over solifluction soils
18	K2	Pseudogley-Kolluvium, Gley-Kolluvium	Planosol	From alluvial deposits
19	L1	Parabraunerde	Luvisol	From loess and sand loess
20	L2	Parabraunerde, Braunerde-Parabraunerde, Pseudogley-Parabraunerde	Luvisol	Of loess loam and loess-loam-rich solifluction soils
21	L3	Parabraunerde, Pelosol-Parabraunerde, Terra fusca-Parabraunerde, Pseudogley-Parabraunerde	Luvisol	From solifluction soils and slope debris
22	L5	Parabraunerde, Parabraunerde-Braunerde, Pseudogley-Parabraunerde	Luvisol	From terrace sediments, river and meltwater gravels
23	N1	Ranker und Braunerde-Ranker	Leptosol-Cambisol	From sandstone
24	P1	Podsol und Braunerde-Podsol	Podsol-Cambisol	Of sandstone, sandstone and flint rubble and solifluction soils
25	Q1	Regosol, partly (Locker)Syrosem	Regosol	Of slope debris, partly anthropogenically redeposited debris
26	R1	Rendzina	Leptosol	From limestone and dolomite, partly from slope or alluvial debris
27	R2	Rendzina und Pararendzina	Leptosol	From slope debris, partly from landslide debris
28	R3	Rendzina und Terra fusca-Rendzina	Leptosol	From river gravels
29	R4	Rendzina	Leptosol	From calcareous tuff and tertiary freshwater limestone
30	S1	Pseudogley, Braunerde-Pseudogley, Pelosol-Pseudogley	Planosol-Cambisol	From solifluction soils, partly Pleistocene alluvial debris
31	S2	Pseudogley, Parabraunerde-Pseudogley	Planosol-Luvisol	Of loess loam and loess-loam-rich solifluction soils
32	S3	Pseudogley, Kolluvium-Pseudogley	Planosol	From alluvial deposits
33	SS1	Stagnogley, Moorstagnogley	Gleysol	From solifluction soils, basin sediments and alluvial deposits
34	X1	Disturbed terrain		Original soils often heavily modified
35	Y1	Rigosol	Anthrosol	From solifluction soils, loess and various solid rocks
36	YY1	Deposit soil		From different substrates
37	Z1	Pararendzina, Pelosol-Pararendzina, Braunerde-Pararendzina	Leptosol-Vertisol	From solifluction soils and slope debris, partly from landslide masses
38	Z2	Pararendzina	Leptosol	Of loess and sandy loess, partly washed away or periglacially redeposited
39	Z4	Pararendzina	Leptosol	From flood deposits, alluvial debris, river and meltwater gravels
40	Z7	Pararendzina, Braunerde-Pararendzina	Leptosol-Cambisol	From volcanic weathering, partly covered by sedimentary material

Table 2
Overview of the covariates for the neural network.

Environmental input data		Definition after
Topographic variables	Altitude above channel network	Behrens et al. (2010), Conrad et al. (2015)
	Eastness	
	Northness	
	Catchment area	
	Convergence index	
	Crest index for lowlands	
	Crest index for mountain areas	
	Diffuse radiation	
	Direct radiation	
	Elevation	
	Elevation below culmination line for lowlands	
	Elevation below culmination line for mountain areas	
	Horizontal curvature	
	Mean slope	
	Plan curvature	
	Profile curvature	
	Projected distance to stream	
	Relative elevation	
	Relative hillslope position for lowlands	
	Relative hillslope position for mountain areas	
	Steepest slope	
	Terrain classification index for lowlands	
	Topography	
Spectral variables	Brightness index	Hounkpatin et al. (2018)
	Colouration index	
	Hue index	
	Normalized difference vegetation index	
	Redness index	
	Saturation index	
Geological variables	Geological map	LGRB Baden-Württemberg

Table 3
Architecture of the Artificial Neural Network.

Layer	Number of neurons	Activation function
Input layer	30	ReLU
Layer 1	395	ReLU
Layer 2	510	ReLU
Layer 3	489	ReLU
Output layer	41	Softmax

Table 4
Distribution of soil types in data subsets.

	Ground truth	Training and validation area	Test area
classes in the area concerned	0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40	0, 1, 2, 3, 4, 6, 7, 10, 13, 15, 17, 18, 19, 20, 21, 23, 25, 30, 31, 32, 35, 36, 37, 38	0, 2, 3, 6, 7, 10, 13, 15, 17, 18, 20, 21, 23, 30, 31, 32, 33, 34, 36
Total number of classes	41	24	19
classes that change compared to previous column	–	5, 8, 9, 11, 12, 14, 16, 22, 24, 26, 27, 28, 29, 33, 34, 39, 40	1, 4, 19, 25, 33, 34, 35, 37, 38

(between $1e-7$ and $1e-1$). We used 500 configurations as the optimization budget and trained each network using Adam (Kingma and Ba, 2015; Loshchilov and Hutter, 2018) with a fixed batch size of 1024 on 70 % of the training data and used the obtained accuracy on the 30 % validation set (obtained via a stratified split from the subsampled train dataset) as the optimization objective. For the remaining analysis, the best performing network configuration, shown in 3 with a weight decay of 0.008394464526246698 and a learning rate of 0.006203468269518196 was used.

After optimization, we made the decision to use the ANN with an early stop triggered when the ANN predicted the training accuracy, measured as the percentage of pixels where the model correctly predicts the outcome, to be above 90 % and the accuracy of the test data set did not increase significantly.

3. Results & discussion

3.1. Loss and accuracy of the ANN

It is important to note that our research's aim is not to outperform other state-of-the-art ANNs in digital soil mapping. Instead, our goal was to provide a realistic representation of the model's capabilities and limitations, especially when dealing with predictions in new areas with different soil types. We deliberately kept the model architecture simple and performed only a brief tuning process to simulate a real-life situation where quick predictions with a prepared ANN are required. At the same time, we focused on a specific scenario where we had two well-sampled areas: one for training the model and the other for testing. For the rest of the study area, we assume that we have no information on soil types. The developed model achieved a low loss of 0.07, which can indicate its effectiveness. The training accuracy of 95.11 % and the validation accuracy of 94.37 % also indicate that the model is capable of performing well within the specific training area, which only includes a deliberate selection of soil types. By limiting the training data to a geographically enclosed area with similar soil characteristics, we could assess how well the model generalizes to unseen similar regions. The test area, while geographically close to the training area, had some differences in soil types and their distribution, as shown in Fig. 3. Even with these variations, the model's test accuracy was 47.95 %, which was expected and still can be considered as good. Comparing our model to other ANNs used in digital soil mapping, particularly for predicting soil classes, it performed at an average level, which aligns with our expectations based on previous studies (Bagheri et al., 2015; Behrens et al., 2005; Bodaghabad et al., 2015; Boruvka and Penizek, 2006; Zhu, 2000).

