

An innovative approach for integrating two-dimensional conversion of Vis-NIR spectra with the Swin Transformer model to leverage deep learning for predicting soil properties

Xiu Jin ^{a,b,1,*}, Jun Zhou ^{a,b,1}, Yuan Rao ^{a,b}, XiaoDan Zhang ^{a,b}, Wu Zhang ^{a,b}, WenJing Ba ^{a,b}, Xiaohu Zhou ^{a,b}, Tong Zhang ^{a,b}

^a College of Information and Computer Science, Anhui Agricultural University, Anhui, Hefei 230031, China

^b Key Laboratory of Agricultural Sensors, Ministry of Agriculture and Rural Affairs, Anhui, Hefei 230031, China

ARTICLE INFO

Handling Editor: Budiman Minasny

Keywords:

Deep learning
Soil spectroscopy
Gramian angular difference field
Spectrograms
Swin Transformer

ABSTRACT

The application of visual-near-infrared and shortwave-infrared (VNIR-SWIR) diffuse reflectance spectroscopy for soil properties analysis is increasingly gaining popularity due to its rapid, cost-effective, and non-destructive nature. In particular, deep learning models have been found to perform exceptionally well for large spectra libraries. This study proposes a novel approach to enhance the deep learning approach that involves converting one-dimensional spectra into two-dimensional (2D) spectral images. We investigated several methods, such as cutting reshape (CR), Gramian angular difference field (GADF), Gramian angle sum field (GASF), and Markov transition field (MTF). We then combine these converted images with the Swin Transformer model to predict a range of soil properties. Furthermore, we compare our proposed method with existing techniques reported in the literature. The results showed that the root mean square error (RMSE) of predicting soil organic carbon content (OC ($\text{g}\cdot\text{kg}^{-1}$)), nitrogen content (N ($\text{g}\cdot\text{kg}^{-1}$)), cation exchange capacity (CEC ($\text{cmol}^{+}\cdot\text{kg}^{-1}$)), pH, and sand (%) and clay content (%) using Gram's angular difference field (GADF) and Swin Transformer were 23.25, 1.26, 8.55, 0.54, 15.33, 6.14, and determination coefficients R² of 0.95, 0.92, 0.79, 0.90, 0.74 and 0.84, respectively. This study introduces a new perspective to enhance deep learning models for soil spectroscopy.

1. Introduction

Soil is an important natural resource that is vital for achieving sustainable development goals and supporting ecosystem services (Adhikari and Hartemink, 2016; Keesstra et al., 2016). However, soil is a finite resource and is under increasing pressure in the context of climate change and land degradation, with problems such as soil erosion and low replenishment of the agricultural production capacity (Lehmann et al., 2020; Rhodes, 2014). Thus, it is necessary to obtain timely information on soil conditions to help policy makers assess the impacts of land management practices on soil and adjust soil policies in a timely manner (Panagos et al., 2012).

Traditional methods of acquiring soil information primarily involve field sampling and laboratory testing, which often lead to delays in obtaining the necessary data. Moreover, these methods are costly in terms of labor and material resources, making them less commonly

employed in agricultural production (Nocita et al., 2015). Compared with traditional analytical techniques, infrared spectroscopy is a low-cost, non-destructive and rapid analytical technique that can be used to measure a wide range of soil properties in a single scan using a small number of soil samples in the laboratory; therefore, this approach is now widely used in soil characterization (Stenberg et al., 2010). Since the soil spectrum contains information about various soil components, a large soil spectrum library is being established at the global scale; this library contains physical and chemical properties and spectral characteristics measured with analytical techniques, and suitable models can be developed from this information for subsequent analyses based on spectral properties (Chang et al., 2001; Viscarra Rossel et al., 2016).

The Land Use/Cover Area frame Survey (LUCAS) soil spectral database developed by the European Union is considered the largest unified open-access dataset of topsoil properties in the world and is a continuously developing database; notably, the European Statistical Office

* Corresponding author at: College of Information and Computer Science, Anhui Agricultural University, Anhui, Hefei 230031, China.

E-mail address: jinxiu123@ahau.edu.cn (X. Jin).

¹ These two authors contributed equally to this work.

(EUROSTAT) provides a constantly updated version of this dataset to assess changes in land use (Orgiazzi et al., 2018; Panagos et al., 2012). It is therefore reasonable to infer that the LUCAS soil spectral library should provide a high degree of reliability for predictions of soil characteristic (Smith et al., 2020). However, spectral data often contain thousands of wavelengths with strong crosstalk and complex interrelationships (Beattie and Esmonde-White, 2021). In order to solve these problems, experts are constantly trying to use more accurate analysis techniques. Stevens et al. used the support vector machine regression (SVMR) method to predict organic carbon and showed that SVMR can achieve good prediction results (Stevens et al., 2013). However, most conventional prediction models have yield good predictions for only a relatively small number of soil types. Moreover, there is an increasing demand for the expansion of prediction models that can be applied to large-scale soil spectral data (Tian et al., 2020). Due to the diversity of soil sources sampled, an excessive sample size may increase the prediction error (Liu et al., 2018). To solve this problem, deep learning methods have been introduced for visual near-infrared spectral analyses of large-scale soil databases (Veres et al., 2015).

Padarian et al. began to explore the application of deep learning to soil spectra, successfully using convolutional neural networks (CNNs) to predict multiple soil components while avoiding dimensionality reduction and complex pre-processing steps (Padarian et al., 2019). Tsakiridis et al. developed localized multichannel 1-D CNNs for simultaneous prediction of multiple soil properties (Tsakiridis et al., 2020). Zhou et al. introduced an attention mechanism combined with CNN to construct an integrated regression network, which was successfully applied to estimate soil organic carbon content using hyperspectral data (Zhao et al., 2022). CNNs process data through local connectivity and parameter sharing, which in turn support the extraction of features from raw spectral data (Singh and Kasana, 2019). Usually, spectral data are considered sequential data (Zhang et al., 2020). The sequential nature of spectral data influences the accuracy of soil property prediction since the same feature bands often exist at different locations in the spectral data and may express different information (Williams and Zipser, 1989). However, CNNs tend to ignore location information during feature extraction, which can lead to poor model accuracy. Recurrent neural networks (RNNs) are commonly used to process sequential data because they can use feedback connections to preserve before-and-after history information (Ndikumana et al., 2018). Yang et al. proposed a joint CNN and RNN architecture that was compared with individual CNN and RNN models. The advantage of the joint architecture is that CNN is used to extract the original soil spectral features first, and then the RNN is used to explore the correlations among the features (Jin et al., 2018; Ng et al., 2019). Tsimpouris et al. used a stacked autoencoder to transform the initially recorded spectra in a new compressed space, helping the chemometric model to improve the accuracy of predictions (Tsimpouris et al., 2021). Tavakoli et al. used simple ratio index (SRI) and normalized difference index (NDI) to transform the visible-NIR data, combined

with a stacked model for calibration, to reduce the model computation (Tavakoli et al., 2023). However, these methods do not fully use the features of the sequences after extraction or in the model construction. It is difficult to parallelize the training of sequence data, and for long-sequence data, extracting features also requires human expertise, complicating the workflow and making it challenging to build end-to-end models. Capture long-range dependencies with good parallelism and scalability. It brings new ideas and breakthroughs to the field of deep learning (Khan et al., 2022). Vision Transformer (ViT) is a model that uses Transformer architecture to process image data. It brings new ideas and breakthroughs to the field of deep learning (Dosovitskiy et al., 2020). ViT divides an image into several blocks, each block is a sequence, and learns image features through Transformer's self-attentive mechanism. Swin Transformer is a new Transformer model that solves the information loss problem of traditional Transformer models by introducing a windowing mechanism to further improve model performance (Yang et al., 2022).

The image processing task, as one of the main tasks of deep learning, mostly involves two-dimensional data. In order to make more extensive use of algorithms related to deep learning, this paper attempts to convert a one-dimensional sequence into a two-dimensional image. Jin et al. used cutting reshape (CR) to convert 1D spectral data of wheat into 2D images and feed the images into a deep network for wheat disease identification; this approach provided slightly higher accuracy than did a 1D-CNN (Jin et al., 2018). Padarian et al. converted 1D data into 2D images by recombining the images with Hamming frames and inputting them into a CNN network for prediction (Padarian et al., 2019); although the method is simple, the prediction was not satisfactory. Wartini et al. compared unprocessed soil spectral data with reconstructed two-dimensional images as inputs for a CNN and then compared the results, concluding that the prediction accuracy of the former was better than that of the latter (Ng et al., 2019). We want to further explore the use of a two-dimensional approach for predicting soil properties.

Therefore, it is necessary to explore how other transformation methods can be used to generate images while preserving the features of 1D serial data. Wang et al. proposed two methods, namely, the Gramian angular field (GAF) and Markov transition field (MTF) methods, for the two-dimensional visualization of time series and used deep networks to process time series data (Wang and Oates, 2015). Despite the differences between spectral wavelength points and time series data, both data types have similar continuity properties. In this paper, to better predict soil properties from spectral data, 1D spectral data are transformed into 2D spectral images, and the optimal deep model for this process is then identified. Therefore, the objectives of this paper are as follows: 1. Converting 1D spectral data into 2D spectral images and comparing the effects of different conversion methods, 2. Predicting multiple soil properties from spectral images using the Swin Transformer model, and 3. Comparing proposed method with other current techniques.

The remainder of this paper is arranged as follows. Section 2

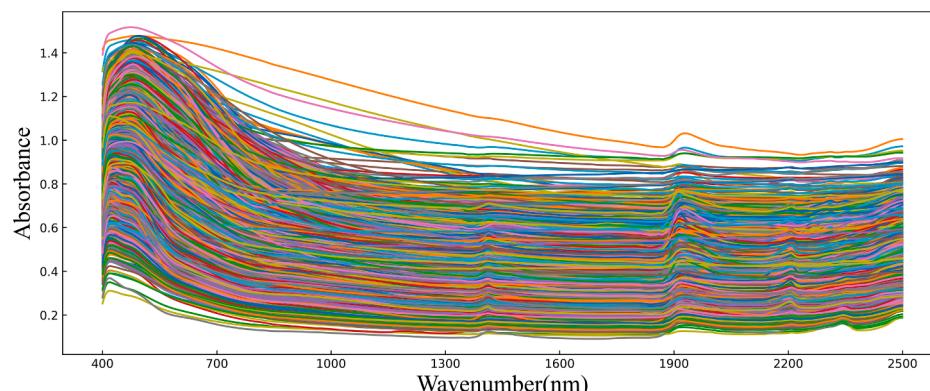
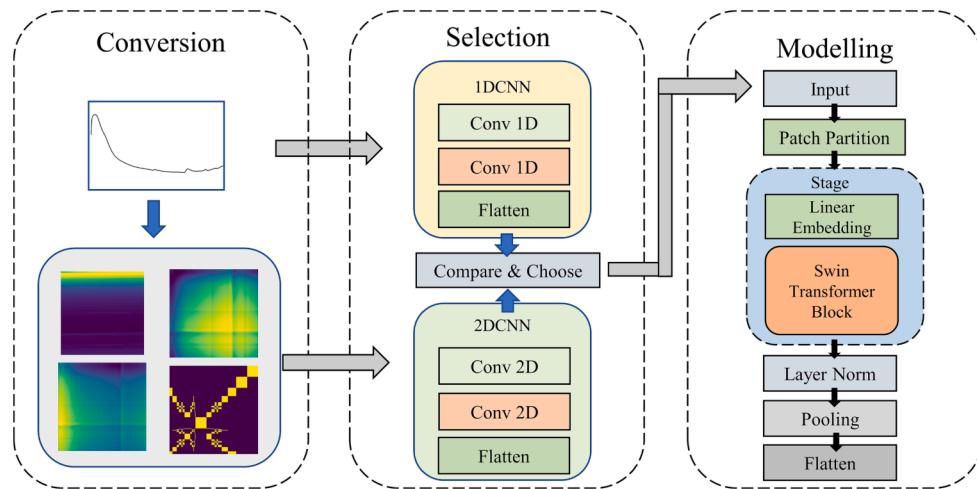


Fig. 1. VNIR-SWIR spectra of the LUCAS soil samples.

Table 1

Summary statistics for soil samples from the LUCAS 2009 topsoil database used; Q1 and Q3 are the first and third quartiles, std represents standard deviation and skew represents skewness.

Property	Total	Min	Q1	Median	Q3	Max	Mean	Std	Skew
OC ($\text{g}\cdot\text{kg}^{-1}$)	19,036	0	12.70	20.80	39.70	586.80	50.00	91.31	3.68
N ($\text{g}\cdot\text{kg}^{-1}$)	19,036	0	1.20	1.70	2.90	38.60	2.92	3.76	3.76
CEC ($\text{cmol}^+\cdot\text{kg}^{-1}$)	19,036	0	7.00	12.40	20.30	234.00	15.76	14.48	4.24
pH	19,036	3.21	5.02	6.21	7.47	10.08	6.20	14.48	-0.08
Sand (%)	17,939	1.00	19.00	42.00	64.00	99.00	42.88	26.10	0.19
Clay (%)	17,939	0	8.00	17.00	26.50	79.00	18.88	13.00	0.91

**Fig. 2.** Experimental flow chart.

describes the LUCAS 2009 topsoil database. Then, an overview of the modelling frameworks used, a list of methods used to encode 1D spectra into 2D spectral images, the deep model and the relevant evaluation methods are described. Section 3 presents the results of the proposed method, an analysis of the results and a comparison of the proposed approach and different state-of-the-art methods. The conclusions of this work are given in Section 4.

2. Materials and methods

2.1. The LUCAS dataset

The LUCAS soil spectral dataset was used in this study. The data (from the 2009 survey) contains 19,036 topsoil samples from 23 EU member states (Panagos et al., 2012). The LUCAS dataset is one of the largest and most diverse soil spectral databases, covering a wide range of land use types, land cover types, and geographic sources. All soil samples were standardized and chemically analysed. The spectral data were then recorded using a diffuse reflectance spectrometer (XDS™ Rapid Content Analyzer) with a spectral interval of 400–2500 nm and a spectral resolution of 0.5 nm, with 4200 data points obtained for each sample (Orgiazzi et al., 2018).

The objective of this study was to predict multiple soil characteristics, the focus was only on these characteristics. Six soil characteristics were investigated as target predictor variables in this paper: clay particle size fraction (Clay, %), sand particle size fraction (sand, %), pH in H_2O (pH), organic carbon content (OC, $\text{g}\cdot\text{kg}^{-1}$), total nitrogen content (N, $\text{g}\cdot\text{kg}^{-1}$), and cation exchange capacity (CEC, $\text{cmol}^+\cdot\text{kg}^{-1}$). These characteristics were chosen because they are important for soil fertility and stability. We considered all available soil samples in the dataset, including mineral and organic soil samples, and do not differentiate between sample classes; since sand and clay are not present in the organic samples (about 1000 samples), we do not consider this part of the samples in predicting sand and clay. In Fig. 1, the VNIR-SWIR

spectra of the LUCAS soil samples are shown. A summary of the statistical information on the physicochemical properties of the soil samples we used in the LUCAS 2009 topsoil database is shown in Table 1.

2.2. Methods

The overall experiment is described below and illustrated in Fig. 2. After the image transformation process, and after a simple CNN comparison to select the optimal transformation method, the next step is to construct a depth model using Swin Transformer.

2.2.1. Two-dimensional conversion and convolutional neural networks

To exploit the advantages of deep learning in the field of NIR spectroscopy and explore inputs with different dimensions suitable for processing with deep learning algorithms, it is necessary to study methods for converting one-dimensional spectral sequences into two-dimensional images and then combine them with deep learning algorithms for subsequent data analysis. For assessing one-dimensional spectral sequences, Jin et al. proposed the cutting reshape (CR) method, which can be used to reorganize one-dimensional spectral data into two-dimensional images while preserving the integrity of the data (Jin et al., 2018). For processing time series data, Wang et al. proposed the Gramian angular field (GAF) and Markov transition field (MTF) algorithms, which can encode 1D time series into 2D images without feature loss (Wang and Oates, 2015).

2.2.1.1. Cutting reshape. Cutting and recombining spectral data according to wavelength is a relatively simple technique in spectral image processing. In this process, each spectral data is first normalised to a value range in the (-1, 1) range to ensure consistency in the subsequent processing. Next, in order to better reconstruct the image, we need to cut the wavelengths according to the appropriate window size. The window size here should be similar to the number of cuts made to ensure that the resulting image is easier to process subsequently. We use a window size

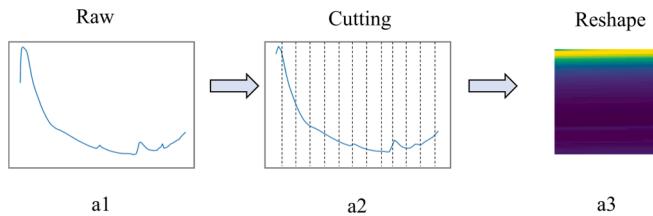


Fig. 3. Reconstruction of NIR spectral sequences into 2D images: NIR spectral sequence (a1), The result after cutting (a2), and the image generated after data reorganization (a3).

of 65, the total wavelength of the data is 4200, and there are 65 bands after the cut. The undersized bands are filled with zeros and reconstructed into a 65*65 two-dimensional matrix, and the conversion process is shown in Fig. 3.