3.2. Prediction of the ANN

3.2.1. Predicted soil type map

When we directly compare the map generated by the neural network prediction in Fig. 4 (Plot A) with the ground truth map derived from the LGRB in Fig. 2 (Plot A), several significant results emerge. Not all soil types occur in the prediction, which is expected, as certain soil types were not in the chosen training domain (see Table 4 column two). Consequently, the neural network was unable to predict these missing soil types because it had no knowledge of them. The decision to consider a specific area as our training data, where some soil types are missing, stems from the recognition that even within a relatively small geographical region, there can be significant variability and transitions between different soil types (Warrick, 2001). In such complex landscapes, it is entirely plausible that certain soil types may not have been sampled due to their close proximity or subtle variations that might have been overlooked during the sampling process (Heuvelink and Webster, 2001). By deliberately incorporating this aspect into our training dataset, this approach provides valuable insights into how the ANN responds to such common scenarios and assesses its ability to generalize and extrapolate predictions across the entire area, including regions with

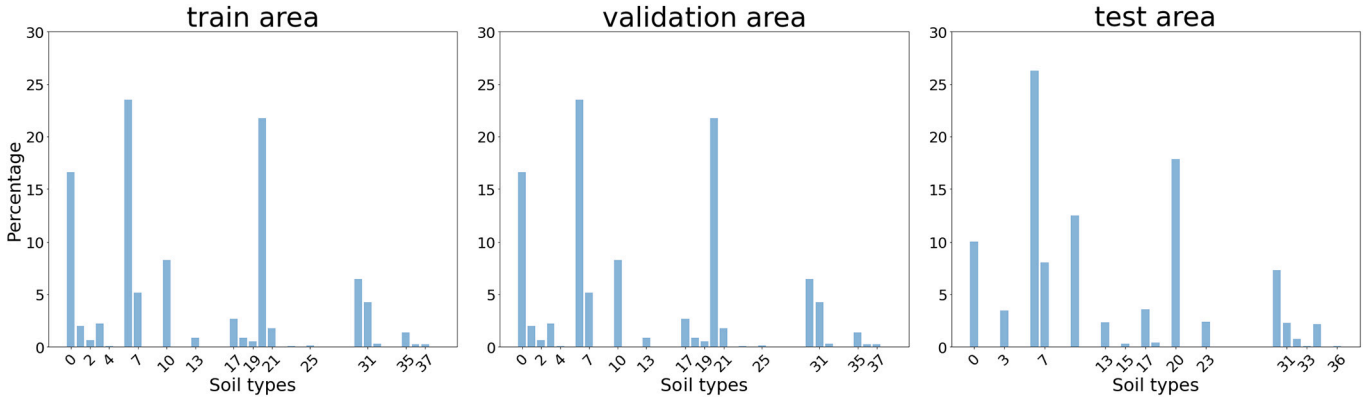


Fig. 3. Histograms of the distribution of soil types in the training, validation and test area.

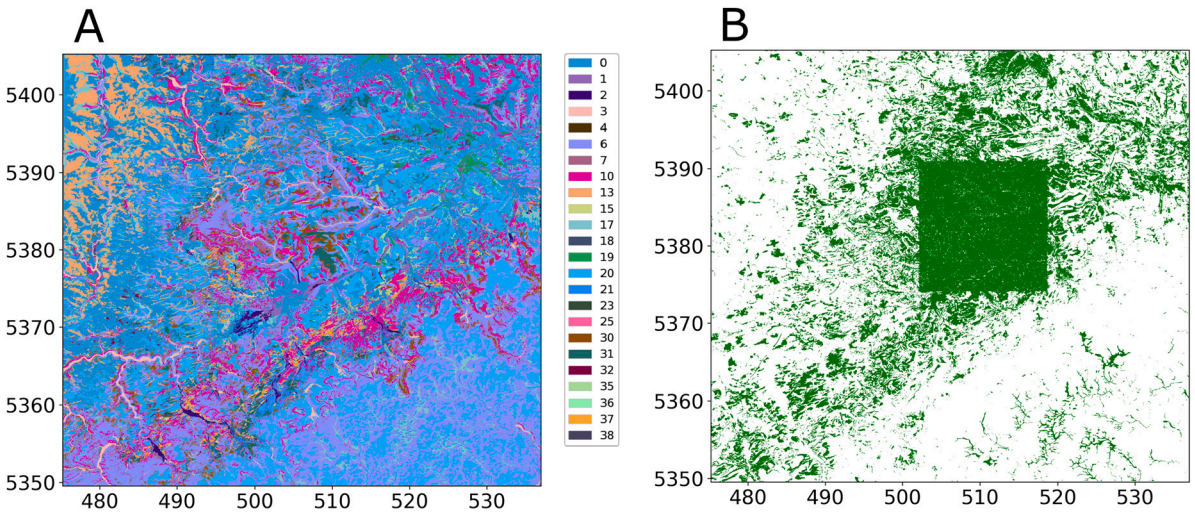


Fig. 4. (A) Prediction map of the soil types in our study area by the ANN, (B) Comparison of the prediction with the ground truth: green means correct prediction of the soil type.

missing soil types (Meyer and Pebesma, 2022). In our study the absent soil types include the soil family Terra fusca (class 5, 8, and 9), a specific soil type of the Rendzina family (class 29) and the Pararendzina family (class 39 and 40), which are typical soil types from SJ. The other part of the soil family Rendzina (classes 26, 27 and 28) is also typical for the SJ region and is also found at the border to the BF, specifically in the Oberen Gäue and along the upper course of the Neckar. In addition, some soil types are only found in specific areas. For instance, Pelosol (classes 11 and 12) and Gley (class 14) occur exclusively within the SJ region and the northern area of the Neckar valley. Conversely, classes 16, 22, 33, and 34 are sparsely distributed throughout the entire area. Lastly, the soil type Podsol (class 24) is typical for the BF. Because these soil types are only found in the specific regions SJ (southeast) and BF (northwest), it was anticipated that the SJ and BF areas have incorrect predictions for these soil types. The results in Fig. 4B illustrate this, where each green pixel represents a correctly predicted soil type.

The square in the centre of the Fig. 4B, with a high number of correctly predicted pixels, corresponds to the training area where the model performed very well. Additionally, when observing the diagonal of the image, which corresponds to areas similar to the training region, the predictions are also accurate. Notably, in the SJ area in the southeast, class 17 is the only one that was partially predicted correctly. Class 17 contains soils that are a result of human land use, including slope sediments from soil erosion (Colluvisol), soils that have been formed by mixing of natural material through human activity (Anthrosol) and alluvial deposits along creeks and rivers. All these azonal soils lack a close relation to natural soil forming factors as described by Jenny (1983) and used in the SCORPAN model by McBratney et al. (2003) except topography. They occur in any populated landscape, which is true for our full prediction area. This result demonstrates that the spatial pattern of the occurrence of azonal soils like Colluvisols and Anthrosols strongly relates to their topographic position (Penizek and Boruvka, 2008). This facilitates a transfer of the model results to other areas. In addition, it can be understood as an indicator for process proximity of a neural network, in that basic process-based rules of soil formation are recognized and reproduced by the model, as indicated before from several studies (Carter and Ciolkosz, 1991; Moore et al., 1993; Osat et al., 2016).

Consequently, as observed from the predicted map, it is evident that certain soil types occur more frequently and are more widely distributed, while others are less prevalent compared to the ground truth map. A detailed breakdown of the soil types classes from the soil truth map and their descriptions can be found in Table 1. The most substantial increase in predicted pixels compared to ground truth pixels, both in absolute numbers and relative proportions, was observed for classes 0, 6, 13, and 20. Additionally, in relative numbers, there was an increase in classes 3, 31, and 15. A more moderate increase was noted for classes 2, 10, 30, 32, and 36. On the other hand, a decrease was observed mainly in absolute numbers for classes 7, 17, 19, 21, and 37, and in relative proportions for classes 1 and 4. A minor decrease was recorded for classes 18, 35, 38, 25, and 23. Despite these decreases, it is essential to note that they are not as significant as the increases observed in some other classes.

The noticeable increase in the occurrence of classes 0, 6, and 20 can be attributed to the fact that our training data set is unbalanced. These particular classes are more frequently represented in the training set compared to other soil types. As a result, the neural network tends to predict them more often in the output (Johnson and Khoshgoftaar, 2019). During the prediction, not only are soil types that are absent in the training area assigned to certain regions, but also less frequently occurring soil types, as clearly shown in Fig. 5. The increase in the occurrence of certain soil classes in our ANN is not a random phenomenon. For instance, even though class 0 exists in the Swabian Jura region in the ground truth, it is not predicted there by our ANN. Instead, classes 6 and 20 dominate in that region. Both represent the two most commonly occurring soil types in Central Europe, making it reasonable for the ANN to classify areas about which it has no information (Amelung et al., 2018). The occurrence of Class 13 in the prediction is not accounted for by the frequency of it the training data. When we closely examine its distribution in the prediction, we find that this class is predicted for the high ridges around the BF. Previously, these areas were dominated by classes 6, 7, and 30, but now they are predicted as class 13, along with classes 17 and 0. Both Class 13 and Class 17 share similar characteristics to those in the ground truth soil types. For example, Gley soils can indeed occur in the BF as associated soils, although not as extensively as depicted in the prediction (Bleich et al., 1982). The overestimation of Class 17 in the region is not surprising, as Colluvium is a correlated sediment of soil erosion, and it can occur independently of climate and geology, primarily in depressions and valleys, just as shown in the prediction (Kopecky-Hermanns et al., 2022). Another significant change is observed in classes 1 to 4, which collectively represent the soils in the floodplains. The increase in class 3 is a result of soil types 1 and 2 being assigned to it. On the other hand, the decrease in class 4 is because this soil type is now primarily predicted as class 2. Despite these shifts, it is important to note that overall, the soil types in classes 1 to 4 have remained within their respective soil family. The prediction map highlights this phenomenon, particularly in the upper course of the Neckar and its tributaries, where class 3 is predicted, and in the Nagold valley, which is located in the BF. Meanwhile, the middle and lower Neckar valley are primarily predicted as class 1. These predictions support the notion that the floodplain soils retain their general characteristics, even with some changes in specific soil type assignments. The repetitive nature of river systems, with their well-defined channels and floodplain areas, provides a distinct and recognizable pattern that can be learned effectively by the neural network during training (Wiechmann, 2000).

3.3. Confidence and uncertainty of the ANN

Based on the previous results, especially the incorrect prediction of soil types in the SJ and BF regions, the next step is to analyse the confidence of the ANN in predicting the soil type. In the case of an ANN, apart from cross-validation methods and other techniques, a common step is to assess the probability of the predicted class (Wadoux et al., 2020). This probability can be interpreted as the model's confidence in its predictions. In Fig. 5A, this confidence value is plotted.

A value of 1 indicates high confidence in the corresponding pixel's predicted class, while a value of 0 indicates low confidence in the prediction. This information provides insights into the reliability of the model's predictions for each specific pixel and the corresponding predicted class. One striking observation is the presence of three areas where pixels with high confidence are concentrated. The central area corresponds to the training region of the ANN, and it aligns with previous results, indicating that over 90 % of the pixel predictions in this region were accurate. However, it is not the area with the highest overall confidence, as we would expected it. Instead, the regions of SJ in the southeast and BF in the northwest show almost uniform high confidence values approaching one, despite the ANN's poor performance in these areas. These regions are geographically and in terms of soil types distant from the training area and in addition the predicted classes often differ from the ground truth labels. As previously mentioned, ANNs tends to be overconfident in situations where it lacks sufficient training data, leading to inaccurate interpretations (Kasiviswanathan et al., 2018; Hein et al., 2019). Looking specifically at the pixels where the ANN correctly predicted soil types, we find that the mean confidence is 93 %. Interestingly, even for the wrong predictions, the mean confidence remains relatively high at 92 %. This indicates that the ANN assigns high confidence to both correct and incorrect predictions, further exacerbating the problem of overconfidence. Furthermore, in the surrounding regions of our training area, there are lower confidence values for the prediction of classes, particularly extending along the diagonal from southwest to northeast. This is also problematic because this area outside the training domain is where the prediction worked well previously.