2.2.1.2. Gramian angle field. In time series processing, the GAF algorithm and MTF algorithm proposed by Wang et al. can be used to encode a one-dimensional time series into a two-dimensional image without feature loss. The GAF approach can be applied to encode data into an image by using a matrix based on polar coordinates, thus maintaining the location information of the data (Ndikumana et al., 2018). First, the data are aggregated by using a piecewise aggregation approximation (PAA) (Guo et al., 2010). To reduce the sequence length, for sequence $Q = \{q_1, q_2, \dots, q_m\}$, another sequence \bar{Q} is established, where $\bar{Q} = \{\bar{q}_1, \bar{q}_2, \dots, \bar{q}_w\}$ and $w < m$. Setting $k = \frac{m}{w}$, the element \bar{q}_i in \bar{Q} can be expressed as:

$$\bar{q}_i = \frac{1}{k} \sum_{j=k(i-1)+1}^{ki} q_j, 1 \leq i \leq w \quad (1)$$

Then, \bar{q}_i can be normalized to the range of 0 to 1:

$$\tilde{\bar{q}}_i = \frac{(\bar{q}_i - \bar{q}_{min})}{\bar{q}_{max} - \bar{q}_{min}} \quad (2)$$

\bar{q}_i is the result of PAA and $\tilde{\bar{q}}_i$ denotes the normalized data. Polar coordinates are used for $\tilde{\bar{q}}_i$:

$$\begin{cases} \phi_i = \arccos(\tilde{\bar{q}}_i), -1 \leq \tilde{\bar{q}}_i \leq 1 \\ r = \frac{i}{n}, i \in n \end{cases} \quad (3)$$

where ϕ_i is the angle value and r is the radius. Formula (3) indicates that the conversion angle ϕ_i is in the range of $[0, \pi]$, and the cosine value monotonically decreases within this range. As the wavelength increases in the spectral interval, \bar{q}_i in each right-angle coordinate system corresponds to a value in the polar coordinate system, and corresponding bending occurs between different angular points on the polar circle. By calculating the cosine of the sum of the angles between two different points with the following formula, we can obtain the Gramian angle sum field (GASF):

$$\text{GASF} = [\cos(\phi_i + \phi_j)] \quad (4)$$

Similarly, by calculating the sines of the angle differences between different points, we can obtain the formula for the Gramian angle difference field (GADF):

$$\text{GADF} = [\sin(\phi_i - \phi_j)] \quad (5)$$

The image position corresponds to the wavelength and is symmetrical along the main diagonal. The conversion process is shown in Fig. 4.

2.2.1.3. Markov angular field. In the Markov angular field (MTF) approach, the Markov transfer probability is used to save position information for a sequence (Wang and Oates, 2015). For a given sequence $X = \{x_1, x_2, \dots, x_n\}$, the value range is divided into Q intervals, and then each x_i is mapped to the corresponding q_i . Therefore, we can obtain a $Q \times Q$ matrix W . Element w_{ij} in the matrix represents the probability that

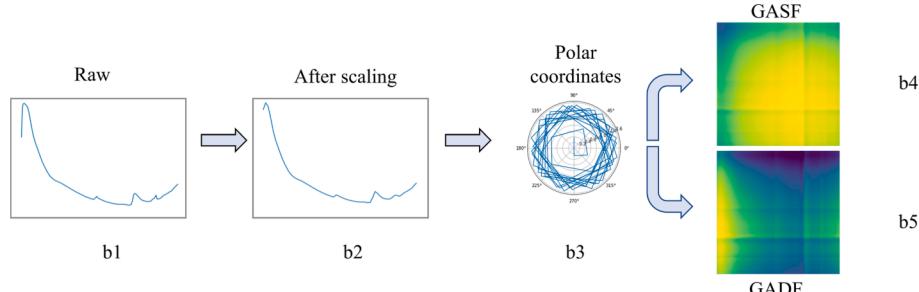


Fig. 4. NIR spectral sequence converted to GAF 2D image, original NIR spectral sequence (b1), processed by PAA (b2), polar transformed (b3) to obtain GASF image (b4) and GADF image (b5).

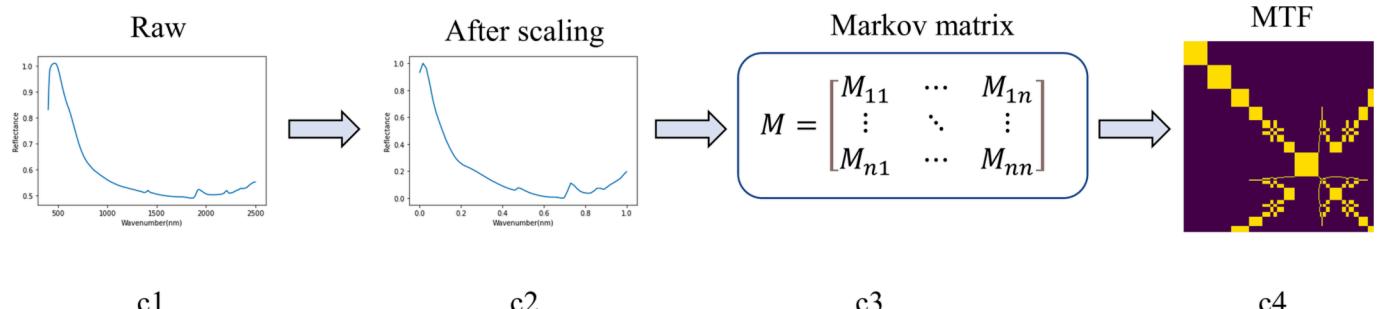


Fig. 5. Conversion of NIR spectral sequences into MTF 2D images, (c1) Near-infrared spectral sequences, after PAA treatment (c2) to the (c3) Markov matrix to generate the MTF image of the soil (c4).