Our examination of the confidence distribution of the probability of the predicted classes for each individual class reinforces a similar pattern, as visually depicted in Fig. 6. The confidence distribution is represented by the blue curve on the left-hand side of the axis at each class. The soil types shown here correspond to those in the training set, as only these could be taken into account by the ANN in its prediction and thus in the confidence distribution. Strikingly, for each class, there is a peak near the value of 1, indicating that most confidence values for

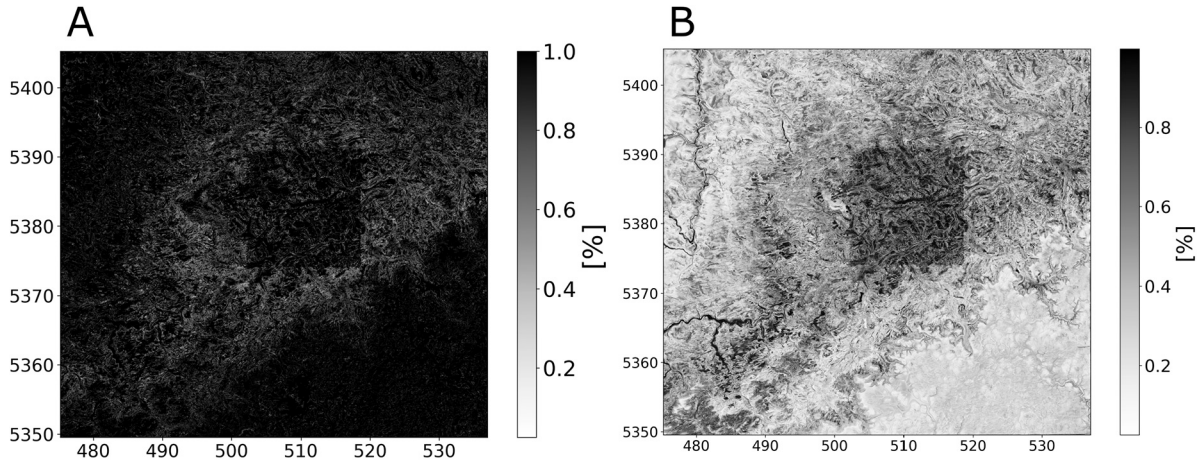


Fig. 5. (A) probability of the class in our study area predicted by the ANN calculated with the SoftMax function, interpreted as confidence of the ANN (B) probability after applying the LLLA, interpreted as uncertainty of the ANN.

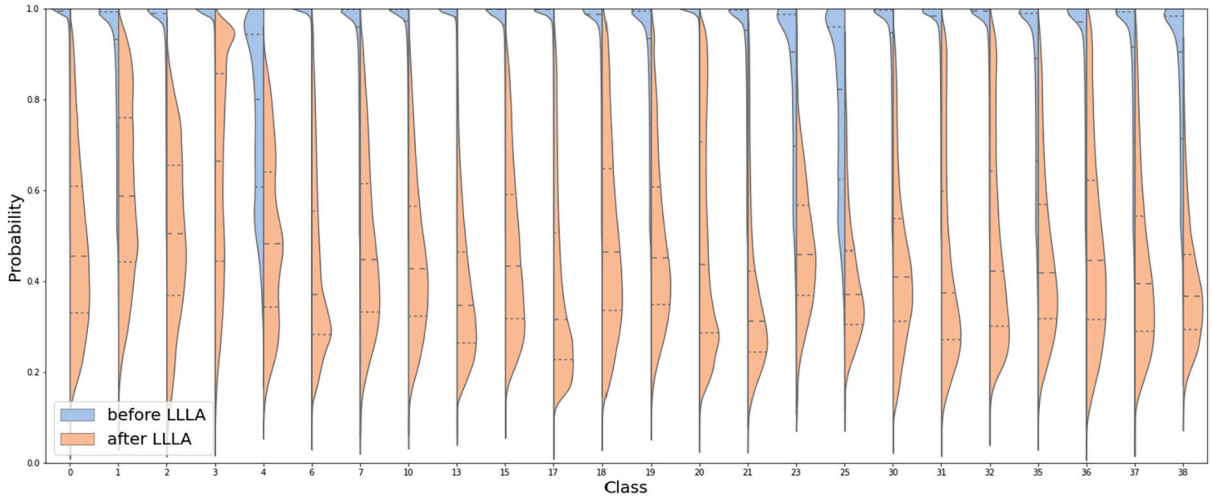


Fig. 6. Confidence and uncertainty distribution of the soil type classes individually, illustrated with the help of a kernel density estimate and additionally by quartiles of the distribution of each predicted class (blue shows the distribution before the application of LLLA, salmon afterwards).

these classes lying there. As well, as the confidence values decreases, the number of points lying around the lower values rapidly decreases, approaching zero. However, we noticed a distinct pattern for classes 4, 23, and 25, where a wider distribution of confidence values was observed. This can be attributed to the fact that these soil types appear infrequently in the predictions and are located in a diagonal region on the map, where we have a more heterogeneous representation of the model's confidence. Interestingly, classes 1, 19, 21, 35 to 38, which are also situated in this diagonal region, exhibit a weakly spread confidence distribution. Given the characteristics of the confidence distributions observed, it is evident that conducting a reliable analysis of the predictions made by the ANN is challenging. The presence of sharp peaks near 1 for most classes, suggests that the ANN is generally highly confident in its predictions. But as shown before, this high confidence may not always be well-calibrated or reliable, which is to be expected due to the absence of these soil types in the training area (Kramer and Leonard, 1990; Gal et al., 2016). These findings emphasize the need to address the issue of overconfidence in the ANN, particularly for distant regions from the training area, and to improve the model's confidence prediction with an uncertainty measurement in the surrounding regions where it previously showed correct predictions. These further investigations and adjustments are necessary to enhance the reliability and applicability of ANNs.

3.4. Uncertainty of the ANN due to Last-Layer Laplace Approximation

As described in Section 2.5, the application of the LLLA method has allowed us to obtain uncertainty statements for our model's predictions, effectively addressing the issue of overconfidence commonly associated with ANNs (Kristiadi et al., 2020). Following the post-processing of the model confidence with the LLLA method, we obtained Plot B in Fig. 5, where lighter shades represent higher uncertainty. In the plot, we observe distinctive patterns among different regions. The training area stands out prominently with low uncertainty, boasting the lowest overall uncertainty levels across the map. This aligns with the predictions in the training area, as it has the highest concentration of correctly predicted soil types. Conversely, the SJ region now exhibits the highest uncertainty in the plot, indicating a higher level of uncertainty for the predicted soil types in that area. This observation is consistent with our knowledge that this area could not be predicted accurately because its main soil types were not provided to the ANN during training (Gal et al., 2016). However, it is worth noting that the few correctly predicted pixels belonging to class 17 in the SJ area have the lowest uncertainty there. This is again consistent with the previously observed fact that the model correctly observed the soils and remains confident after applying LLLA with the correctly predicted class 17 representing the colluvisols, which also simplifies the application of the model by adding LLLA to different regions. Similar to the SJ area, we also observe the BF region with a relatively high uncertainty, except for the Nagold valley, where the uncertainty is notably lower. This indicates that the model's predictions in the BF region are generally less reliable, except for the specific area of the Nagold valley where the model exhibits higher confidence. In the southwest region, we encounter a more diverse picture with varying levels of uncertainty. As in the previous cases, there is surprisingly little uncertainty in the Neckar Valley. Upon closer examination, we find that the predicted soil type for both the Neckar valley and the Nagold valley corresponds to the previously described soil type number 3. Although this prediction was incorrect, the low uncertainty in these cases can be explained by the fact that the original soil types in both valleys belong to the same soil type family as class 3. Even though they are ultimately incorrect, this similarity in soil type family allows the model to be more certain in its predictions (Rossiter et al., 2017). The area in the diagonal of the map, which is the most similar to the training area, has now slightly higher uncertainty after the application of LLLA. Nevertheless, these uncertainty values are still higher than those observed in the SJ and BF regions. The slight increase in uncertainty after LLLA is acceptable, even if it is the most correct prediction after the training range. Considering the other results we have obtained, this increase in uncertainty is still within an acceptable range. Upon revisiting the mean uncertainty of all correctly predicted pixels, we find that it has dropped to 60.4 %, which is significantly less than before. If we instead examine the mean uncertainty of incorrectly predicted values, we observe a decrease of nearly 20 %, indicating that we are able to detect the uncertainty even on average.

When we delve into more detail by examining the corresponding part in Fig. 6, which is depicted in orange on the right side of the axis for each class, we gain a more diverse and informative understanding compared to the previous representation. The plots are wider, indicating a broader distribution of uncertainty values. Notably, Class 3 stands out distinctly, as it is the only class with a peak of uncertainty surpassing 90 %. This aligns with the previous knowledge that Class 3 is found in the valleys, precisely representing the locations where the uncertainty is lowest. Additionally, the only other wider plot at higher values belongs to Class 1, which belongs to the soil family of floodplains as Class 3, and exhibits similar behaviour. It has been further observed that class 2 and class 4, also belonging to this soil family, show a notable concentration of uncertainty values around the average, unlike all the rest of the classes. These clear patterns of uncertainty within the alluvial soil family

underline the ability of the LLLA uncertainty measurement to recognize subtle geographical and pedological features. This shows that the model not only recognizes that these soil types belong to the alluvial soil family, but also distinguishes them from other soil types with different characteristics.

The analysis of Class 0 reveals that its uncertainty values span a wide range from nearly 0 to 1, but there is a significant concentration at low values. Similar patterns are observed for other classes, specifically Class 6 and Class 20. This is a positive outcome compared to the previous situation, as these soil types were previously consistently overpredicted with high confidence at the SJ, but at the same time still well predicted in the other areas. The uncertainty statement thus covers both situations well for these classes. This shows that the uncertainty measurement can better classify the spatial variability of the soil classes and is not only based on the fact that these classes occur with the highest number of data points in the training data, which is important for a reliably prediction with an imbalanced dataset (Meyer and Pebesma, 2022). Another important observation is related to Classes 13 and 17, which now exhibit distributions concentrated at the lowest confidence values, indicating higher uncertainty for these classes. This increased uncertainty now captures the fact that these classes were completely overestimated in the BF region, leading to their relative increase in the number of pixels in the prediction range. In summary, the integration of the Bayesian method LLLA has allowed us to obtain more reliable uncertainty estimates for the model's predictions, which has proven to be in line with other Baysian methods for predicting soil variables (Poggio et al., 2016). The distinct patterns observed in the uncertainty values for different regions provide valuable insights into the model's performance and its reliability in various areas.

4. Implication on soil management

In the field of soil management, the integration of artificial neural networks (ANNs) is promising to improve precision and efficiency (Khaledian and Miller, 2020). However, a major challenge is to address the problem of overconfidence in ANNs, which can lead to inaccurate predictions and suboptimal decision-making. To combat this, we can leverage the Last-Layer Laplace Approximation method, which generates reliable uncertainty statements from ANNs. This method allows us to create uncertainty maps that play a crucial role in decision-making processes (McBratney, 1992; Heuvelink and Webster, 2023). Uncertainty maps allow soil mappers to target and prioritize areas to increase efficiency. By incorporating targeted sampling in regions with high uncertainty into the prediction workflow, we can quickly improve the quality of soil maps generated by ANNs (Stumpf et al., 2017; Richer-de et al., 2019). Thereby optimizing soil management practices by providing the dual benefit of broader spatial coverage and targeted resampling. In addition, exploring the potential transferability of a trained ANN to similar regions with comparable soil classes is of great interest (Mirzaeitalarposhti et al., 2022). When an ANN trained in one region is applied to another with similar soil characteristics, it is crucial to understand how well the model performs in the new context. A map indicating areas of greater uncertainty can directly show where the ANN may not function effectively. This visualization of uncertainty is crucial as it helps us understand the limitations and constraints of the model in practical applications. Such information provides essential insights for future investigations and improvements in both data collection and model development.

5. Conclusion

The primary objective of this research was to develop a reliable and straightforward method for quantifying uncertainty in Artificial Neural Networks (ANNs) used in digital soil mapping, particularly for soil type predictions in order to make the predictions more reliable and interpretable. This also includes the correction of overconfidence of ANNs, a tendency of ANNs to make overconfident predictions, especially in regions with limited data. This issue is particularly concerning in soil mapping because accurate and precise soil information is essential for effective soil management. To tackle the issue of insufficient uncertainty measurements for ANNs, we introduced a technique named Last-Layer Laplace Approximation (LLLA). The LLLA method is designed to produce more trustworthy uncertainty statements for the specified study area predicted by ANN. By incorporating LLLA into the ANN modelling process, the study demonstrates its effectiveness in showing of the areas, where the ANN erroneously has a high confidence and thereby preserves the low uncertainty in the correctly predicted areas. This improvement in uncertainty estimation ensures that the model's predictions are more reliable and trustworthy. This is crucial, because it helps users understand the level of confidence they can place in the model's predictions and therefore make decisions. Additionally, this research provides valuable insights into identifying knowledge gaps by analyzing areas with limited data support and high uncertainty. This information guides researchers in prioritizing data collection efforts in regions where the model's predictions are less reliable. Moving forward, we plan to apply the ANN with the LLLA approach to real soil samples and compare the results with maps generated by soil mappers. This comparison will further enhance our understanding of the underlying processes. In conclusion, the incorporation of Last-Layer Laplace Approximation into ANN modelling offers a promising solution to improve uncertainty estimation, making digital soil mapping predictions more reliable, interpretable, and actionable for effective soil management.