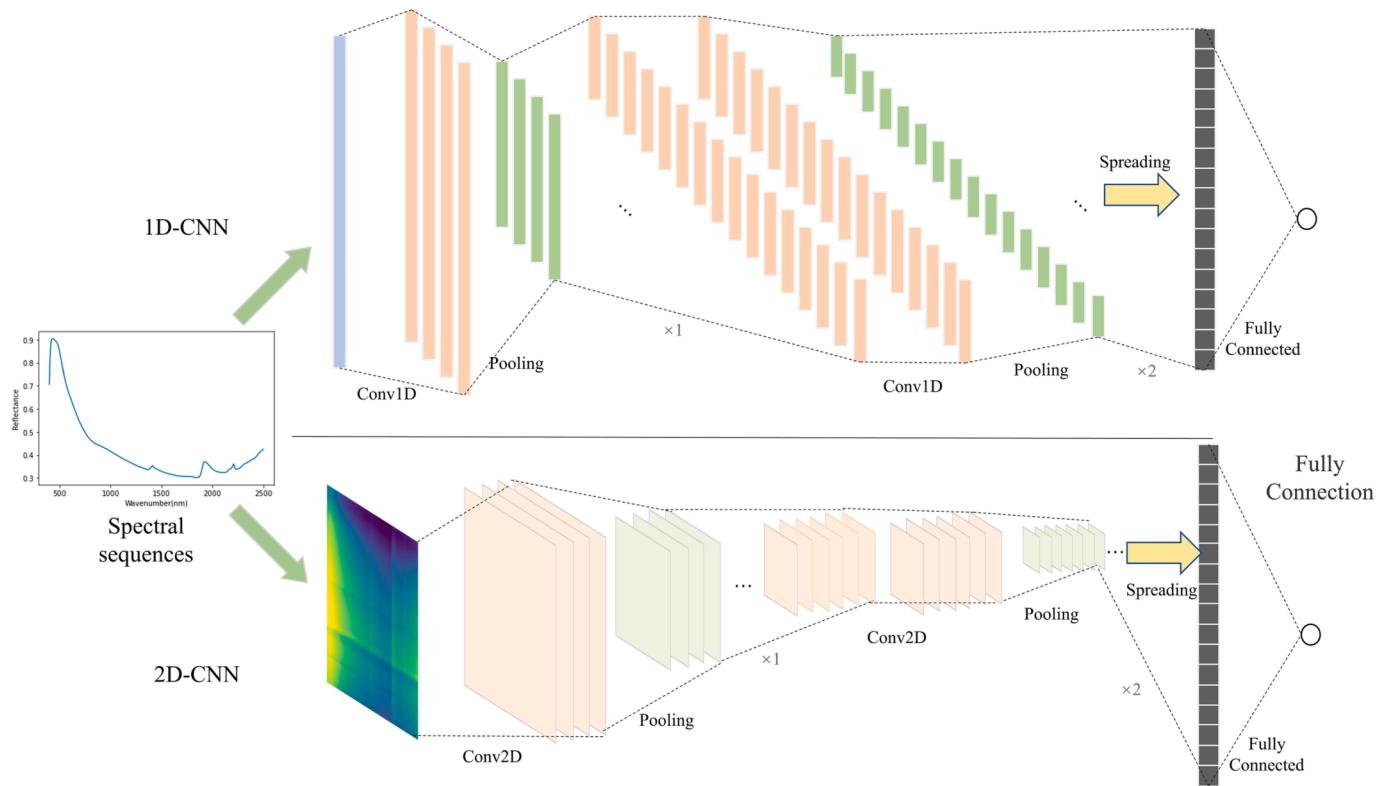


Fig. 6. Convolutional neural networks with two different dimensions. The top structure is a one-dimensional convolutional neural network (1D-CNN) for spectral sequence processing, and the bottom structure is a two-dimensional convolutional neural network (2D-CNN) for spectral image processing.

the element in interval q_j is followed by q_i in the same interval and satisfies the condition $\sum_{j=1}^Q w_{ij} = 1$. W is regarded as a Markov transfer matrix. However, its dependence on the distribution and sequence X is not strong, and the matrix w experiences extensive information loss for the original time series. The cell M_{ij} in matrix M is shown in formula (6).

$$M_{ij} = \begin{bmatrix} w_{ij} | x_1 \in q_i, x_1 \in q_j & \cdots & w_{ij} | x_1 \in q_i, x_n \in q_j \\ \vdots & \ddots & \vdots \\ w_{ij} | x_n \in q_i, x_1 \in q_j & \cdots & w_{ij} | x_n \in q_i, x_n \in q_j \end{bmatrix} \quad (6)$$

To merge the time information into m , M_{ij} is used to denote a series from q_i to q_j . The matrix M is represented as shown in formula (7). The conversion process is shown in Fig. 5.

$$M = \begin{bmatrix} M_{11} & \cdots & M_{1n} \\ \vdots & \ddots & \vdots \\ M_{n1} & \cdots & M_{nn} \end{bmatrix} \quad (7)$$

In the process of encoding spectral sequence data into different types of images, CR does not require a complex mathematical transformation, but both the GAF and MTF methods require certain mathematical transformations. The GAF can be recovered for the original sequence based on the principle of polar transformation, and the MTF is established based on the probability of element movement; therefore, it cannot be recovered to the original sequence.

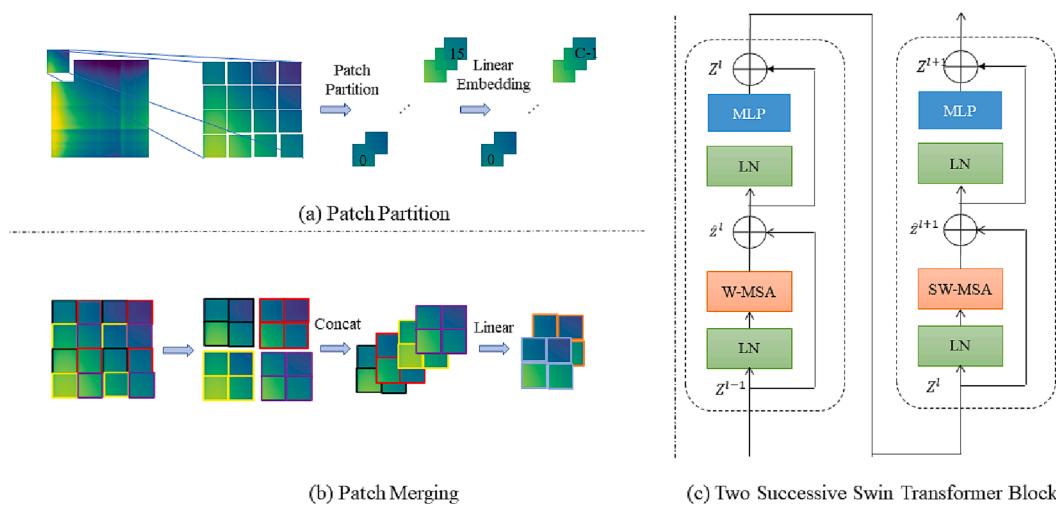


Fig. 7. Swin Transformer model, which consists of three main parts: the patch partitioning layer (a), the patch merging layer (b) and the Swin Transformer block (c).

2.2.2. Convolution and Swin Transformer networks

We created two similar convolutional network architectures for two different inputs (1D and 2D). The networks were used to process spectral data in both dimensions. The frameworks of these two convolutional neural networks are shown in Fig. 6, with different convolution and pooling operations for data with different dimensions. The fully connected layer is used to get the final output. The hyperparameters used for both networks are: the optimiser used was Adam, with a learning rate of 0.001; the loss function was mean square error loss (MSELoss); the training batch size was 128 and the training epoch was 100.

We attempted to extract information from spectral images using a transformer approach. Since the spectral images generated from spectral sequence data contain important positional information, the Swin Transformer was used to obtain global features (Han et al., 2021). The structure of the Swin Transformer consists of three main parts, as shown in Fig. 7: the patch partitioning layer, the patch merging layer and the Swin Transformer block. The patch partitioning layer is responsible for dividing the spectral map. Then, the results are distributed through channels, and down-sampling is performed in the patch merging layer. Next, the attention mechanism for each patch is computed with the Swin Transformer block. The result is passed to a norm layer, a global pooling layer and a fully connected layer to obtain the final output. The three components of the Swin Transformer are described below.

For a 2D spectral image, the data format is a two-dimensional matrix, which is not ideal for the Swin Transformer. Therefore, the data need to be transformed in the embedding layer. As shown in Fig. 7a, an image is first divided into sub-images of a given size and then flattened in the channel direction. Thus, after patch partitioning, the image shape is reduced by a factor of 16, and there are 16 sub-images. The linear embedding layer then performs a linear transformation of the channel data for each pixel. Next, multiple stages are passed, each consisting of a Linear Embedding with multiple Swin Transformer blocks. In addition to that with the linear embedding layer in the first stage, down-sampling with path merging layers is performed again in the remaining three stages.

In each stage, down-sampling is performed with the patch merging layer (except for the first stage). As shown in Fig. 7b, the input to the patch merging layer is a 4*4 feature map, and each adjacent 2*2 pixel is divided into a patch. Blocks of pixels at the same position in each patch are then stitched together to obtain four feature maps. The four feature maps are then stitched together in the depth direction and then input into the normalization layer. Finally, the depth direction of the feature map is linearly adjusted in the fully connected layer, and the depth, height, and width of the feature map are reduced by half.

Each Swin Transformer block consists of a multi-head self-attention (MSA) mechanism and a multilayer perceptron (MLP) (Voita et al., 2019). In addition, the normalization layer (LN) is applied before the MSA block and the MLP, and a residual connection is applied after each block. The Swin Transformer uses window-based multi-head self-attention (W-MSA) and shifted-window multi-head self-attention (SW-MSA) mechanisms to improve efficiency, as shown in Fig. 7c. In the W-MSA mechanism, the input feature map is divided into windows, each containing 7*7 patches. The W-MSA mechanism is only used for self-attention calculations within the windows, thus simplifying the computational effort; however, information transfer between windows is limited. To solve this problem, the SW-MSA (W-MSA with shifting) mechanism is introduced, and the windows are shifted so that information is exchanged between them. Moreover, the two structures are used in pairs. First, the W-MSA structure and then the SW-MSA structure are used. Finally, the LN, global pooling layer and fully connected layer are connected to obtain the final output.

2.3. Evaluation

The following indicators are used to verify the performance of the regression prediction model for the validation set: the 1) coefficient of

Table 2

Configuration of two convolutional neural networks.

1D-CNN	2D-CNN
Input (1*4200)	Input (1*65*65)
Conv1d 3-64	Conv2d 3*3-64
Maxpooling 2	Maxpooling 2*2
Conv1d 3-128	Conv2d 3*3-128
Maxpooling 2	Maxpooling 2*2
Conv1d 3-256	Conv2d 3*3-256
Conv1d 3-256	Conv2d 3*3-256
Maxpooling 2	Maxpooling 2*2
Conv1d 3-512	Conv2d 3*3-512
Conv1d 3-512	Conv2d 3*3-512
Maxpooling 2	Maxpooling 2*2
Conv1d 3-512	Conv2d 3*3-512
Conv1d 3-512	Conv2d 3*3-512
Maxpooling 2	Maxpooling 2*2
FC Dense	FC Dense

determination R², 2) root mean square error (RMSE), 3) ratio of performance to interquartile distance (RPIQ). The RMSE is a measure that represents the standard deviation of the model prediction error. A low RMSE usually indicates good performance. The RMSE is derived from the following formula:

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{N}} \quad (8)$$

R² is the percentage of change in the dependent variable explained by the regression model and varies between 0 and 1; a value close to 1 indicates strong performance. The R² formula is as follows:

$$R^2 = 1 - \frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{\sum_{i=1}^N (y_i - \bar{y}_i)^2} \quad (9)$$

RPIQ is used to measure the deviation between predicted and observed values. Where IQR is the interquartile range of the observed values and RMSE is the root mean square error between the predicted and observed values. The calculation formula of RPIQ is:

$$RPIQ = \frac{IQR}{RMSE} \quad (10)$$

All deep learning models are implemented based on PyTorch (1.12.1). An NVIDIA Tesla P100 PCIe 16 GB GPU is used to accelerate the training process and perform the experiments.

3. Results and discussion

3.1. Analysis of the 1D- and 2D-CNN models

The original soil NIR spectral sequence was used to generate different 2D images, of which the best conversion method was chosen. We first converted the spectral data from the LUCAS dataset into two-dimensional images based on four methods and then input the spectral sequences and the obtained images into a 1D-CNN and a 2D-CNN for training. The performance of the models during the training process was assessed, and the results were compared. With the CNN structure and hyperparameters fixed, only the input data affected the prediction results. The 1D spectral sequences in the LUCAS dataset were converted to 2D images in four ways, and then the 1D spectral sequences and the generated images were input into the 1D-CNN and 2D-CNN, respectively.

We divided the dataset, including the spectra with the selected soil properties, into two random groups, with 70% of the data used for training and 30% for independent testing. The training set is then divided into five parts using a five-fold cross-validation method, which allows the trained model to have better generalization performance, and

Table 3

Testing set results for six soil properties (OC, N, CEC, pH, Sand, and Clay) based on different convolutional models and different data forms.

Soil properties	Raw_1DCNN		CR_2DCNN		GASF_2DCNN		MTF_2DCNN		GADF_2DCNN	
	RMSE	R ²	RMSE	R ²	RMSE	R ²	RMSE	R ²	RMSE	R ²
OC	26.34	0.91	24.20	0.93	24.89	0.92	27.42	0.86	24.18	0.93
N	2.28	0.81	2.09	0.87	1.94	0.86	2.62	0.76	1.82	0.89
CEC	10.39	0.56	9.53	0.64	9.52	0.70	13.35	0.51	8.71	0.78
pH	0.88	0.76	0.64	0.85	0.63	0.87	0.78	0.71	0.60	0.88
Sand	19.52	0.60	18.95	0.62	17.86	0.61	21.96	0.43	18.41	0.63
Clay	8.56	0.72	7.49	0.79	7.85	0.77	12.19	0.62	7.24	0.80

the hyperparameters of the model are fixed at training time.

By comparing the differences between the soil NIR spectral sequences and the generated 2D images, the best conversion method was then selected. We first converted the spectral data from the LUCAS dataset into different 2D images according to four methods; the original spectral sequences were used as input to a 1D CNN to train the model and tune the parameters; and the obtained images were used as input to a 2D CNN using the same network structure and hyperparameters as the 1D CNN. With the CNN structure and hyperparameters fixed, the input data will be used as a factor to influence the prediction results. Table 2 shows the network framework we used.

Supplementary Table 1 and Supplementary Table 2 show the average loss values for the training and validation of the model over 100 training cycles and provide a comparison of the five different inputs for the six tasks. The loss values of each deep neural network decrease during training and stabilize after 100 training epochs. For the six soil property prediction tasks, rapid convergence was observed after twenty epochs in five cases (all except for sand, for which the loss values declined more slowly during training). During training, the 2D-CNN model with the GADF method yielded the lowest and most stable loss values for training and validation and thus better performance than the other methods. In contrast, the 1D-CNN model with a 1D spectrum and the 2D-CNN model with the MTF method displayed slightly worse performance.

The prediction results are shown in Table 3 and indicate that the 2D spectral images obtained after CR processing and the 1D spectral sequence data can be used to effectively train the model; however, the 1D spectral sequence is less effective on the test set. Because only the data dimension is changed in CR processing, it can be assumed that the 2D spectral images are advantageous for predicting soil properties. In the case of the conversion to 2D spectrograms, both the GAF and MTF methods were used to increase the dimensions of the original data and achieve data enhancement.

The results in Table 3 show that the worst results were obtained with the MTF method, especially for the prediction of Sand (RMSE = 21.96, R² = 0.43), and this prediction was significantly different from the prediction of CR for Sand (RMSE = 18.95, R² = 0.62). Thus, the use of the location information preserved in the MTF method did not enhance the results. By contrast, the two-dimensional spectral images obtained using the GADF and GASF treatments in the GAF method both enhanced the prediction of soil properties. The results obtained with the GADF method corresponded to the lowest RMSE and the highest coefficient of determination, and the results obtained with the GASF method were slightly worse than those obtained with the GADF approach. Still, both methods yielded good OC predictions (RMSE = 24.18, R² = 0.93). The predictions for the six soil properties using the GASF approach were worse than those using the GADF method, with differences in the RMSE between 0 and 0.81 and differences in R² between 0.01 and 0.08. Moreover, the GADF approach notably improved CEC prediction (RMSE = 8.71, R² = 0.78) compared to the other methods. This finding suggests that using the spatial location correlations preserved in the GADF method can enhance information extraction with deep neural networks.

Table 4

R² values with different transformer module stages + Swin Transformer block structures for six prediction tasks.

Stage + Block structure	OC	N	CEC	pH	Sand	Clay
Stage *1(Block *1)	0.93	0.91	0.75	0.85	0.67	0.77
Stage *1(Block *2)	0.93	0.92	0.76	0.87	0.69	0.79
Stage *1(Block *3)	0.92	0.91	0.73	0.87	0.68	0.79
Stage *1(Block *4)	0.92	0.91	0.73	0.86	0.68	0.79
Stage *1(Block *2) + Stage *2 (Block *2)	0.94	0.93	0.77	0.88	0.72	0.80
Stage *1(Block *2) + Stage*2 (Block *2) + Stage*3(Block *2)	0.94	0.93	0.78	0.89	0.73	0.82
Stage *1(Block *2) + Stage *2 (Block *2) + Stage *3(Block *2) + Stage *4(Block *2)	0.95	0.94	0.79	0.90	0.74	0.84

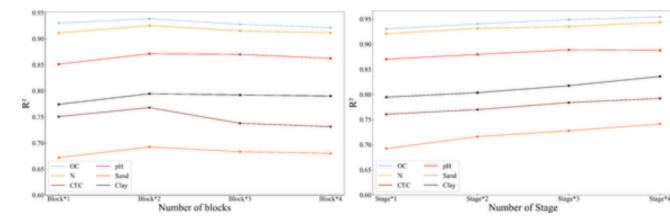


Fig. 8. On the left is the performance of the model when the number of blocks is adjusted, and on the right is the performance of the model when the number of blocks is fixed at 2 and the number of Stages is adjusted.

3.2. Training and evaluating the Swin Transformer model with the GADF method

Based on the results of the first experiment, we propose the GADF-Swin Transformer approach. Images generated using the GADF method are used as inputs to the Swin Transformer model to investigate its performance when processing spectral images that retain positional correlations.

The Swin Transformer model in this study includes a hierarchical construction method similar to that used in CNNs. To investigate the effects of the number of layers in the Swin transformer and the number of Swin Transformer blocks on the performance of the model, one, two, three and four layers of Swin Transformer blocks were used in a given transformer module design to train the model while other parameters were unchanged. The predictions of six soil properties were used to guide model construction, and the prediction performance of the model is shown in Table 4. It is notable that the results of the model do not improve as the number of Swin Transformer blocks increases, with the final predictions of the model being approximately the same at a block size of 2, as shown in the left panel of Fig. 8. Based on these results, we fixed the number of blocks to 2 and adjusted the number of Stages in the model, the results are shown in the right panel of Fig. 8. The accuracy of the model increases as the number of layers increases, and the model performs optimally when 4 layers are used. Considering the model performance, a Swin Transformer model with 4 layers in the transformer module stage and 2 Swin Transformer blocks is used for training and to

Table 5

The best Swin Transformer model configuration.