CRediT authorship contribution statement

Kerstin Rau: Writing – original draft, Visualization, Validation, Software, Methodology, Data curation, Conceptualization. **Katharina Eggensperger:** Writing – review & editing, Validation, Software. **Frank Schneider:** Software, Formal analysis. **Philipp Hennig:** Supervision, Funding acquisition. **Thomas Scholten:** Writing – review & editing, Supervision, Funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgements

Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy – EXC number 2064/1 – Project number 390727645 and the Tübingen AI Center (FKZ 01IS18039A). We also thank the Landesamt für Geologie, Rohstoffe und Bergbau Baden-Württemberg (LGRB), Freiburg, Germany, for providing soil data.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.scitotenv.2024.173720.

References

Adhikari, K., Minasny, B., Greve, M.B., Greve, M.H., 2014. Constructing a soil class map of Denmark based on the fao legend using digital techniques. Geoderma 214, 101–113.

Albrecht, C., Jahn, R., Huwe, B., 2005. Bodensystematik und bodenklassifikation teil i: Grundbegriffe. J. Plant Nutr. Soil Sci. 168 (1), 7–20.

Amelung, W., Blume, H.-P., Fleige, H., Horn, R., Kandeler, E., Kögel-Knabner, I., Kretzschmar, R., Stahr, K., Wilke, B.-M., 2018. Scheffer/Schachtschabel Lehrbuch der Bodenkunde. Springer-Verlag.

Bagheri, Bodaghabadi M., Martínez-Casasnovas, J., Salehi, M.H., Mohammadi, J., Esfandiarpoor, Borujeni I., Toomanian, N., Gandomkar, A., 2015. Digital soil mapping using artificial neural networks and terrain-related attributes. Pedosphere 25 (4), 580–591. https://doi.org/10.1016/S1002-0160(15)30038-2. https://www.sciencedirect.com/science/article/pii/S1002016015300382.

Behrens, T., Förster, H., Scholten, T., Steinrücken, U., Spies, E.-D., Goldschmitt, M., 2005. Digital soil mapping using artificial neural networks. J. Plant Nutr. Soil Sci. 168 (1), 21–33.

Behrens, T., Zhu, A.-X., Schmidt, K., Scholten, T., 2010. Multi-scale digital terrain analysis and feature selection for digital soil mapping. Geoderma 155 (3–4), 175–185.

Bleich, K., Hädrich, F., Hummel, P., Müller, S., Ortlam, D., Werner, J., 1982. Paläoböden in baden-württemberg. Geologisches Jahrbuch Reihe F, Bodenkunde 14, 61–99.

Bodaghabad, M.B., Martínez-Casasnovas, J., Salehi, M.H., Mohammadi, J., Borujeni, I.E., Toomanian, N., Gandomkar, A., 2015. Digital soil mapping using artificial neural networks and terrain-related attributes. Pedosphere 25 (4), 580–591.

Boruvka, L., Penizek, V., 2006. Chapter 30 a test of an artificial neural network allocation procedure using the czech soil survey of agricultural land data. In: Lagacherie, P., McBratney, A., Voltz, M. (Eds.), Digital Soil Mapping, Developments in Soil Science, vol. 31. Elsevier, pp. 415–424. https://doi.org/10.1016/S0166-2481(06)31030-6. https://www.sciencedirect.com/science/article/pii/S0166248106310306.

Breiman, L., 2001. Statistical modeling: the two cultures (with comments and a rejoinder by the author). Stat. Sci. 16 (3), 199–231.

Bridle, J.S., 1990. Probabilistic interpretation of feedforward classification network outputs, with relationships to statistical pattern recognition. In: Soulié, F.F., Hérault, J. (Eds.), Neurocomputing. Springer, Berlin Heidelberg, Berlin, Heidelberg, pp. 227–236.

Brungard, C.W., Boettinger, J.L., Duniway, M.C., Wills, S.A., Edwards Jr., T.C., 2015. Machine learning for predicting soil classes in three semi-arid landscapes. Geoderma 239, 68–83.

Carter, B.J., Ciolkosz, E.J., 1991. Slope gradient and aspect effects on soils developed from sandstone in Pennsylvania. Geoderma 49 (3), 199–213. https://doi.org/10.1016/0016-7061(91)90076-6. https://www.sciencedirect.com/science/article/pii/0016706191900766.

Conrad, O., Bechtel, B., Bock, M., Dietrich, H., Fischer, E., Gerlitz, L., Wehberg, J., Wichmann, V., Böhner, J., 2015. System for automated geoscientific analyses (saga) v. 2.1. 4. Geosci. Model Dev. 8 (7), 1991–2007.

Daxberger E., Kristiadi A., Immer A., Eschenhagen R., Bauer M., Hennig P., 2021. Laplace redux - effortless bayesian deep learning. CoRR abs/2106.14806. URL https://arxiv.org/abs/2106.14806, https://arxiv.org/abs/2106.14806https://arxiv.org/abs/2106.14806.

Dramsch, J.S., 2020. Chapter one - 70 years of machine learning in geoscience in review. In: Moseley, B., Krischer, L. (Eds.), Machine Learning in Geosciences, Advances in Geophysics, vol. 61. Elsevier, pp. 1–55. https://doi.org/10.1016/bs.agph.2020.08.002. https://www.sciencedirect.com/science/article/pii/S0065268720300054.

Eckelmann, W., Sponagel, H., Grottenthaler, W., 2005. Bodenkundliche Kartieranleitung.-5. verbesserte und erweiterte-Auflage. Schweizerbart'sche Verlagsbuchhandlung.

Gal, Y., et al., 2016. Uncertainty in Deep Learning.

Garnett, R., 2022. Bayesian Optimization. Cambridge University Press.

Gascon, F., Bouzinac, C., Thépaut, O., Jung, M., Francesconi, B., Louis, J., Lonjou, V., Lafrance, B., Massera, S., Gaudel-Vacaresse, A., et al., 2017. Copernicus sentinel-2a calibration and products validation status. Remote Sens. (Basel) 9 (6), 584.

Goodfellow, I., Bengio, Y., Courville, A., 2016. Deep Learning. MIT Press. http://www.deeplearningbook.org.

Grinand, C., Arrouays, D., Laroche, B., Martin, M.P., 2008. Extrapolating regional soil landscapes from an existing soil map: sampling intensity, validation procedures, and integration of spatial context. Geoderma 143 (1–2), 180–190.

Grunwald, S., Thompson, J., Boettinger, J., 2011. Digital soil mapping and modeling at continental scales: finding solutions for global issues. Soil Sci. Soc. Am. J. 75 (4), 1201–1213.

Guo, C., Pleiss, G., Sun, Y., Weinberger, K.Q., 2017. On Calibration of Modern Neural Networks. International conference on machine learning, PMLR, In, pp. 1321–1330.

Hahnloser, R.H., Sarpeshkar, R., Mahowald, M.A., Douglas, R.J., Seung, H.S., 2000. Digital selection and analogue amplification coexist in a cortex-inspired silicon circuit. nature 405 (6789), 947–951.

Hartemink, A.E., Bockheim, J., 2013. Soil genesis and classification. CATENA 104, 251–256. https://doi.org/10.1016/j.catena.2012.12.001. https://www.sciencedirect.com/science/article/pii/S0341816212002548.

Haykin, S., 1998. Neural Networks: A Comprehensive Foundation. Prentice Hall PTR.

Hein, M., Andriushchenko, M., Bitterwolf, J., 2019. Why Relu Networks Yield High-Confidence Predictions Far Away from the Training Data and how to Mitigate the Problem. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, In, pp. 41–50.

Hengl, T., Mendes de Jesus, J., Heuvelink, G.B., Ruiperez, Gonzalez M., Kilibarda, M., Blagotić, A., Shangguan, W., Wright, M.N., Geng, X., Bauer-Marschallinger, B., et al., 2017. Soilgrids250m: global gridded soil information based on machine learning. PloS One 12 (2), e0169,748.

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

Heuvelink G., Webster R., 2001. Modelling soil variation: past, present, and future. Geoderma 100(3):269–301. doi:https://doi.org/10.1016/S0016-7061(01)00025-8, URL https://www.sciencedirect.com/science/article/pii/S0016706101000258, developments and Trends in Soil Science.

Heuvelink, G., Webster, R., 2023. Uncertainty assessment of spatial soil information. Encyclopedia of Soils in the Environment 4, 671–683.

Hewitt, A., 1993. Predictive modelling in soil survey. Soils and Fertilizers 56 (3), 305–314.

Hounkpatin, K.O., Schmidt, K., Stumpf, F., Forkuor, G., Behrens, T., Scholten, T., Amelung, W., Welp, G., 2018. Predicting reference soil groups using legacy data: a data pruning and random forest approach for tropical environment (Dano catchment, Burkina Faso). Sci. Rep. 8 (1), 1–16.

Jenny, H., 1983. Factors of Soil Formation: A System of Quantitative Pedology. McGraw-Hill New York.

Johnson, J.M., Khoshgoftaar, T.M., 2019. Survey on deep learning with class imbalance. J. Big Data 6 (1), 1–54.

Kasiviswanathan, K., Sudheer, K., He, J., 2018. Probabilistic and ensemble simulation approaches for input uncertainty quantification of artificial neural network hydrological models. Hydrol. Sci. J. 63 (1), 101–113. https://doi.org/10.1080/02626667.2017.1393686. URL doi:10.1080/02626667.2017.1393686, https://arxiv.org/abs/doi:10.1080/02626667.2017.1393686https://arxiv.org/abs/doi:10.1080/02626667.2017.1393686.

Khaledian, Y., Miller, B.A., 2020. Selecting appropriate machine learning methods for digital soil mapping. App. Math. Model. 81, 401–418. https://doi.org/10.1016/j.apm.2019.12.016. https://www.sciencedirect.com/science/article/pii/S0307904X19307565.

Kingma, D., Ba, J., 2015. Adam: A method for stochastic optimization. In: Proceedings of the International Conference on Learning Representations (ICLR'15), pp. 1–15 published online: iclr.cc.

Kopecky-Hermanns, B., Vogt, R., Berg, S., 2022. Kolluvien. Geoarchäologie. Springer, In, pp. 207–216.

Kramer, M., Leonard, J., 1990. Diagnosis using backpropagation neural networks—analysis and criticism. Comput. Chem. Eng. 14 (12), 1323–1338. https://doi.org/10.1016/0098-1354(90)80015-4. https://www.sciencedirect.com/science/article/pii/0098135490800154.

Kristiadi A., Hein M., Hennig P., 2020. Being bayesian, even just a bit, fixes overconfidence in ReLU networks. In: III H. D., Singh A. (eds) Proceedings of the 37th International Conference on Machine Learning, Proceedings of Machine Learning Research, vol 119. PMLR, pp 5436–5446, URL https://proceedings.mlr.press/v119/kristiadi20a.html.

Lindauer, M., Eggensperger, K., Feurer, M., Biedenkapp, A., Deng, D., Benjamins, C., Ruhkopf, T., Sass, R., Hutter, F., 2022. SMAC3: a versatile bayesian optimization package for Hyperparameter optimization. J. Mach. Learn. Res. 23 (54), 1–9.

Loshchilov, I., Hutter, F., 2018. Fixing weight decay regularization in Adam. In: Proceedings of the International Conference on Learning Representations (ICLR'18), pp. 1–13 published online: iclr.cc.

MacKay, D.J., 1995. Probable networks and plausible predictions-a review of practical bayesian methods for supervised neural networks. Netw. Comput. Neural Syst. 6(3): 469.

McBratney, A.B., 1992. On variation, uncertainty and informatics in environmental soil management. Soil Research 30 (6), 913–935.

McBratney, A., Mendonça, Santos M., Minasny, B., 2003. On digital soil mapping. Geoderma 117, 3–52. https://doi.org/10.1016/S0016-7061(03)00223-4.

Meyer, H., Pebesma, E., 2022. Machine learning-based global maps of ecological variables and the challenge of assessing them. Nat. Commun. 13 (1), 2208.

Minasny B., McBratney A., 2016. Digital soil mapping: A brief history and some lessons. Geoderma 264:301–311. doi:https://doi.org/10.1016/j.geoderma.2015.07.017, URL https://www.sciencedirect.com/science/article/pii/S0016706115300276, soil mapping, classification, and modelling: history and future directions.

Mirzaeitalarposhti, R., Shafizadeh-Moghadam, H., Taghizadeh-Mehrjardi, R., Demyan, M.S., 2022. Digital soil texture mapping and spatial transferability of machine learning models using sentinel-1, sentinel-2, and terrain-derived covariates. Remote Sens. (Basel) 14 (23). https://doi.org/10.3390/rs14235909. https://www.mdpi.com/2072-4292/14/23/5909.