Down-sampling Rate (Output size)		GADF-Swin Transformer
Stage 1	4 (16 × 16)	concat 4 × 4, 96-d, LN $\left[\begin{smallmatrix} \text{win.sz.7} \times 7, \\ \text{dim96}, \text{head3} \end{smallmatrix} \right] \times 2$
Stage 2	8 (8 × 8)	concat 2 × 2, 192-d, LN $\left[\begin{smallmatrix} \text{win.sz.7} \times 7, \\ \text{dim192}, \text{head6} \end{smallmatrix} \right] \times 2$
Stage 3	16 (4 × 4)	concat 2 × 2, 384-d, LN $\left[\begin{smallmatrix} \text{win.sz.7} \times 7, \\ \text{dim384}, \text{head12} \end{smallmatrix} \right] \times 2$
Stage 4	32 (2 × 2)	concat 2 × 2, 768-d, LN $\left[\begin{smallmatrix} \text{win.sz.7} \times 7, \\ \text{dim768}, \text{head24} \end{smallmatrix} \right] \times 2$
Normalization layer		
AdaptiveAvgPooling		
FC dense		

Table 6 R^2 results for predictions of the GRDF-Swin Transformer and other models.

Soil properties	GADF-ResNet	GADF-MobileNet	GADF-EfficientNet	GADF-ViT	GADF-Swin Transformer
OC	0.80	0.79	0.93	0.93	0.95
N	0.82	0.75	0.85	0.78	0.94
CEC	0.40	0.61	0.71	0.67	0.79
pH	0.61	0.73	0.86	0.83	0.90
Sand	0.41	0.43	0.68	0.48	0.74
Clay	0.71	0.66	0.76	0.71	0.84

learn the spectral images for subsequent soil property detection.

Stage: Transformer module stage; Block: Swin Transformer block. Except for the first module, which consists of linear embedding and the Swin Transformer block, the subsequent modules involve patch merging and the Swin Transformer block.

In our experiments, the loss function was based on the mean squared error, the stochastic gradient descent (SGD) optimizer and GELU activation function were used, and the batch size was 128. The best model configuration for different swin transformer structures is shown in Table 5. After an image was down-sampled and the channel number was adjusted, it was passed to the LN and two stacked Swin Transformer blocks, where `win.sz` indicates the size of the window used, `dim` indicates the channel depth of the feature map, and `head` indicates the number of heads in the multi-head attention module.

`win.sz` indicates the size of the window used, `dim` indicates the channel depth of the feature map, and `head` indicates the number of heads in the multi-head attention module.

The loss of the model is observed for 100 epochs during training, as shown in Supplementary Fig. 8. Notably, the losses from the training and validation of each model converge rapidly as the number of iterations increases. In subsequent iterations, the training and validation loss curves for OC, N, CEC, and pH predictions largely overlap, while the training and validation loss curves for the Clay and Sand predictions fluctuate slightly. The loss of each model decreases as the number of training epoch increases, verifying the superior stability and performance of the model used.

Regression predictions for six soil properties (OC, N, CEC, pH, Sand, and Clay) were obtained using the GADF-Swin Transformer model. The performance results of the model were assessed based on the R^2 , RMSE and RPIQ values obtained. In the Table 6, the transformer performed well in OC, N and pH prediction ($R^2 = 0.90\text{--}0.95$); additionally, the Clay predictions were moderately accurate ($R^2 = 0.84$), and the CEC and Sand predictions were poor ($R^2 = 0.78$, $R^2 = 0.73$). Fig. 9 shows scatter plots of the measured and predicted values for the six soil properties based on the transformer model, and the distributions of measured and predicted values are clearly shown.

3.3. Comparing the soil property predictions of different methods

To demonstrate that our model is suitable for processing images generated from spectral data, images generated by the GADF method were used as inputs to train other image processing models, and the results were compared. We have selected several representative algorithmic models: Resnet50 (He et al., 2016), a deep convolutional network; MobileNet (Howard et al., 2017), a lightweight convolutional network; EfficientNet (Tan and Le, 2019), an efficient convolutional network that introduces reinforcement learning; and Visual Transformer (ViT) (Dosovitskiy et al., 2020), which unifies natural language processing and image processing. We used the hyperparameters of the above model in the image classification dataset CIFAR-10 task and changed the output of the model to accomplish the task of soil property prediction. In Fig. 10, we graphically show the prediction performance results R^2 for the soil property predictions. The results show that in this experiment, the Swin Transformer with images generated from spectral data is superior to the other models and can be used to process images and predict soil properties.

To further assess the performance of different modelling approaches, the results of the GADF-Swin Transformer model proposed in this paper were compared with those of the multi-task convolutional neural networks (multi-CNNs) used by Padarian et al. (Padarian et al., 2019), the long short-term memory (LSTM) used by Singara et al. (Singh and Kasana, 2019), and the combines CNN and RNN (CCNVR) used by Yang et al. (Yang et al., 2020). The complete set of results is presented in Table 7. Due to the correlation between different soil properties, related studies have shown that neural networks can improve the performance of models when correlations are considered in the m-model, which in turn can be used to predict multiple soil properties simultaneously (Padarian et al., 2019; Tsakiridis et al., 2020). In our study, on the other hand, we focused only on single property output methods.

When comparing our results with those of previous studies on the LUCAS dataset (as shown in Table 7), we should consider that, as in our study, some work used both organic and mineral soils from the dataset (Padarian et al., 2019; Singh and Kasana, 2019), while some other work has used only mineral soils (Tsakiridis et al., 2020; Yang et al., 2020). In addition, the data sets divided in the work of predecessors are also different, and different preprocessing has been applied in previous work (2D spectrograms in Padarian et al. (2019), and multiple spectral preprocessing in Tsakiridis et al., 2020). Our approach treats organic and mineral soils as a whole dataset to improve the generalizability of the model and does not perform feature extraction on soil spectra, focusing only on the generalizability of the proposed method.

1D- and 2D-CNN: one- and two-dimensional convolutional neural networks; LSTM: long short-term memory network; RMSE: root mean squared error; R^2 : coefficient of determination; RPIQ: ratio of performance to interquartile range; NA: not available.

Our proposed method does not require complex pre-processing steps for feature extraction, thereby taking advantage of the end-to-end features of deep learning for prediction tasks. Overall, the results show that the GADF-Swin Transformer is capable of predicting each soil property and that the GADF method can be combined with 1D NIR spectra to generate 2D images that can be effectively processed with the Swin Transformer, even in the case of highly similar and covariant soil spectra.

4. Conclusions

In this paper, we first analyse the challenges associated with predicting soil properties using the LUCAS spectral database, explore and discuss the 2D spectral image-based methods for 1D spectral sequence transformation, and propose the GADF-Swin Transformer method in conjunction with the Swin Transformer. In this approach, 2D spectral images generated based on 1D spectral sequences are used to predict soil properties with networks of different depths. Then, the proposed