Moore, I.D., Gessler, P.E., Nielsen, G.A., Peterson, G.A., 1993. Soil attribute prediction using terrain analysis. Soil Sci. Soc. Am. J. 57 (2), 443–452. https://doi.org/10.2136/sssaj1993.03615995005700020026x.

Nguyen, A., Yosinski, J., Clune, J., 2015. Deep Neural Networks Are Easily Fooled: High Confidence Predictions for Unrecognizable Images. Proceedings of the IEEE conference on computer vision and pattern recognition, In, pp. 427–436.

Osat, M., Heidari, A., Karimian, Eghbal M., Mahmoodi, S., 2016. Impacts of topographic attributes on soil taxonomic classes and weathering indices in a hilly landscape in northern Iran. Geoderma 281, 90–101. https://doi.org/10.1016/j.geoderma.2016.06.020. https://www.sciencedirect.com/science/article/pii/S0016706116302683.

Penizek, V., Boruvka, L., 2008. The Digital Terrain Model as a Tool for Improved Delineation of Alluvial Soils. Springer, Netherlands, Dordrecht, pp. 319–326. https://doi.org/10.1007/978-1-4020-8592-5_28. URL doi:10.1007/978-1-4020-8592-5_28.

Poggio, L., Gimona, A., Spezia, L., Brewer, M.J., 2016. Bayesian spatial modelling of soil properties and their uncertainty: the example of soil organic matter in Scotland using r-inla. Geoderma 277, 69–82. https://doi.org/10.1016/j.geoderma.2016.04.026. https://www.sciencedirect.com/science/article/pii/S0016706116301860.

Rentschler, T., Bartelheim, M., Behrens, T., Behrens, T., Díaz-Zorita, Bonilla M., Teuber, S., Scholten, T., Schmidt, K., 2022. Contextual spatial modelling in the

horizontal and vertical domains. Sci. Rep. 12 (1), 9496. https://doi.org/10.1038/s41598-022-13514-5. https://europepmc.org/articles/PMC9184608.
Richer-de Forges A. C., Arrouays D., Bardy M., Bispo A., Lagacherie P., Laroche B., Lemercier B., Sauter J., Voltz M., 2019. Mapping of soils and land-related environmental attributes in France: analysis of end-users' needs. Sustainability 11 (10). doi:https://doi.org/10.3390/su11102940, URL https://www.mdpi.com/2071-1050/11/10/2940.
Rossiter, D.G., Zeng, R., Zhang, G.-L., 2017. Accounting for taxonomic distance in accuracy assessment of soil class predictions. Geoderma 292, 118–127. https://doi.org/10.1016/j.geoderma.2017.01.012. https://www.sciencedirect.com/science/article/pii/S0016706116303901.
Schmidhuber, J., 2015. Deep learning in neural networks: an overview. Neural Netw. 61, 85–117.
Scull, P., Franklin, J., Chadwick, O.A., McArthur, D., 2003. Predictive soil mapping: a review. Progress in Physical Geography: Earth and Environment 27 (2), 171–197.
Silveira, C.T., Oka-Fiori, C., Santos, L.J.C., Sirtoli, A.E., Silva, C.R., Botelho, M.F., 2013. Soil prediction using artificial neural networks and topographic attributes. Geoderma 195, 165–172.
Stumpf, F., Schmidt, K., Goebes, P., Behrens, T., Schönbrodt-Stitt, S., Wadoux, A., Xiang, W., Scholten, T., 2017. Uncertainty-guided sampling to improve digital soil maps. CATENA 153, 30–38. https://doi.org/10.1016/j.catena.2017.01.033. https://www.sciencedirect.com/science/article/pii/S0341816217300401.
Taghizadeh-Mehrjardi, R., Mahdianpari, M., Mohammadimanesh, F., Behrens, T., Toomanian, N., Scholten, T., Schmidt, K., 2020. Multi-task convolutional neural networks outperformed random forest for mapping soil particle size fractions in Central Iran. Geoderma 376 (114), 552. https://doi.org/10.1016/j.geoderma.2020.114552. https://www.sciencedirect.com/science/article/pii/S0016706119312777.
Taghizadeh-Mehrjardi, R., Emadi, M., Cherati, A., Heung, B., Mosavi, A., Scholten, T., 2021a. Bio-inspired hybridization of artificial neural networks: an application for mapping the spatial distribution of soil texture fractions. Remote Sens. (Basel) 13 (5). https://doi.org/10.3390/rs13051025. https://www.mdpi.com/2072-4292/13/5/1025.
Taghizadeh-Mehrjardi, R., Hamzehpour, N., Hassanzadeh, M., Heung, B., Ghebleh, Goydaragh M., Schmidt, K., Scholten, T., 2021b. Enhancing the accuracy of machine learning models using the super learner technique in digital soil mapping. Geoderma 399 (115), 108. https://doi.org/10.1016/j.geoderma.2021.115108. https://www.sciencedirect.com/science/article/pii/S0016706121001889.
Veres, M., Lacey, G., Taylor, G.W., 2015. Deep learning architectures for soil property prediction. In: 2015 12th Conference on Computer and Robot Vision, pp. 8–15.
Wadoux, A.M.-C., Minasny, B., McBratney, A.B., 2020. Machine learning for digital soil mapping: applications, challenges and suggested solutions. Earth Sci. Rev. 210 (103), 359.
Warrick, A.W, 2001. Soil Physics Companion. CRC press.
Wiechmann, H., 2000. Die bodensystematische kennzeichnung von auenböden. Böden und Hydrologie, Schadstoffe, Bewertungen, Stoffhaushalt von Auenökosystemen, pp. 19–25.
WRB I. W. G, 2022. World Reference Base for soil resources: International soil classification system for naming soils and creating legends for soil maps. In: International Union of Soil Sciences (IUSS), Vienna, Austria, 4th edn.
Zhang, W., Gu, X., Tang, L., Yin, Y., Liu, D., Zhang, Y., 2022. Application of machine learning, deep learning and optimization algorithms in geoengineering and geoscience: comprehensive review and future challenge. Gondw. Res. 109, 1–17. https://doi.org/10.1016/j.gr.2022.03.015.
Zhou, B., Zhang, X.-g., Wang, R.-c., 2004. Automated soil resources mapping based on decision tree and bayesian predictive modeling. Journal of Zhejiang University-SCIENCE A 5, 782–795.
Zhu, A.-X., 2000. Mapping soil landscape as spatial continua: the neural network approach. Water Resour. Res. 36 (3), 663–677.