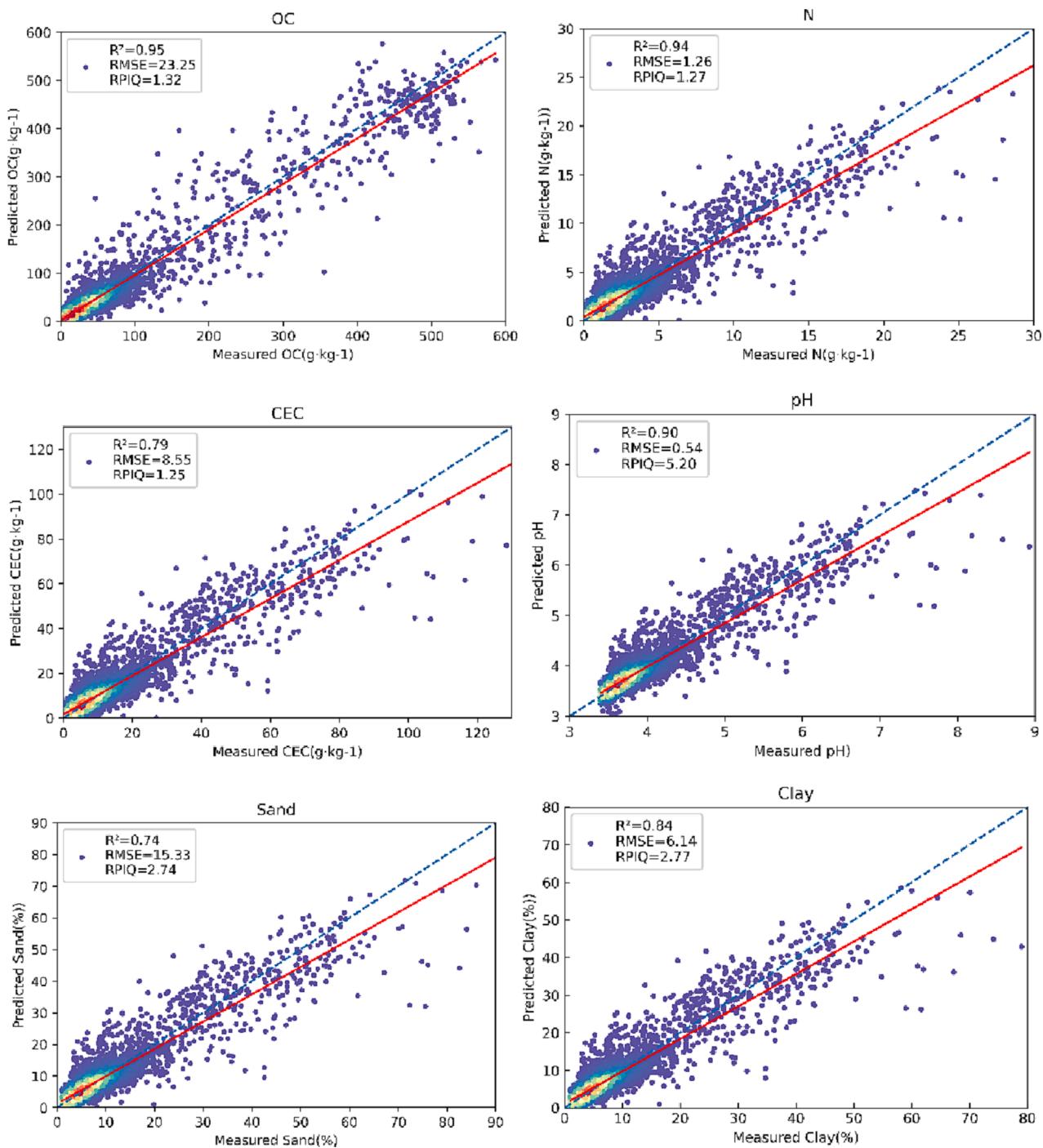


Fig. 9. GADF-Swin Transformer scatter plot of measured and predicted values for the six soil properties. The blue dashed line is the 1:1 line, while the red solid line is the curve according to the simple linear regression model.

approach was compared with methods used in the literature, and the advantages of our approach were verified. It was confirmed that the transformation-based 2D approach is suitable for deep network models and is beneficial for improving the prediction accuracy. Additionally, the GADF-Swin Transformer method was explored in detail to demonstrate the training process and assess results based on soil spectral images. The model was able to successfully predict soil properties. Compared to the commonly used modelling methods in the literature, our approach based on the LUCAS database yielded good results.

Future work may focus on improving two-dimensional coding techniques. Different spectral pre-processing techniques should be explored to reduce the number of unwanted features in the generated spectral

images. As soil spectral databases continue to emerge from different regions of the world, different deep learning algorithms should be explored in the future to predict soil properties more accurately from large spectral datasets.

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.

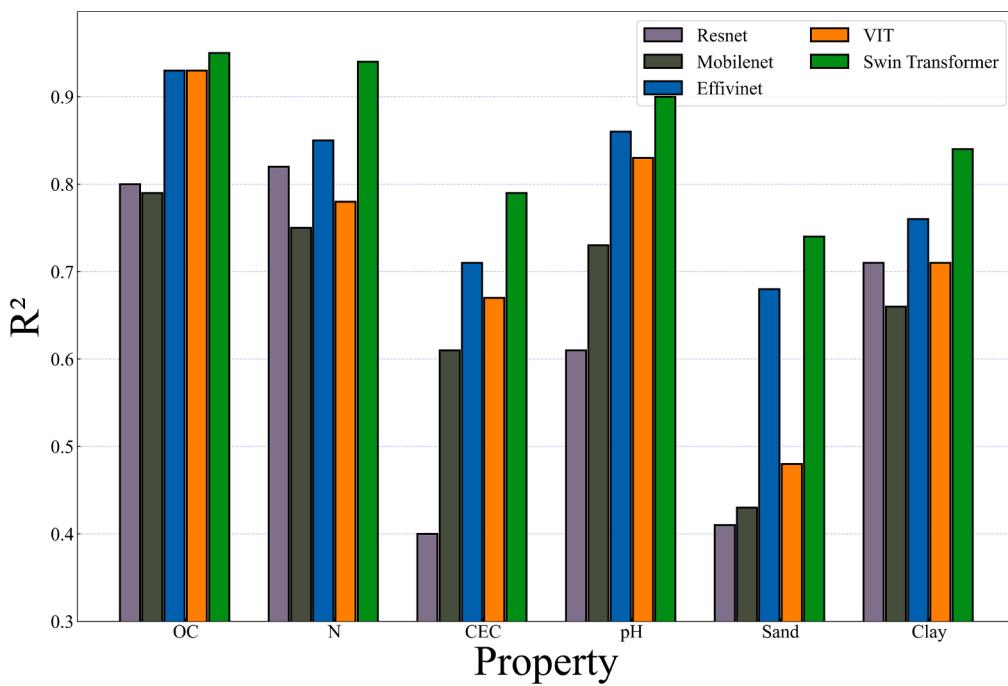


Fig. 10. Results of the GRDF-Swin Transformer and other models for soil property prediction.

Table 7

Comparison between the best models found in this work with deep learning methods proposed by the previous studies for prediction of soil properties using the Vis-NIR spectral data of the LUCAS database.

Soil property	Best model			2D-CNN (Padarian et al., 2019)		1D LSTM (Singh and Kasana, 2019)		1D-CNN (Yang et al., 2020)		1D-CNN (Tsakiridis et al., 2020)		
	RMSE	R ²	RPIQ	RMSE	R ²	RMSE	R ²	RMSE	R ²	RMSE	R ²	RPIQ
OC (g·kg ⁻¹)	23.25	0.95	1.32	32.14	0.88	23.25	0.94	6.40	0.73	11.59	0.84	1.97
N (g·kg ⁻¹)	1.26	0.94	1.27	1.54	0.83	1.15	0.91	0.45	0.70	0.69	0.82	2.04
CEC (cmol ⁺ ·kg ⁻¹)	8.55	0.79	1.25	8.58	0.66	6.75	0.77	3.30	0.73	4.44	0.81	2.61
pH (in water)	0.54	0.90	5.20	0.50	0.87	0.42	0.90	0.35	0.86	0.36	0.93	7.01
Sand (%)	15.33	0.74	2.74	18.15	0.53	NA	NA	NA	NA	13.01	0.74	3.46
Clay (%)	6.14	0.84	2.77	7.55	0.70	NA	NA	NA	MA	5.14	0.84	3.50

Data availability

The authors do not have permission to share data.

Acknowledgements

The LUCAS topsoil dataset used in this work was made available by the European Commission through the European Soil Data Centre managed by the Joint Research Centre (JRC), <https://esdac.jrc.ec.europa.eu/content/lucas-2009-topsoil-data>.

This study is supported by Key Research and Development Project of Anhui Province in 2022 (Project No: 202204c06020022), Key Research Project of Anhui Province in 2022(Project No: 2022AH040125) and Nature Science Research Project of Education Department in Anhui Province (Project No: KJ2020A0108).

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.geoderma.2023.116555>.

References

- Adhikari, K., Hartemink, A.E., 2016. Linking soils to ecosystem services—a global review. *Geoderma* 262, 101–111.

- Beattie, J.R., Esmonde-White, F.W., 2021. Exploration of principal component analysis: deriving principal component analysis visually using spectra. *Appl. Spectrosc.* 75 (4), 361–375.
- Chang, C.-W., Laird, D.A., Mausbach, M.J., Hurlburgh, C.R., 2001. Near-infrared reflectance spectroscopy–principal components regression analyses of soil properties. *Soil Sci. Soc. Am. J.* 65 (2), 480–490.
- Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., 2020. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv: 2010.11929*.
- Guo, C., Li, H., Pan, D., 2010. An improved piecewise aggregate approximation based on statistical features for time series mining. In: Knowledge Science, Engineering and Management: 4th International Conference, KSEM 2010, Belfast, Northern Ireland, UK, September 1–3, 2010. Proceedings 4. Springer, pp. 234–244.
- Han, K., Xiao, A., Wu, E., Guo, J., Xu, C., Wang, Y., 2021. Transformer in transformer. *Adv. Neural Inf. Proces. Syst.* 34, 15908–15919.
- He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 770–778.
- Howard, A.G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M., Adam, H., 2017. Mobilenets: efficient convolutional neural networks for mobile vision applications. *arXiv preprint arXiv:1704.04861*.
- Jin, X., Jie, L., Wang, S., Qi, H.J., Li, S.W., 2018. Classifying wheat hyperspectral pixels of healthy heads and Fusarium head blight disease using a deep neural network in the wild field. *Remote Sens. (Basel)* 10 (3), 395.
- Keesstra, S.D., Bouma, J., Wallinga, J., Tittonell, P., Smith, P., Cerdà, A., Montanarella, L., Quinton, J.N., Pachepsky, Y., van der Putten, W.H., Bardgett, R.D., Moolenaar, S., Mol, G., Jansen, B., Fresco, L.O., 2016. The significance of soils and soil science towards realization of the United Nations Sustainable Development Goals. *Soil* 2 (2), 111–128.
- Khan, S., Naseer, M., Hayat, M., Zamir, S.W., Khan, F.S., Shah, M., 2022. Transformers in vision: a survey. *ACM Computing Surveys (CSUR)* 54 (10s), 1–41.

- Lehmann, J., Bossio, D.A., Kögel-Knabner, I., Rillig, M.C., 2020. The concept and future prospects of soil health. *Nat. Rev. Earth Environ.* 1 (10), 544–553.
- Liu, L., Ji, M., Buchroithner, M., 2018. Transfer learning for soil spectroscopy based on convolutional neural networks and its application in soil clay content mapping using hyperspectral imagery. *Sensors* 18 (9), 3169.
- Ndikumana, E., Ho Tong Minh, D., Baghdadi, N., Courault, D., Hossard, L., 2018. Deep recurrent neural network for agricultural classification using multitemporal SAR Sentinel-1 for Camargue, France. *Remote Sens.* 10 (8), 1217.
- Ng, W., Minasny, B., Montazerolghaem, M., Padarian, J., Ferguson, R., Bailey, S., McBratney, A.B., 2019. Convolutional neural network for simultaneous prediction of several soil properties using visible/near-infrared, mid-infrared, and their combined spectra. *Geoderma* 352, 251–267.
- Nocita, M., Stevens, A., van Wesemael, B., Aitkenhead, M., Bachmann, M., Barthès, B., Dor, E.B., Brown, D.J., Clairotte, M., Csorba, A., 2015. Soil spectroscopy: an alternative to wet chemistry for soil monitoring. *Adv. Agron.* 132, 139–159.
- Orgiazzi, A., Ballabio, C., Panagos, P., Jones, A., Fernández-Ugalde, O., 2018. LUCAS Soil, the largest expandable soil dataset for Europe: a review. *Eur. J. Soil Sci.* 69 (1), 140–153.
- Padarian, J., Minasny, B., McBratney, A.B., 2019. Using deep learning to predict soil properties from regional spectral data. *Geoderma Reg.* 16, e00198.
- Panagos, P., Van Liedekerke, M., Jones, A., Montanarella, L., 2012. European soil data centre: response to European policy support and public data requirements. *Land Use Policy* 29 (2), 329–338.
- Rhodes, C.J., 2014. Soil erosion, climate change and global food security: challenges and strategies. *Sci. Prog.* 97 (2), 97–153.
- Singh, S., Kasana, S.S., 2019. Estimation of soil properties from the EU spectral library using long short-term memory networks. *Geoderma Reg.* 18, e00233.
- Smith, P., Soussana, J.-F., Angers, D., Schipper, L., Chenu, C., Rasse, D.P., Batjes, N.H., Egmond, F., McNeill, S., Kuhnen, M., Arias-Navarro, C., Olesen, J.E., Chirinda, N., Fornara, D., Wollenberg, E., Álvaro-Fuentes, J., Sanz-Cobena, A., Klumpp, K., 2020. How to measure, report and verify soil carbon change to realize the potential of soil carbon sequestration for atmospheric greenhouse gas removal. *Glob. Chang. Biol.* 26 (1), 219–241.
- Stenberg, B., Rossel, R.A.V., Mouazen, A.M., Wetterlind, J., 2010. Visible and near infrared spectroscopy in soil science. *Adv. Agron.* 107, 163–215.
- Stevens, A., Nocita, M., Tóth, G., Montanarella, L., van Wesemael, B., Chen, H.Y.H., 2013. Prediction of soil organic carbon at the European scale by visible and near infrared reflectance spectroscopy. *PLoS One* 8 (6), e66409.
- Tan, M., Le, Q., 2019. Efficientnet: rethinking model scaling for convolutional neural networks. International conference on machine learning. PMLR 6105–6114.
- Tavakoli, H., Correa, J., Sabetizade, M., Vogel, S., 2023. Predicting key soil properties from Vis-NIR spectra by applying dual-wavelength indices transformations and stacking machine learning approaches. *Soil Tillage Res.* 229, 105684.
- Tian, H., Wang, T., Liu, Y., Qiao, X., Li, Y., 2020. Computer vision technology in agricultural automation—a review. *Inform. Process. Agric.* 7 (1), 1–19.
- Tsakiridis, N.L., Keramiris, K.D., Theocharis, J.B., Zalidis, G.C., 2020. Simultaneous prediction of soil properties from VNIR-SWIR spectra using a localized multi-channel 1-D convolutional neural network. *Geoderma* 367, 114208.
- Tsimpouris, E., Tsakiridis, N.L., Theocharis, J.B., 2021. Using autoencoders to compress soil VNIR-SWIR spectra for more robust prediction of soil properties. *Geoderma* 393, 114967.
- Veres, M., Lacey, G., Taylor, G.W., 2015. Deep learning architectures for soil property prediction. In 2015 12th Conference on Computer and Robot Vision. IEEE, pp. 8–15.
- Viscarra Rossel, R.A., Behrens, T., Ben-Dor, E., Brown, D.J., Dematté, J.A.M., Shepherd, K.D., Shi, Z., Stenberg, B., Stevens, A., Adamchuk, V., Aichi, H., Barthès, B.G., Bartholomaeus, H.M., Bayer, A.D., Bernoux, M., Böttcher, K., Brodsky, L., Du, C.W., Chappell, A., Fouad, Y., Genot, V., Gomez, C., Grunwald, S., Gubler, A., Guerrero, C., Hedley, C.B., Knadel, M., Morrás, H.J.M., Nocita, M., Ramirez-Lopez, L., Roudier, P., Campos, E.M.R., Sanborn, P., Sellitto, V.M., Sudduth, K.A., Rawlins, B.G., Walter, C., Winowiecki, L.A., Hong, S.Y., Ji, W., 2016. A global spectral library to characterize the world's soil. *Earth Sci. Rev.* 155, 198–230.
- Voita, E., Talbot, D., Moiseev, F., Sennrich, R., Titov, I., 2019. Analyzing multi-head self-attention: Specialized heads do the heavy lifting, the rest can be pruned. arXiv preprint arXiv:1905.09418.
- Wang, Z., Oates, T., 2015. Encoding time series as images for visual inspection and classification using tiled convolutional neural networks. Workshops at the twenty-ninth AAAI conference on artificial intelligence, AAAI Menlo Park, CA, USA.
- Williams, R.J., Zipser, D., 1989. A learning algorithm for continually running fully recurrent neural networks. *Neural Comput.* 1 (2), 270–280.
- Yang, Y., Jiao, L., Liu, X., Liu, F., Yang, S., Feng, Z., Tang, X., 2022. Transformers meet visual learning understanding: a comprehensive review. arXiv preprint arXiv: 2203.12944.
- Yang, J., Wang, X., Wang, R., Wang, H., 2020. Combination of convolutional neural networks and recurrent neural networks for predicting soil properties using Vis-NIR spectroscopy. *Geoderma* 380, 114616.
- Zhang, R., Xie, H., Cai, S., Hu, Y., Liu, G.-K., Hong, W., Tian, Z.-Q., 2020. Transfer-learning-based Raman spectra identification. *J. Raman Spectrosc.* 51 (1), 176–186.
- Zhao, W., Wu, Z., Yin, Z., Li, D., 2022. Attention-based CNN ensemble for soil organic carbon content estimation with spectral data. *IEEE Geosci. Remote Sens. Lett.* 19, 1–5.